

Entwicklung eines SOA-basierten und anpassbaren Bewertungsdienstes für Inhalte aus sozialen Medien

Christian Reuter, Michael Ritzkatis, Thomas Ludwig

Institut für Wirtschaftsinformatik, Universität Siegen
Hölderlinstr. 3, 57076 Siegen

christian.reuter@uni-siegen.de; michael.ritzkatis@student.uni-siegen.de;
thomas.ludwig@uni-siegen.de

Abstract: Dieser Beitrag soll aufzeigen, wie ein anpassbarer Bewertungsdienst die Nutzung bürgergenerierter Inhalte aus sozialen Medien unterstützen kann. Dabei soll insbesondere geklärt werden, wie dieser gestaltet werden kann und wie Nutzer die Qualitätskriterien angemessen artikulieren können. Nach einer Darstellung von Grundlagen und verwandten Arbeiten wird anhand einer empirischen Vorstudie der Umgang von Behörden und Organisationen mit Sicherheitsaufgaben (BOS) mit bürgergenerierten Informationen betrachtet. Basierend auf den dort gewonnen Erkenntnissen wurde ein service-orientierter Bewertungsdienst entwickelt und in eine Anwendung integriert, welche so den Zugang zu bürgergenerierten Informationen aus verschiedenen sozialen Medien inklusive einer anpassbaren Qualitätsbewertung ermöglicht. Eine abschließende Evaluation illustriert deren mögliche Anwendung in der Praxis.

1 Einleitung

Bürgergenerierte Inhalte spielen in Zeiten weiter Verbreitung interaktiver Technologien und sozialer Medien eine immer größere Rolle. Bürgergenerierte Inhalte sind dabei laut der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung [Orga07] im Internet publizierte und eine gewisse kreative Eigenleistung erfordernde Inhalte, die außerhalb professioneller Routinen und Praktiken erstellt werden. In Schadenslagen, wie zuletzt dem Hurrikan Sandy 2012 oder auch dem mitteleuropäischen Hochwasser 2013, wurden soziale Medien von Betroffenen und freiwilligen Helfern aus der Bevölkerung genutzt, um sich untereinander auszutauschen und zu koordinieren [KaRe14]. Da das Engagement der Bürger weitgehend ungesteuert ist und Inhalte somit nicht zwingend strukturiert entstehen, muss die Vielzahl der entstehenden Daten analysiert werden. Zur Analyse sind geeignete Bewertungsmethoden unabdingbar, wobei eine konsistente Bewertung der Informationsqualität komplex sein kann [FrPK10]. Gerade wenn aus einer Vielzahl von Informationsquellen und -formaten eine Auswahl getroffen werden muss, ist es hilfreich, wenn verschiedene Qualitätskriterien die Bewertung erleichtern können. Dieser Beitrag adressiert die Herausforderungen der Integration bürgergenerierter Inhalte [ReMP12], insbesondere die Bewertung von Informationen aus sozialen Medien. Er

stellt die Ergebnisse einer Design-Fallstudie [Wulf09] dar, die basierend auf einer empirischen Vorstudie [ReRi14] (Kap. 2) und anhand der Entwicklung von IT (Kap. 3 und 4) sowie deren Evaluation (Kap. 5) untersucht, wie unterschiedliche Akteure durch die Verwendung eines anpassbaren Dienstes zur Bewertung bürgergenerierter Informationen aus sozialen Medien unterstützt werden können.

2 Verwandte Arbeiten: Informationsqualität

Es existieren verschiedene Definitionen für Informationsqualität (IQ), deren Gemeinsamkeit ist, dass eine Information dann eine hohe Qualität besitzt, wenn sie den subjektiven Ansprüchen des Adressaten entspricht. Dieses Maß lässt sich anhand verschiedener Kriterien messen. Während [Wolf99] beispielsweise Zeit, Ort, Menge, Form sowie Inhalt, Entscheidungsrelevanz und Wirtschaftlichkeit als Maßstab berücksichtigt, verwenden [ZeHe96] Übertragbarkeit, Gebrauchstauglichkeit, Wartbarkeit, Effizienz, Reliabilität und Funktionalität. Basierend auf 179 Attributen entwickelten [WaSG96] ein Framework mit 15 Dimensionen in vier Kategorien, die für Datennutzer von Relevanz sind.

