

Text Mining in Ideencommunities: Gestaltungsempfehlungen eines Text Mining Artefakts zur Unterstützung der Ideenevaluation

Thomas Maier¹, Sebastian Gronich¹, Ann-Katrin Rickert¹ and Nikolaus Lipusch²

¹University of Kassel, Kassel, Germany
thomas_maier10@web.de, Basti.Oliver@gmx.de, anka.rickert@googlemail.com

²University of Kassel, Information Systems, Kassel, Germany
nikolaus.lipusch@uni-kassel.de

Abstract. Ideencommunities haben sich als effektives Tool zur Erhöhung der Innovationskraft von Unternehmen etabliert. Ein Aspekt, dem bislang wenig Aufmerksamkeit geschenkt wurde, ist die Ideenevaluation in Ideencommunities. Ein Problem im Rahmen der Evaluation ist die Vielzahl an Ideen, die in solchen Communities entsteht und Unternehmen zunehmend vor neue Herausforderungen stellt. So verfügen Unternehmen in der Regel nur über eine begrenzte Kapazität, um das vollständige Potenzial solcher Ideen nutzbar zu machen. Eine Lösung für dieses Problem bieten sogenannte Text Mining Verfahren, die in Verbindung mit neuen Informations- und Kommunikationstechnologien Unternehmen dabei unterstützen können, Ideen effizienter auszuwerten sowie neue Datenzusammenhänge und somit wertvolle Information zu erkennen. Vor diesem Hintergrund untersucht der Beitrag welche Phasen der Ideenevaluierung am effizientesten durch ein Text Mining Tool unterstützt werden können und leitet daraus konkrete Gestaltungsempfehlungen für ein Text Mining Tool für Ideencommunities ab. Der Beitrag umfasst dabei konkrete Gestaltungsempfehlungen, die Praktiker bei der Umsetzung eines solchen Tools unterstützen sollen. Der theoretische Beitrag liegt in der Erweiterung des Erkenntnisstandes hinsichtlich neuer Methoden zur Unterstützung der effizienten Auswertung von Ideen im Crowdsourcing

Keywords: Ideencommunities, Absorptive Capacity, Text Mining, Design Science Research

13th International Conference on Wirtschaftsinformatik,
February 12-15, 2017, St. Gallen, Switzerland

Maier, T.; Gronich, S.; Rickert, A.-K.; Lipusch, N. (2017): Text Mining in Ideencommunities: Gestaltungsempfehlungen eines Text Mining Artefakts zur Unterstützung der Ideenevaluation, in Leimeister, J.M.; Brenner, W. (Hrsg.): Proceedings der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), St. Gallen, S. 1405-1416

1 Einleitung

In jüngerer Vergangenheit hat Crowdsourcing immer mehr an Popularität gewonnen [1]. Eine spezielle Form des Crowdsourcing, die von Unternehmen zunehmend aktiv eingesetzt wird, sind sogenannte Ideencommunities. Es handelt sich dabei zumeist um Plattformen, die es Unternehmen ermöglichen bestimmte Probleme zu posten und in Zusammenarbeit mit der Crowd Lösungen in Form von Ideen zu generieren. Das Ziel von Ideencommunities ist es dabei, von den Erfahrungen und Ideen der Kunden zu profitieren und dadurch die Innovationskraft des Unternehmens zu erhöhen. Die Vorteile von Ideencommunities wurden in der Literatur umfangreich belegt [2]. Das Unternehmen Lego kann hierbei exemplarisch für den Aufbau einer solchen Ideencommunity hervorgehoben werden [3]. Durch das interaktive Kollaborieren und Zusammenwirken der Crowd wird der dem Ansatz zugrunde liegende Kerngedanke des Nutzbarmachens der sogenannten „Weisheit der Vielen“ deutlich [4, 5].

Trotz der genannten Vorteile, die Ideencommunities mit sich bringen, ist deren effiziente Nutzung auch mit großen Herausforderungen verbunden. So entsteht durch die rege Teilnahme an Ideencommunities eine Vielzahl an Beiträgen. Das Resultat ist eine große Menge an unstrukturierten Daten. My Starbucks Idea umfasst beispielsweise über mehrere Kategorien hinweg ca. 150.000 Ideenvorschläge, Dell Idea Storm über 25.000 (Stand: September 2016). Die große Menge an textuellen Informationen macht es dabei in der Regel schwierig das volle Potenzial von Ideencommunities zu nutzen [6, 7]. Unternehmen, als Betreiber solcher Ideencommunities, sehen sich dabei mit der Herausforderung konfrontiert, dass sie nur über begrenzte Ressourcen und Kapazitäten verfügen, um diese Daten auszuwerten. Die beschriebene Problematik wird in der Literatur unter dem Begriff Absorptive Capacity thematisiert. Hierunter wird die Fähigkeit verstanden, Wissen aus externen Quellen zu identifizieren, assimilieren und anschließend anzuwenden [8]. Die Mengen an unstrukturierten Daten, die im Rahmen von Ideencommunities generiert werden, stellen eine besondere Herausforderung für die Absorptive Capacity von Unternehmen dar. Eine Möglichkeit, Unternehmen bei der Auswertung von Ideencommunities zu unterstützen, bieten moderne Informations- und Kommunikationstechnologien. Neben der effektiven Einbeziehung von Kunden bergen diese das Potenzial Organisationsroutinen und die Koordinationsaktivitäten von Unternehmen zu verbessern und Kosten zu senken [9, 10].