1. **Intrinsische IQ:** Das Kriterium Präzision gibt an, ob die Information den beschriebenen Inhalt hinreichend genau wiedergibt [FrPK10]. Objektivität drückt aus, ob der Sachverhalt unvoreingenommen von persönlichen Eindrücken und Wertungen widerspiegelt wird; die Glaubwürdigkeit, ob die Informationen aus der Sicht des Empfängers plausibel und verwertbar sind. Dieses Maß entspricht nicht zwingend der objektiven Glaubwürdigkeit [FrPK10], die durch die Reputation ausgedrückt wird.
2. **Kontextabhängige IQ:** Neben den obigen eher allgemeinen Attributen werden solche beschrieben, die eine Information auf ihre Gültigkeit im jeweiligen Kontext überprüfen. Eine Information muss für den aktuellen Kontext einen Mehrwert liefern [ChRe13]. Dieser ist eng an die Aktualität gekoppelt; so ist ein veralteter Plan über Rettungswege nicht geeignet. Zusätzlich sollte sie ein hohes Maß an Relevanz für den Empfänger besitzen. Damit Informationen bewertet werden können, muss eine ausreichende Datenmenge vorhanden sein, die eine Einordnung in einen Kontext überhaupt erst ermöglicht [BZAZ09]. Gleichzeitig sollten Informationen vollständig sein, um die geschilderte Situation umfassend darzustellen [FrPK10].
3. **Begriffliche IQ:** Zusätzlich werden Informationen anhand des Informationsformats und ihrer Bedeutung klassifiziert. Am wichtigsten ist hierbei, dass diese in ihrer dargestellten Form für den Empfänger einfach zu verstehen und darüber hinaus interpretierbar sind. Somit kann der Empfänger die Information eigenständig in den entsprechenden Kontext einordnen. Für das Informationsformat gelten die Konsistenz des Formats und prägnante Darstellung, durch welche eine Eindeutigkeit von Abkürzungen und Symbolen sowie der Information selbst gewährleistet wird [WaSG96].
4. **Zugangs-IQ:** Die vierte Kategorie umfasst Eigenschaften der Sicherheit. Auf Informationen muss zugegriffen werden können, um sie verwenden zu können. Der Zugriff sollte aber nur unter Einhaltung der Sicherheitsanforderungen geschehen, so dass eine Person nur berechtigte Informationen benutzen kann [WaSG96].

Wie dargestellt, lässt sich Informationsqualität aus verschiedenen Perspektiven bestimmen. Im Folgenden wird sie als hoch betrachtet, wenn die Informationen den subjektiven

Ansprüchen des Empfängers genügen. Daraus motiviert sich, dass der Empfänger der Information selbst bestimmen muss, welche Dimensionen für ihn relevant sind.

Im Laufe der letzten Jahre beschäftigten sich verschiedene Vorhaben mit der Umsetzung obiger und weiterer Kriterien zur Bewertung und Integration bürgergenerierter Informationen in Schadenslagen. Eine Herangehensweise ist die Beschränkung der Informationsmenge. Dazu existieren Verfahren, die nicht auf der freien Informationseingabe der Benutzer basieren, sondern den Benutzern entweder eine spezielle Eingabesyntax vorschreiben [StPa11] oder aber konkrete Informationsanfragen an einzelne Benutzer stellen [DARO13]. Durch diese Einschränkung besteht allerdings die Gefahr, dass möglicherweise relevante Informationen verloren gehen. Andere Ansätze betrachten ebenfalls nur Teilmengen der Informationsbasis: [TVSP12] berücksichtigen lediglich die Daten eines sozialen Netzwerks. [ViBo12] setzen auf eine breitere Informationsbasis, ermöglichen jedoch nur BOS Zugang zu ihren Informationen. Gängige Verfahren zur Klassifizierung von Informationen reichen derzeit von maschinellem Lernen auf Basis von Wörterbüchern oder Sprachklassifizierungen [CACI13] über Text Mining [ZMTW13] und Bestimmung des Gefühlsausdrucks der Information durch den normalisierten „Sentiment Value“, sowie Einteilung der Nachrichten in verschiedene Klassen mit Bayesschen Netzen [NaSt12] und Bestimmung von Ähnlichkeiten anhand regulärer Ausdrücke [ZiBü12] bis hin zu Bestimmung des Informationsgehalts der Nachricht durch Entropie [VVCP11]. Einige Arbeiten beschränken sich auf die Ermittlung von Metainformationen, mit denen eine bessere Bewertung der Informationslieferanten möglich sein kann [StMP12]. Dennoch weist zum aktuellen Zeitpunkt jede Bewertungsmethode Fehlklassifizierungen auf. Weiterhin existieren zur Selektion und Nutzung bürgergenerierter Informationen einige Ansätze, die allerdings keine umfassende Qualitätsbewertung unterstützen [ReRi14].

3 Vorstudie: Nutzung bürgergenerierter Inhalte

Um ein tiefgehendes Verständnis über die Rolle bürgergenerierter Informationen in Schadenslagen zu bekommen, wurde das Datenmaterial einer empirischen Studie über Informations- und Kommunikationspraktiken von BOS (hier: Feuerwehr, Polizei) im Falle einer Großschadenslage in zwei geographisch und strukturell verschiedenen Kreisen Nordrhein-Westfalens verwendet. Die Ergebnisse der Vorstudie werden in [ReRi14] ausführlich behandelt und für diesen Beitrag nur zusammengefasst dargestellt.

Es wurden verschiedene qualitative Methoden wie Dokumentenanalysen, 4 Beobachtungen, 26 teilstrukturierte Interviews und 4 Gruppendiskussionen durchgeführt und ausgewertet. Ziel der Dokumentenanalyse war die Untersuchung offizieller Arbeitspraktiken mit Hilfe von Gesetzen und Richtlinien. Die Beobachtungen dienten dazu, Erkenntnisse über die tatsächliche praktische Arbeit zu erhalten. Diese wurden in der Leitstelle während eines normalen Arbeitstages (9h), im Krisenstab und der Einsatzleitung einer Krisenübung (4h) und eines Großereignisses (6h) durchgeführt. Die semi-strukturierten Interviews (n=22 je 1–2h; bezeichnet als I01-22) dienten der Reflexion der Arbeitspraxis, der Identifikation möglicher IT-Unterstützung und interorganisationaler Informationsbedarfe und wurden mit Einsatzkräften verschiedener Ebenen durchgeführt. Sie fo-

kussierten im ersten Teil die Arbeitstätigkeiten in normalen Lagen, im zweiten Teil potentielle Aufgaben im vorgelegten Szenario und im dritten Teil Informations- und Kommunikationstechnologien. Ergänzende Interviews (n=5 je 1h; bezeichnet als IM01-05) fokussierten mobile Kollaborationspraktiken. Die Gruppendiskussionen (n=4 je 4h, bezeichnet als W1-4) dienten der Erfassung kommunikativer Arbeitspraxen. Bei der Auswertung des vorliegenden Datenmaterials wurde sich an der Grounded Theory [Stra04] orientiert. Dazu wurden die Transskripte offen kodiert und die Aussagen in Textbausteine und spätere Kategorien aufgeteilt und Herausforderungen hergeleitet:

Plattformübergreifende Nutzung sozialer Medien: In Schadenslagen müssen sich Einsatzkräfte einen Überblick über die Situation verschaffen, um Tendenzen sowie Entwicklungen der Lage zu erkennen (I07, I18). Der Versuch, genügend Informationen vorzuhalten, scheitert dabei oftmals aufgrund der Ungewissheit der Situation (I01). In einigen Bereichen, wie der Beobachtung von Großevents oder Demonstrationen, werden externe Quellen mit dem Ziel der Erweiterung der Informationsbasis genutzt (W3, I02, I03). Hierzu gehören u.a. soziale Medien: „Diese elektronischen Medien, [...] werden immer mehr Einfluss gewinnen, auch auf die Art der Einsatzbearbeitung [...] und auf die Art der Informationserhebung“ (I02).

Automatisierte Verarbeitung: Gerade in nicht planbaren Einsätzen entstehen schnell Zeit- und Personalengpässe, welche die Notwendigkeit computerunterstützter Verarbeitung deutlich machen: „Wenn wir oben nur mit drei oder vier Mann sitzen und führen einen Einsatz, dann habe ich keine Zeit, mich noch mit Fremden zu beschäftigen“ (I2). Eine Vielzahl der Informationen soll idealerweise automatisiert und nicht manuell bearbeitet werden, um eine Informationsüberflutung der Einsatzkräfte, die diese paralyisiert und weitere Handlungen behindert, zu vermeiden (I02, I03). Denn „*die Gefahr besteht*, mit Informationen überfrachtet zu werden und dass man dann letztendlich gar nichts versteht, weil man so viel Input bekommt“ (I03).

Flexible und anpassbare Qualitätsbewertung: Insbesondere externe Quellen sind für Fehlinformationen anfällig (I14, I15) und müssen deshalb überprüft werden (I15): „Da zeigt sich natürlich immer wieder, dass wir da aufs Glatteis geführt werden sollen [...] und wir müssen die Informationen bewerten für uns“ (I02). Trotzdem scheitert der Einsatz von Bürgerinformationen aus sozialen Netzwerken besonders an der Bewertung der Informationen durch Einsatzkräfte, denn „*da ist einfach irgend so ein Flaschenhals, über den wir nicht drüber kommen*“ (I02). Im Allgemeinen stellt die Bewertung von Informationen eine große Herausforderung dar, da sowohl die Herkunft der Informationen als auch die zur Bewertung notwendigen Metainformationen meist unbekannt sind (W3, W4). Diese Überprüfung benötigt Ressourcen, damit die Einsatzkraft nicht einer Informationsflut ausgesetzt wird: „Dann müssen sie noch 3-4 Filter-Funktionen davor schalten, damit Sie nur die wichtigen Informationen ankommen lassen“ (I01). Insbesondere muss sichergestellt sein, dass die gefilterten Informationen einen Mehrwert für die Einsatzkräfte darstellen (I02, I03). Erschwerend kommt hinzu, dass nicht im Vorhinein bestimmt werden kann, welche Hierarchieebene welche Informationen benötigt (IM01). Daher muss eine Flexibilität der Bewertungskriterien gegeben sein (I19). Letztlich muss die Einsatzkraft die Verantwortung über die Handlungen übernehmen und entscheiden, ob eine Information verwendet wird oder nicht (I15).

4 Konzept: Bewertung von Inhalten aus soziale Medien

Verwandte Arbeiten sowie die empirische Studie haben gezeigt, dass die Bewertung der Masseninformationen und die Extraktion relevanter Informationen eine große Herausforderung darstellen. Es scheint nachvollziehbar, dass verschiedene Situationen verschiedene Bewertungsmethoden erfordern und somit die Möglichkeit der Kombination derselben zur Verbesserung der Qualitätsbewertung beitragen könnten. Verschiedene Bewertungsmethoden wurden bereits in Kapitel 2 aufgezeigt. In Kapitel 3 wurde deutlich, dass unterschiedliche Bewertungsmethoden die subjektive Qualitätsbewertung in verschiedenen Situationen unterstützen können. Der Bewertungsdienst für soziale Medien ermöglicht eine Bewertung der Nachricht nach 15 Bewertungsmethoden (Tabelle 1), die sich an den oben dargestellten Kriterien orientieren und sich hinsichtlich der technischen Umsetzung in vier Kategorien unterteilen lassen:

- A Die **Bewertung der Metadaten** beinhaltet fünf Bewertungsmethoden (Autorenhäufigkeit, zeitliche Nähe, örtliche Nähe, Anzahl Follower/Likes, Anzahl Metadaten), bei denen entweder die Abweichung zu den eingegebenen Suchkriterien oder aber das absolute Auftreten durch die Ermittlung der Differenz bestimmt wird.
- B Die **Bewertung auf Basis des Inhalts** stellt zwei Bewertungsmethoden (Suchbegriffhäufigkeit, Stopwords) zur Verfügung, die das Vorkommen bestimmter Wörter anhand einer Wortliste bzw. einer Synonymliste ermitteln.
- C Die **Bewertung auf Basis der Klassifikation der Nachricht** nutzt sechs Bewertungsmethoden (Sentiment Analyse, Fear Factor, Happiness Factor, Named Entity Recognition, Emoticon, Slang), die mittels Wortlisten das Auftreten von Wörtern untersuchen und die Information so in verschiedene Klassen einteilen.
- D Die **Bewertung auf Basis wissenschaftlicher Methoden** verwendet zwei Bewertungsmethoden (Shannonscher Informationsgehalt, Term Frequency – Inverse Document Frequency).

Von diesen Methoden kann der Benutzer wählen, welche er in welcher Gewichtung verwenden möchte. Somit kann eine subjektive Informationsqualität bestimmt werden. Darüber hinaus ermöglichen diese Auswahl und die verschiedenen Kategorien einen Einsatz des Bewertungsdienstes in einer Vielzahl unterschiedlicher Szenarien. Generell gilt: Zuerst werden die einzelnen Nachrichten hinsichtlich der jeweiligen Methode absolut bewertet. Anschließend wird durch Suche des größten und kleinsten absoluten Wertes der relative Wert einer jeden Nachricht bestimmt. Die Nachricht mit dem größten absoluten Wert erhält den Score 1.0 (100%), diejenige mit dem kleinsten den Score 0.0 (0%). Im Anschluss werden Einzelscores gewichtet und ein Gesamt-Score gebildet.

Nr.	Methode / Kriterium	Beschreibung
A	Bewertung der Metadaten	
1	Autorenhäufigkeit (Reputation)	Anzahl von Nachrichten desselben Autors im Nachrichtenset. Je mehr Nachrichten ein Autor verfasst, desto mehr Wissen über die Situation wird unterstellt.
2	Zeitliche Nähe (Aktualität)	Zeitliche Nähe der Nachrichten mit dem Mittelpunkt des Suchzeitraums. Je näher die Nachricht am gesuchten Zeitpunkt ist, desto sicherer kann