Ein vielversprechender Ansatz, um Unternehmen bei der IT-gestützten Auswertung von Ideen zu unterstützen, bilden sogenannte Text Mining Anwendungen [41]. Es handelt sich dabei um Tools, die eine Kombination aus computergestützten linguistischen, statistischen und mathematischen Verfahren anwenden, um Wissen aus großen Mengen an unstrukturierten Textdaten heraus zu filtern und neue Zusammenhänge zu erkennen [11]. Obwohl von einem hohen Potenzial von Text Mining Lösungen für Ideencommunities ausgegangen werden kann, blieb eine kombinierte Betrachtung dieser beiden Forschungsgebiete bislang unberücksichtigt [1, 12]. Demnach wurde lediglich partiell abgedeckt, wie sich eine Text Mining Lösung systematisch für Ideencommunities nutzen lässt. Vor diesem Hintergrund ist es Ziel dieses Papers potenzielle Einsatzgebiete und Unterstützungspotenziale eines Text

Mining Tools zur Ideenevaluierung in Ideencommunities zu identifizieren und Unternehmen Gestaltungsempfehlungen für ein Tool zur automatisierten und somit effizienteren Ideenevaluation an die Hand zu geben. Das generelle Vorgehen im Rahmen der Studie orientiert sich dabei am Design Science Research (DSR)-Ansatz nach Vaishnavi und Kuechler [13].

Der sich hieraus erschließende Beitrag bezieht sich auf zwei Bereiche. Zum einen geben wir Praktikern Empfehlungen zum Einsatz und zur Gestaltung eines Text Mining Tool an die Hand, die darauf abzielen den Prozess der Ideenevaluierung in Ideencommunities zu automatisieren und Unternehmen bei der Analyse von Ideen zu unterstützen. Der theoretische Beitrag dieser Arbeit liegt in einer Erweiterung des Erkenntnisstandes zur Ideenauswertung im Crowdsourcing. Das Ziel hierbei ist es die Phase der Ideenauswertung näher zu beleuchten. Speziell die einzelnen Prozessschritte, die mit der Phase der Ideenevaluation einhergehen, werden in der vorliegenden Arbeit betrachtet, da die genaue Untersuchung dieser Thematik in der Literatur bisweilen noch keinen Eingang fand. Darüber hinaus soll gezeigt werden, wie ein Text Mining Tool in den einzelnen Prozessschritten am besten unterstützen kann. Wir setzen hierbei an vorangehenden Arbeiten zum Crowdsourcing an, die zu dem Erkenntnis kommen, dass die Phase der Ideenevaluation ein großes Potenzial für die Effizienzsteigerung und die breite Anwendung von Crowdsourcingaktivitäten birgt.

Das folgende Paper gliedert sich wie folgt. Nach der Einführung des theoretischen Hintergrunds wird die zugrunde gelegte Methode in Form des DSR-Ansatzes genauer erläutert. Im Anschluss an die Ermittlung des Einsatzpotenzials eines Text Mining Tools zur Unterstützung der Ideenevaluierung und der konkreten Ableitung von Gestaltungsempfehlungen, welche ein Text Mining Tool erfüllen muss, werden die weiteren geplanten Schritte für die Entwicklung dieses Tools dargelegt. Das Paper schließt mit einer Zusammenfassung und dem erwarteten Beitrag für Praxis und Wissenschaft.

2 Theoretischer Hintergrund

2.1 Absorptive Capacity Theorie

Das Grundkonzept der Absorptive Capacity wird in der Fachliteratur als Fähigkeit von Unternehmen definiert „to recognize the value of new, external information, assimilate it, and apply it to commercial ends“ [8]. Aus der Definition ist abzuleiten, dass sich die Absorptionsfähigkeit in sequentielle Teilfähigkeiten differenzieren lässt [14].

Zunächst beschreibt es die Fähigkeit eines Unternehmens neue, externe Informationen zu identifizieren und dabei den zu schöpfenden, potenziellen Wert zu erkennen. In einem weiteren Schritt müssen im Zuge der Assimilationsfähigkeit die Informationen an die Bedürfnisse adäquat angepasst bzw. an das bestehende Wissen des Unternehmens angeknüpft werden. Abschließend kann das Wissen angewendet bzw. kommerziell eingesetzt werden [8, 15]. Dieses klassische Konzept zur Absorptive Capacity nach Cohen und Levinthal [8] wurde in den darauffolgenden Jahren insbesondere durch die Beiträge von Zahra und George [16] sowie Todorova