		man sein, dass es sich um eine neue Information handelt.
3	Örtliche Nähe	Entfernung zwischen dem Erstellungsort einer Nachricht zu dem übergebenen Ort. Je geringer die Entfernung ist, desto höher ist die Wahrscheinlichkeit, dass die Nachricht die aktuelle Schadenslage thematisiert.
4	Follower/Likes (Glaubwürdigk.)	Mit steigender Anzahl an Likes bzw. Followern einer Nachricht / eines Autors wird ein steigendes Maß an Glaubwürdigkeit unterstellt.
5	Metadaten (Bilder/Links) (Ausreichende Datenmenge)	Gerade wenn eine bessere Betrachtung der Schadenslage fokussiert werden soll, ist es neben textuellen Informationen hilfreich, ein Bild oder andere mediale Informationen zu verwenden. Daher misst dieses Bewertungskriterium die Anzahl der Metadaten.
B	Bewertung auf Basis des Inhalts	
6	Suchbegriffhäufigkeit (Interpretierbarkeit)	Es wird sichergestellt, dass die Nachricht nicht nur zufällig den Suchbegriff enthält, sondern auch tatsächlich den Sachverhalt thematisiert. Dazu wird die Nachricht auch nach Synonymen durchsucht.
7	Stopwords	Die Anzahl an sogenannten Stopwords (bspw: „also“) erhöht nicht den Aussagegehalt der Nachricht, da diese Wörter keine Information liefern. Deshalb steigt mit sinkender Anzahl an Stopwords in der Nachricht ihr Nutzen.
C	Bewertung auf Basis der Klassifikation der Nachricht	
8	Sentiment Analyse (Objektivität)	Die Nachricht wird hinsichtlich ihrer emotionalen Eigenschaft bewertet. Emotionale Nachrichten können den Sinn verfremden, insbesondere wenn sie durch Angst motiviert sind.
9	Fear Factor	Bemisst den Grad des Ausdrucks von Angst in der Nachricht durch die Häufigkeit von Wörtern, die mit der Thematik Angst verbunden sind.
10	Happiness Factor	Bemisst hingegen den Ausdrucksgrad von Freude in der Nachricht durch die Häufigkeit von Wörtern, die mit der Thematik Freude verbunden sind.
11	Named Entity Recognition (NER)	Anzahl an Entitäten in der Nachricht. Eine Entität deutet dabei auf die Verknüpfung des Inhalts der Informationen zu einer anderen Informationsquelle hin. Daher steigt die Qualität der Information mit der Anzahl der Entitäten einer Nachricht.
12	Emoticon	Anzahl an Emoticons, wie Smileys, die ironisch formulierte Nachrichten kennzeichnen. Zum Verständnis werden die Emoticons durch ihre Bedeutung in der Nachricht ersetzt. Daraus motiviert sich, dass Informationen mit wenig Emoticons einen eindeutigeren Inhalt haben können.
13	Slang	Anzahl an Abkürzungen, dessen Einsatz durch Zeichenbeschränkungen ein beliebtes Instrument ist, um mehr Inhalt zu vermitteln. Wie Emoticons erschweren Abkürzungen eine automatisierte Verarbeitung.
D	Bewertung auf Basis wissenschaftlicher Methoden	
14	tf-Idf (term frequency – Inverse document frequency)	Das Auftreten der einzelnen Suchbegriffe (term frequency) mit der Auftrittshäufigkeit in der gesamten Nachrichtenanzahl (Inverse document frequency). Die Methode ist bei der Verwendung von mehreren Suchbegriffen hilfreich, da das Auftreten eines Teils der Suchphrase, welches in wenigen Dokumenten häufig auftritt, höher gewichtet wird, als wenn es in vielen Dokumenten absolut nicht so häufig vorkommt.
15	Informationsgehalt	Shannonscher Informationsbegriff. Je größer der Informationsgehalt einer Nachricht ist, desto mehr Informationen kann sie transportieren.

Tabelle 1: Bewertungsmethoden

5 Implementierung eines SOA-basierten Bewertungsdienstes

Der eigentliche Bewertungsdienst ist nach dem webbasierten service-orientierten Architekturmodell (SOA) realisiert. Durch dieses Architekturmodell ist es möglich, die Bewertung zentral zur Verfügung zu stellen und durch die Bereitstellung der Ergebnisse inklusive der Originaldaten im Format JSON (JavaScript Object Notation) eine Einbindung in verschiedene Anwendungen zu ermöglichen. Der Aufruf der Schnittstelle funktioniert per HTTP-GET (RESTful-Webservice). Der URL werden die Parameter (durch "&" getrennt) angehängen. Die serverseitige Bewertung der Informationen soll die Rechenlast des Clients verringern.