und Durisin [17] theoretisch diskutiert, konkretisiert und erweitert. Lane et al. [18] verstehen Absorptive Capacity als eines der zentralen Konstrukte, wie Unternehmen oder Organisationseinheiten Wissen einarbeiten, welches sich außerhalb ihrer Grenzen befindet. Im Zuge des Trends Richtung Open Innovation werden Kunden in die Gestaltung von Innovationsprozessen der Unternehmen aktiv integriert [19]. In diesem Zusammenhang werden Ideencommunities als Instrumente eingesetzt, um Ideen bzw. Wissen zu generieren. Die Herausforderung liegt explizit darin, die Unmengen an unstrukturierten Daten den Bedürfnissen eines Unternehmens anzupassen, um anschließend relevantes Wissen absorbieren und anwenden zu können. Zur effizienten Steuerung der Informationen im Unternehmen legen Roberts et al. [9] den Einsatz von Informationstechnologien dar, welche in Form von automatisierten Anwendungstools in den Absorbierungsprozess integriert werden. Das vorliegende Paper setzt sich mit der Verwendung eines Text Mining Tools auseinander, welches unter Berücksichtigung der Effizienzsteigerung im Absorbierungsprozess eingesetzt werden soll.

2.2 Grundlagen zum Crowdsourcing

Der Neologismus aus den beiden Wörtern „outsourcing“ und „crowd“ wird in Anlehnung an Leimeister [20] als ein Interaktionsprozess definiert. Eine Unternehmung, Organisation, Gruppe oder ein Individuum, auch Crowdsourcer genannt, bietet einer nicht genauer definierten großen Gruppe, den sogenannten Crowdsources, die Möglichkeit der Bearbeitung einer Aufgabe über einen offenen Aufruf. Ziel ist es, sich das aggregierte Wissen der Masse zu Nutze zu machen. Die Crowdsources, welche sich ebenfalls aus Individuen, formellen sowie informellen Gruppen, Organisationen oder Unternehmen zusammensetzen können, übernehmen anschließend die Aufgabenbearbeitung [20].

Eine besondere Ausprägungsform dieses genannten Ansatzes sind von Unternehmen initiierte und betriebene, IT-basierte Ideencommunities. Sie ermöglichen die Teilnahme von Individuen außerhalb des Unternehmens, insbesondere von Kunden, am unternehmerischen Innovationsprozess [7]. Kunden haben auf Ideencommunities die Möglichkeit, eigene Erfahrungen, Verbesserungsvorschläge sowie Innovationsideen bekanntzugeben, Vorschläge anderer zu bewerten, interaktiv mit Hilfe von Kommentar- und Wiki-Funktionen weiterzuentwickeln und auf diese Weise aktiv zur Produkt- und Serviceentwicklung beizutragen [7].

Die Anwender und Nutzer der Ideencommunities generieren im Rahmen dieses interaktiven Prozesses eine Menge an Ideen, die als textuelle Informationen gespeichert werden. Dieses wird bisher in einem ressourcenintensiven, manuellen Verfahren durch User sowie Experten evaluiert. Die auf eine Person begrenzte und wertvolle Arbeitskraft, die für diesen Auswertungsprozess benötigt wird, steht anschließend nicht mehr für die angemessene Analyse der vielen Beiträge und Ideen zur Verfügung. Darüber hinaus kann die Analyse der Ideen aufgrund der Masse an textueller Information nicht adäquat vorgenommen werden, denn häufig fokussieren die am Analyseprozess beteiligten Experten unbewusst auf einige wenige Ideen, ohne dabei die Gesamtheit der Ideen zu berücksichtigen. Aufgrund dieses zeitintensiven

und subjektiven Auswertungsvorgehens gelingt es Unternehmen folglich nicht das gesamte Potenzial der Ideen zu nutzen. Eine Möglichkeit den zuvor genannten Problemen zu begegnen und dabei die Absorbierbarkeit von Unternehmen zu unterstützen, bildet der Einsatz eines automatisierten Text Mining Tool.

2.3 Grundlagen zum Text Mining

Der Begriff Text Mining bezeichnet einen Ansatz, welcher sich aus verschiedenen Verfahren zusammensetzt. Durch die Anwendung unterschiedlicher Techniken werden neue Informationen aus unstrukturierten Texten generiert und diese automatisch extrahiert [21]. Text Mining wird dem Bereich des Data Mining zugeordnet, mit dem Unterschied, dass Informationen im Gegensatz zum Data Mining aus unstrukturierten Daten gewonnen werden [21, 22]. In der Literatur wird überwiegend ein Prozessverlauf thematisiert, welcher das Text Mining Verfahren in zwei Phasen aufteilt [23, 24].

Die erste Phase stellt hierbei die Datenvorbereitungsphase dar. Der unstrukturiert vorliegende Text muss in ein Format transferiert werden, aus dem ein Text Mining Algorithmus Informationen gewinnen kann. Dieser Vorbereitungsprozess bedient sich dabei computerlinguistischer Methoden in Form der morphologischen, syntaktischen und semantischen Analyse [23]. Im Bereich der morphologischen Analyse wird durch die sogenannte Tokenisierung (Entfernen von Satz- und Leerzeichen; Generierung von Tokens), das Filtern (Stoppwortentfernung) sowie das Stemming (Rückführung der Wörter zu Wortstämmen) der Text für den weiteren Verlauf vorbereitet [24, 25]. Ferner behandelt die syntaktische Analyse den Textinhalt und versucht durch Worterkennungen mit Hilfe syntaktischer Regeln einzelne Phasen in strukturierte Syntaxbäume zu bringen. Darüber hinaus werden hierdurch das Verhältnis sowie die gegenseitige Modifikation einzelner Phrasen erkannt [26]. Die semantische Analyse hingegen behandelt Mehrdeutigkeiten der einzelnen Wörter und deren Bezug zum Gesamtkontext [27]. Das Ziel dieser ersten Phase ist demnach die Komplexitätsreduktion, um einen bereinigten und nur aus potenziell interessanten Wörtern bestehenden Text zu erhalten, der in einem nächsten Schritt bearbeitet werden kann [28].

In der zweiten Phase ist es nun durch das Nutzen von Mining-Techniken möglich, wichtige Informationen aus dem überarbeiteten Text zu gewinnen. Daten können hierbei vordefinierten Klassen durch Verfahren wie Naïve Bayes, Logistic regression, Entscheidungsbäume, Neuronale Netzwerke, Support Vector Machines, MARSplines, k-nearest neighbors zugeordnet werden [29, 30, 31, 32, 33]. Clustering-Methoden wie das k-means Clustering [25] können sowohl einzeln, als auch in Kombination mit anderen Methoden angewandt werden, um Gruppen mit denselben Inhalten zu bilden.

Abschließend wurde herausgearbeitet, wie sich Text Mining Methoden auf den Kontext von Ideencommunities anwenden lassen, um aus einer unstrukturierten Menge an Daten wichtige Informationen zu extrahieren, diese in einem Prozess für die weiteren Bearbeitungsschritte zu selektieren, um schlussendlich einen Beitrag für eine effiziente Ideenevaluation leisten zu können.

3 Methode

3.1 Design Science Research

Das vorliegende Paper bezieht sich auf den DSR-Ansatz nach Vaishnavi und Kuechler [13]. Dieser Ansatz konzentriert sich auf die Konstruktion sowie Evaluation von sogenannten Artefakten, welche zur Lösung eines vorab definierten Problems beitragen [35, 36]. Artefakte können hierbei mehrere Ausprägungen annehmen, wie beispielsweise Konstrukte, Modelle, Methoden, Instanziierungen oder wie unserem Fall Design-Guidelines [37].

Das Ziel dieses Papers ist es, Einsatzgebiete und Unterstützungspotenziale eines Text Mining Tools im Rahmen der Ideenevaluierung zu identifizieren, sowie entsprechende Gestaltungsempfehlungen in Form von Design-Guidelines für Unternehmen abzuleiten. Die Design-Guidelines dienen als Basis für die Gestaltung eines Text Mining Tools zur Unterstützung der Ideenanalyse in Ideencommunities. Das Tool soll dabei konzipiert werden um eine große Menge an unstrukturiert vorliegenden Textdaten automatisiert aufzubereiten und auszuwerten. Es dient dabei, Experten in der Ideenevaluierung zu unterstützen, indem es eine effektivere Nutzung des Potenzials vorhandener Daten erlaubt.

Vaishnavi und Kuechler [13] untergliedern den DSR-Ansatz in fünf Phasen: Awareness of Problem, Suggestion, Development, Evaluation und Conclusion. In unserem Fall wurden die ersten drei Phasen konkret bearbeitet.

In der ersten Phase, *Awareness of Problem*, wurde zunächst der Prozess der Ideenevaluierung in Ideencommunities mit Hilfe einer systematischen Inhaltsanalyse abgeleitet, um anschließend Probleme in den einzelnen Prozessphasen identifizieren zu können. Insgesamt wurden im Rahmen der Analyse die folgenden fünf Ideencommunities untersucht: My Starbucks Idea, SAPiens, Dell Ideastorm, LEGO Ideas und Tchibo Ideas. Bei den Betreibern der Plattformen handelt es sich um Unternehmen aus unterschiedlichen Branchen, die mit ihren Ideencommunities primär Endnutzer (B2C-Bereich) ansprechen. Darüber hinaus ähneln sich die betrachteten Communities stark im Hinblick auf ihren Aufbau und ihre Struktur. So erlauben es alle Communities Probleme zu posten und Ideen durch User generieren und bewerten zu lassen. Die Analyse folgte einem iterativen Vorgehen, indem 3 Autoren die oben genannten Plattformen unabhängig voneinander untersuchten und anschließend Probleme im Hinblick auf die Ideenevaluierung identifizierten.