Für die Extraktion der Daten aus den sozialen Netzwerken werden APIs der jeweiligen Networkbetreiber verwendet [ReSc14]. Insbesondere werden im Rahmen dieses Beitrags die sozialen Netzwerke Twitter und Facebook, begründet durch die Existenz einer notwendigen API, betrachtet. Die jeweiligen Betreiber-APIs bieten eine Fülle von Möglichkeiten, Daten sowohl in das soziale Netzwerk zu importieren als auch zu exportieren. Für die Erkennung von Begriffen aus dem Inhalt einer Nachricht wird ein NER (Nr. 11) verwendet. Der Stanford NER (<http://nlp.stanford.edu/software/CRF-NER.shtml>) steht kostenlos als Java-Bibliothek zur Verfügung. Für den NER wurde der Korpus „deWac generalized classifier“ eingesetzt, weil dieser vor allem auf deutsche Nachrichten aus sozialen Netzwerken trainiert wurde. Für die Erstellung eines Bayes Classifiers (Nr. 8), mit dem die Informationen in verschiedene Klassen eingeteilt werden können, wurde die Bibliothek Classifier4J (<http://classifier4j.sourceforge.net/>) verwendet, da dieser mit Wortlisten trainiert werden kann. Die Synonymliste (Nr. 6) wird mittels des Open Thesaurus Webservices (<http://www.openthesaurus.de/>) erzeugt.

Zur Visualisierung der Informationen wird ein geographischer Bezug benötigt. Da ein Großteil der Informationen keine geographischen Metadaten beinhalten, müssen diese geocodiert werden. Der Gisgraphy Geocoder (<http://www.gisgraphy.com/>) lässt sich über einen REST (Representational State Transfer) Webservice verwenden und geokodiert die Ortsinformationen für beliebiges Kartenmaterial. Aus Geschwindigkeitsgründen wird jedoch für jeden bereits geokodierten Ort eine Liste angelegt, aus der die Koordinaten auch ohne Geokodierung ermittelt werden können. GSON (<https://code.google.com/p/google-gson/>) bietet die automatische Erstellung eines JSON Objektes anhand eines Java-Objektmodells und wird deshalb zur Konvertierung verwendet.

Der Bewertungs-Service wurde in eine Facebook-Anwendung, die ungebundene Helfer in ihrer Arbeit der Schadensbewältigung durch die medienübergreifende Informationssuche und die -generierung unterstützt, integriert [KaRe14]. Innerhalb dieser Anwendung besteht die Möglichkeit, Informationen nach verschiedenen Parametern zu suchen und wahlweise zusätzlich eine Qualitätsbewertung durchzuführen (Abbildung 1). Dabei bestimmt der Benutzer welche Bewertungskriterien er verwenden möchte. Durch diese Art der Benutzerintegration werden die Anforderungen der Vorstudie nach einer flexiblen und kontrollierbaren Qualitätsbewertung realisiert. Dazu werden bereitgestellte Schnittstellen des Bewertungsdienstes verwendet. Die Methoden Nr. 8, 11 und 13 wurden zur Reduzierung der Bewertungskriterien hier nicht integriert. Die Ergebnisse dieser

Informationssuche und Bewertung werden sowohl tabellarisch mit verschiedenen Metainformationen, wie dem Erfüllungsgrad einzelner Methoden, die als MouseOver-Text angezeigt werden, als auch visuell auf einer Lagekarte zur direkten Bestimmung der Nähe der Information zu dem Suchstandort, dargestellt (Abbildung 2). Somit kann der Benutzer aus verschiedenen Verfahren der Ergebnisbetrachtung wählen und erhöht damit die Flexibilität der Anwendung. Dabei stellt diese Benutzeroberfläche nur eine von vielen Varianten zur Benutzung der Qualitätsbewertung dar.