In der zweiten Phase, *Suggestion*, wurden im Zuge von Interviews mit drei Crowdsourcing Experten potenzielle Einsatzgebiete von Text Mining im Ideenevaluierungsprozess erarbeitet. Bei den Crowdsourcing Experten handelte es sich um Spezialisten mit umfassender Erfahrung hinsichtlich der Durchführung von Crowdsourcingprojekten sowie um Personen, die sich umfassend mit dem Thema aus praktischer und wissenschaftlicher Sicht beschäftigt haben. Als Grundlage für die Interviews dienten der zuvor abgeleitete Prozess und die identifizierten Probleme.

In der dritten Phase, *Development*, wurden auf Basis der zuvor identifizierten Probleme konkrete Gestaltungsempfehlungen in Form von Design-Guidelines abgeleitet. Zur konkreten Ausgestaltung des Text Mining Tools wurden Interviews mit

drei Text Mining Experten geführt. Hierbei handelt es sich um Spezialisten auf dem Gebiet der Entwicklung und Implementierung von Text Mining Lösungen. Ein Interview wurde mit dem Geschäftsführer eines Unternehmens geführt, welches sich auf Text Mining Lösungen spezialisiert hat. Die beiden weiteren Interviewpartner verfügen über langjährige Erfahrung in der Implementierung von Text Mining Lösungen und mit der Betreuung von Text Mining Projekten.

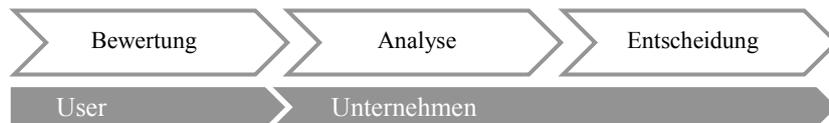
Die weiterführenden Prozessphasen, *Evaluation* und *Conclusion*, werden in Kapitel 5 betrachtet.

3.2 Vorläufige Ergebnisse (Anforderungen an ein Text Mining Tool)

In diesem Paper können wir bereits vorläufige Ergebnisse präsentieren, welche die Ziele und damit einhergehenden Anforderungen eines Text Mining Tools in Bezug auf Ideencommunities kennzeichnen.

In dem ersten Prozessschritt *Awareness of Problem* wurde zunächst, auf Grundlage einer systematischen Inhaltsanalyse von mehreren Ideencommunities, der Prozess zur Ideenevaluierung abgeleitet (siehe Abbildung 1) und damit einhergehend potenzielle Problemfelder beleuchtet.

Abbildung 1. Bisheriger Prozess der Ideenevaluierung



Wie in Abbildung 1 ersichtlich gliedert sich der Prozess der Ideenevaluierung in drei Phasen. In Phase eins, *Bewertung*, erfolgt eine initiale Bewertung der Ideen durch die User. Dies geschieht über das Verfassen von Kommentaren oder durch die Anwendung von Bewertungsmechanismen wie Ratingskalen. In Phase zwei, *Analyse*, werden die Ideen durch das Unternehmen analysiert. Diese Phase ist dadurch gekennzeichnet, dass Unternehmen eine Vielzahl an Beiträgen und Kommentaren analysieren sowie bewerten müssen. In Phase drei, *Entscheidung*, entscheidet das Unternehmen, wie und mit wem die Ideen weiterentwickelt bzw. welche Ideen weiterverfolgt werden. Wie aus der obigen Abbildung ersichtlich werden Beiträge und die daraus resultierende Menge an unstrukturierten Textdaten im bisherigen Prozesses der Ideenevaluierung durch User und durch Experten evaluiert. Dem Prozess liegt dabei eine stark subjektive Betrachtungs- und Vorgehensweise zu Grunde.

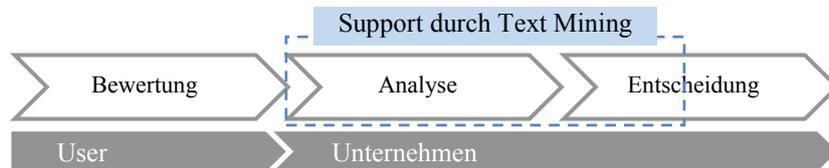
In einem weiteren Schritt wurden die Problemfelder im Prozess der Evaluierung von Ideen näher betrachtet. Während bei My Starbucks Idea, Dell Ideastorm und LEGO Ideas eine hohe Anzahl an generierten Beiträgen auffallend ist, weist SAPIens und Tchibo Ideas bisher noch keine bzw. eine beschränkte Themenkategorisierung der Beiträge auf.

Im anknüpfenden Prozessschritt, *Suggestion*, wurde der abgeleitete Prozess der Ideenevaluierung und die identifizierten Probleme im Rahmen von Interviews mit

Crowdsourcing Experten diskutiert. Die identifizierten Probleme beziehen sich hauptsächlich auf die Analyse Phase des Ideenevaluierungsprozesses.

In den mit den Experten geführten Interviews wurden die zuvor identifizierten Probleme bestätigt und konkretisiert. Die Experten kamen dabei zu dem Schluss, dass vor allem die Vielzahl an Beiträgen und fehlende Themenkategorisierung die Unternehmen vor große Probleme stellt. Durch die subjektive Betrachtung einerseits und das arbeits- sowie zeitintensive manuelle Vorgehen andererseits werden Organisationen an Kapazitätsgrenzen geführt. Der Einsatz von automatisierten Auswertungsverfahren im Sinne von Text Mining Tools war den Experten bei Ideencommunities nicht bekannt. Ferner konnte in den Experteninterviews erarbeitet werden, welche Phasen im bisherigen Ideenevaluationsprozess das höchste Potenzial für die Unterstützung durch ein TextMining Tool bergen.

Abbildung 2. Unterstützende Funktion eines Text Mining Tool in der Ideenevaluierung



Aus den Interviews mit den Experten lässt sich dabei ableiten, dass ein Text Mining Tool insbesondere in der Analysephase einen strukturierten bzw. gesamtheitlichen Überblick über die Ideen darlegt. Indirekt unterstützt das Tool die Entscheidungsphase, da Unternehmen zusätzliche Informationen erhalten, die Unternehmen bei der Wahl der geeigneten Ideen helfen.

Nach der Identifizierung potenzieller Einsatzgebiete und Unterstützungspotenzials eines Text Mining Tools im Rahmen des Prozesses der Ideenevaluierung wurde in der *Development*-Phase in Zusammenarbeit mit Text Mining-Experten die konkrete Ausgestaltung eines Text Mining Tools, in Form von Design-Guidelines, erarbeitet.

Tabelle 1. Darstellung konkreter Probleme und dazugehörige Design-Guidelines

Kategorien	Probleme	Design-Guidelines
Dokumenten-aufbereitung	Unaufbereitete textuelle Daten	Tool soll Textdokument bereinigen.*
Themen-kategorisierung	Vielfalt an aufbereiteten textuellen Daten	Tool soll Schlüsselwörter extrahieren (Keyword Extraction), deren Häufigkeiten ermitteln und anschließend gewichten.

	Überflüssige/identische Informationen	Tool soll identische Ideen eliminieren.
	Unterschiedliche Formulierungen einer identischen Idee	Tool soll identische Themenbereiche erkennen und entsprechende Oberkategorien bilden.
Priorisierung	Vielfalt an gegensätzlichen User-Bewertungen	Tool soll Bewertungstendenzen innerhalb der Themenbereiche erkennen (Sentiment-Analyse).
	Relevante Informationen erkennen	Tool soll Themenbereiche sowie Ideen innerhalb dieser Bereiche priorisieren/ gewichten.

*Morphologische, syntaktische und semantische Analyse: a) Entfernen von Zahlen, Satz- und Leerzeichen, b) Behandeln von Groß- und Kleinschreibung, Rechtschreibung, c) Entfernen von Stoppwörtern, d) Stemming, e) Erkennen von Satzgrenzen, f) Durchführung des Part-of-Speech-Tagging, g) Bildung von Wortgruppen, h) Durchführung des Parsing und Junking, i) Erkennen von Synonymen, Mehrdeutigkeit der Sprache (Dialekt).

Hierzu wurden in einem ersten Schritt Probleme textueller Daten identifiziert. Die Anforderung der Aufbereitung dieser Daten liegt darin, das Textdokument mit Hilfe einer morphologischen, syntaktischen und semantischen Analyse zu bereinigen. Die Probleme zwei bis vier können der Themenkategorisierung zugeordnet werden. Zunächst liegt der Fokus auf der Vielfalt an textuellen Daten. Hierbei soll das Tool Schlüsselwörter extrahieren, indem durch eine Häufigkeitszählung bestimmte Wörter identifiziert, geclustert und entsprechend gewichtet werden. Des Weiteren sind die unterschiedlichen Formulierungen einer identischen Idee sowie überflüssige als auch identische Informationen in einem Evaluationsprozess als problematisch einzuordnen. Das Tool soll zum einen identische Ideen eliminieren können. Zum anderen soll es identische Themenbereiche erkennen und Oberkategorien bilden. Die Probleme fünf und sechs, der Kategorie Priorisierung zuzuordnen, zielen auf die Vielfalt an widersprüchlichen User-Bewertungen sowie auf das Erkennen von relevanten Informationen ab. Demnach soll das Tool Bewertungstendenzen innerhalb der Themenbereiche erkennen (Sentiment-Analyse) und anschließend die Themenbereiche sowie Ideen innerhalb dieser Bereiche priorisieren bzw. gewichten können.

Anhand der angesetzten Strukturierung der Probleme konnte ein Prozess mit Anforderungen aufgestellt werden. Diesen gilt es zu erfüllen, um das gesetzte Ziel einer automatisierten und effizienten Ideenevaluation zu erreichen.

4 Geplante weitere Schritte für die Entwicklung eines Text Mining Tools

In dem vorherigen Abschnitt haben wir die vorläufigen Ergebnisse in Bezug auf die notwendigen Gestaltungsanforderungen eines Text Mining Artefakts konkret herausgearbeitet. Die geplanten weiteren Schritte für die Entwicklung dieses Text Mining Tools in Anlehnung an den DSR-Ansatz sehen wie folgt aus: In der vierten Phase, *Evaluation*, werden die Gestaltungsanforderungen durch zusätzliche Interviews mit Plattformbetreibern bezüglich der potenziellen Anwendung bzw. Nutzbarkeit analysiert. Ferner wird das Tool entwickelt und in einem abschließenden Schritt erneut evaluiert. Das gewählte Vorgehen zielt darauf ab, den Prozess der Ideenevaluation effizienter zu gestalten und folglich die Nutzung bzw. Verwertung der Ideen aus Ideencommunities zu unterstützen. Aus diesem Grund wird das Tool auf Effizienz im Sinne von Geschwindigkeit untersucht. Mit Hilfe einer einfachen Zeitmessung könnte man die benötigte Dauer des Ideenevaluationsprozesses durch Anwendung des Tools ermitteln und anschließend dem zweistufigen manuellen Prozess durch User und Experten gegenüberstellen.

5 Zusammenfassung

In dem vorliegenden Paper haben wir uns mit der Gestaltung eines Artefakts auf Grundlage des DSR-Ansatzes nach Vaishnavi und Küchler [13] auseinander gesetzt. Das Ziel hierbei war es die Phasen der Ideenevaluierung näher zu beleuchten, Unterstützungspotenziale durch ein Text Mining Tool in den einzelnen Phasen zu identifizieren und auf Basis dessen die konkreten Anforderungen an ein solches Tool abzuleiten. Der sich anschließend hieraus ergebene Beitrag bezieht sich auf die zwei Bereiche Praxis und Wissenschaft.

Zum einen liegt der praktische Mehrwert unseres Beitrags darin Unternehmen Handlungsempfehlungen zum effizienten Einsatz und zur Implementierung eines Text Mining Tools zur Unterstützung der Ideenanalyse in Ideencommunities zu geben. Die Vorteile einer solchen Lösung liegen auf der Hand. So erlaubt die Implementierung einer Text Mining Lösung dem Plattformbetreiber die Ausschöpfung von Automatisierungspotenzial, das zu internen Ressourceneinsparungen in Form von Kosten und Arbeitskapazität führt. Zusätzlich bietet eine Text Mining Lösung den Unternehmen die Erkennung von neuen, bisher nicht entdeckten sowie relevanten Zusammenhängen und dient diesen folglich als Informations- und Entscheidungsunterstützung. Folglich kann das vorhandene Potenzial der vorliegenden Daten besser ausgeschöpft werden.

Der theoretische Mehrwert unseres Beitrags liegt in der Erweiterung des Erkenntnisstandes der bestehenden Crowdsourcing Literatur. Dieser wird einerseits erreicht durch die gezielte sowie der tiefergehende Betrachtung und Analyse der Ideenevaluierung im Crowdsourcing. Bisherige Studien zu dem Thema beschäftigten sich primär damit wie sich die Crowd in die Evaluation von Ideen einbinden lässt [38] und welche Individuen sich am besten für die Bewertung von Ideen eignen [39]. Zum

anderen wird die bestehende Literatur durch die in dieser Studie vorgestellten Gestaltungsempfehlungen für ein Text Mining Tool um eine neue Forschungsperspektive erweitert. Unser Beitrag liefert dabei eine Design Theorie [40] zur Gestaltung eines Textmining Artefakts zur effizienten Unterstützung der Ideenevaluation im Crowdsourcing.

Literaturverzeichnis

1. Zogaj, S., Bretschneider, U., Leimeister, J.M.: Managing Crowdsourced Software Testing – A Case Study Based Insight on the Challenges of a Crowdsourcing Intermediary. *Journal of Business Economics (JBE)* (DOI: 10.1007/s11573-014-0721-9) (2014)
2. Poetz, M.K., Schreier, M.: The value of crowdsourcing: can users really compete with professionals in generating new product ideas?. *Journal of Product Innovation Management*, 29(2), 245-256 (2012)
3. Hienrth, C., Lettl, C., Keinz, P.: Synergies among producer firms, lead users, and user communities: The case of the LEGO producer–user ecosystem. *Journal of Product Innovation Management*, 31(4), 848-866 (2014)
4. Blohm, I., Bretschneider, U., Leimeister, J.M., Krcmar, H.: Does collaboration among participants lead to better ideas in IT-based idea competitions? An empirical investigation. *International Journal of Networking and Virtual Organisations*, 9(2), 106-122 (2011)
5. Blohm, I., Leimeister J.M., Krcmar, H.: Crowdsourcing: How to Benefit from (Too) Many Great Ideas. *MIS Quarterly Executive*, 4(12), 199-211 (2013)
6. Majchrzak, A., Malhotra, A.: Towards an information systems perspective and research agenda on crowdsourcing for innovation. *The Journal of Strategic Information Systems*, 22(4), 257-268 (2013)
7. Blohm, I., Leimeister, J.M., Rieger, M., Krcmar, H.: Controlling von Ideencommunities – Entwicklung und Test einer Ideencommunity-Scorecard. *Controlling*, 2(23), 96-103 (2011)
8. Cohen, W.M., Levinthal, D.A.: Absorptive Capacity: A New Perspective on Learning and Innovation. *Administrative Science Quarterly*, 35(1), 128-152 (1990)
9. Roberts, N., Galluch, P., Dinger, M., Grover, V.: Absorptive Capacity and Information Systems Research: Review, Synthesis, and Directions for Future Research. *MIS Quarterly*, 36(2), 625-648 (2012)
10. Malone, T.W.: Modeling coordination in organizations and markets. *Management science*, 33(10), 1317-1332 (1987)
11. Hippner, H., Rentzmann, R.: Text mining. *Informatik-Spektrum*, 29(4), 287-290 (2006)
12. Yang, J., Adamic, L.A., Ackerman, M.S.: Crowdsourcing and knowledge Sharing: Strategic user behavior on taskcn. *Proceedings of ACM Electronic Commerce'08*, 246-255 (2008)
13. Vaishnavi, V., Kuechler, W.: Design research in information systems. (2004)
14. Schreyögg, G., Schmidt, S.: Absorptive Capacity – Schlüsselpraktiken für die Innovationsfähigkeit von Unternehmen. *WiSt Heft 10*, 474-479 (2010)
15. Lichtenthaler, U., Ernst, H.: Opening up the innovation process: The Role of Technology Aggressiveness. *R&D Management*, 39(1), 38-54 (2009)
16. Zahra, S.A., George, G.: Absorptive capacity: A review, reconceptualization, and extension. *Academy of Management Review*, 27(2), 185-203 (2002)
17. Todorova, G., Durisin, B.: Absorptive capacity: Valuing a reconceptualization. *Academy of Management Review*, 32(3), 774-786 (2007)

18. Lane, P.J., Koka, B.R., Pathak, S.: The Reification of Absorptive Capacity: A Critical Review and Rejuvenation of The Construct. *Academy of Management Review*, 31(4), 833-863 (2006)
19. Ebner, W., Leimeister, J.M., Krcmar, H.: Community Engineering for Innovations: The Ideas Competition as a method to nurture a Virtual Community for Innovations. *R&D Management*, 39(4), 342-356 (2009)
20. Leimeister, J.M.: *Einführung in die Wirtschaftsinformatik*. Springer-Verlag. (2015)
21. Hearst, M.: What is text mining. SIMS, UC Berkeley (2003)
22. Feldman, R., Dagan, I.: Knowledge Discovery in Textual Databases (KDT). *KDD.*, 95, 112-117 (1995)
23. Hippner, H., Rentzmann, R.: Text mining. *Informatik-Spektrum*, 29(4), 287-290 (2006)
24. Sullivan, D.: *Document warehousing and text mining: techniques for improving business operations, marketing, and sales*. John Wiley & Sons, Inc. (2001)
25. Hotho, A., Nürnberger, A., Paaß, G.: A brief survey of text mining. *Ldv Forum*, 20(1), 19-62 (2005)
26. Carstensen, K.U., Ebert, C., Ebert, C., Jekat, S., Langer, H., Klabunde, R.: *Computerlinguistik und Sprachtechnologie: Eine Einführung*. Springer-Verlag (2009)
27. Feldman, R., Sanger, J.: *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press (2007)
28. Williams, S.: *A survey of natural language processing techniques for text data mining* (2000)
29. Miner, G.: *Practical text mining and statistical analysis for non-structured text data applications*. Academic Press (2012)
30. Aggarwal, C.C., Zhai, C.: A survey of text classification algorithms. *Mining text data*. Springer US, 163-222 (2012)
31. Weiss, S.M., Indurkha, N., Zhang, T.: *Fundamentals of predictive text mining*, 41. London: Springer (2010)
32. Konchady, M.: *Text mining application programming*. Charles River Media, Inc. (2006)
33. Tan, S.: Neighbor-weighted k-nearest neighbor for unbalanced text corpus. *Expert Systems with Applications*, 28(4), 667-671 (2005)
34. Hotho, A., Nürnberger, A., Paaß, G.: A brief survey of text mining. *Ldv Forum*, 20(1), 19-62 (2005)
35. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45-77 (2007)
36. Simon, H.A.: *The sciences of the artificial*. MIT press. (1996)
37. March, S.T., Smith, G.F.: Design and natural science research on information technology. *Decision support systems*, 15(4), 251-266 (1995)
38. Leimeister, J.M., Huber, M., Bretschneider, U., & Krcmar, H.: Leveraging crowdsourcing: activation-supporting components for IT-based ideas competition. *Journal of management information systems*, 26(1), 197-224 (2009)
39. Poetz, M.K., & Schreier, M.: The value of crowdsourcing: can users really compete with professionals in generating new product ideas? *Journal of Product Innovation Management*, 29(2), 245-256 (2012)
40. Gregor, S., & Jones, D.: The anatomy of a design theory. *Journal of the Association for Information Systems*, 8(5), 312 (2007)
41. Kruse, P., Schieber, A., Hilbert, A., & Schoop, E.: *Idea Mining–Text Mining Supported Knowledge Management for Innovation Purposes* (2013)