Kompakt betrachtet vereint der Bewertungsdienst folgende Funktionalitäten: Die bürgergenerierten Informationen können anhand verschiedener Methoden bewertet werden. Diese Bewertung findet nicht nur auf Basis von Metadaten, sondern vor allem auf Basis des Inhalts statt. Der Benutzer bestimmt dabei jederzeit, welche Methoden er wie gewichten möchte. Die subjektive Qualität einer Nachricht bestimmt sich aus Kombination der Bewertung aller Methoden. Der Bewertungsdienst weist ein hohes Maß an Flexibilität durch die Möglichkeit der einfachen Erweiterung der Quellen und Bewertungsmethoden auf. Durch die SOA-basierte Implementierung wird die Einbindung und Verwendung dieser Informationen in weitere Applikationen ermöglicht.

SUCHEINSTELLUNGEN

Allgemein

Suchbegriff:

Netzwerke festlegen: Facebook Twitter Evaluation

Zeitraum angeben: -

Standort auswählen:

Umkreis der Suche (km):

Bewertung der Metadaten ↑

Mich interessieren Nachrichten, die...

- von einem Autor sind, der viele Nachrichten verfasst hat
- in zeitlicher Nähe zu meinem Zeitraum sind
- in der Nähe meines Ortes erstellt wurden
- andere User als hilfreich befunden haben
- einen Link oder ein Bild enthalten

nicht erwünscht egal erwünscht

Bewertung auf Basis des Inhalts ↑

Mich interessieren Nachrichten, die...

- mein Suchwort und dessen Synonyme am Häufigsten enthalten
- ohne Füllwörter sind

nicht erwünscht egal erwünscht

Bewertung auf Basis der Klassifikation der Nachricht ↑

Mich interessieren Nachrichten, die...

- Emotionen ausdrücken
- Angst ausdrücken
- Freude ausdrücken

nicht erwünscht egal erwünscht

Bewertung auf Basis wissenschaftlicher Methoden ↑

Mich interessieren Nachrichten, die...

- einen hohen Informationsgehalt besitzen
- (nur bei mehreren Suchbegriffen sinnvoll) einen Teil des Suchbegriffs im Verhältnis zu den anderen Teilen am Häufigsten enthalten

nicht erwünscht egal erwünscht

Abbildung 1: Informationssuche mithilfe von Schiebereglern

6 Evaluation: Anpassbare Bewertungsmechanismen

Um die Fragestellung zu beantworten, wie anpassbare Bewertungsmechanismen Benutzern geeignet zur Verfügung stellen können und wie Nutzer die Qualitätskriterien angemessen artikulieren können, wurde der Bewertungsdienst durch potenzielle Endnutzer evaluiert. Die Durchführung fand dabei anhand eines szenarienbasierten Walkthroughs mit anschließenden teilstrukturierten Interviews statt, um möglichst viele Erkenntnisse über das Potenzial des Dienstes und der Qualitätsbewertung bürgergenerierter Informationen zu erhalten. Die Teilnehmer sollten dabei gemäß Thinking Aloud [Niel93] ihre Gedanken während der Durchführung laut äußern, damit zugrundeliegende Gedankengänge und subjektive Eindrücke aufgenommen werden konnten. Die Evaluation dauerte jeweils ca. 45 Minuten und wurde mit insgesamt 19 Personen durchgeführt, wovon sechs Personen eine erste Version der Applikation bewertet haben. Sofern ein Teilnehmer kein ungebundener Helfer war, der sehr aktiv soziale Medien während des Hochwassers 2013 genutzt hat [KaRe14], wurde ihm zusätzlich eine Rollenbeschreibung vorgestellt, sodass er sich in die Lage eines Helfers versetzen konnte. Die Evaluation ergab, dass sich der Bewertungsdienst in vielen Bereichen als hilfreich erwies, jedoch noch weiterer Überarbeitungsbedarf besteht.

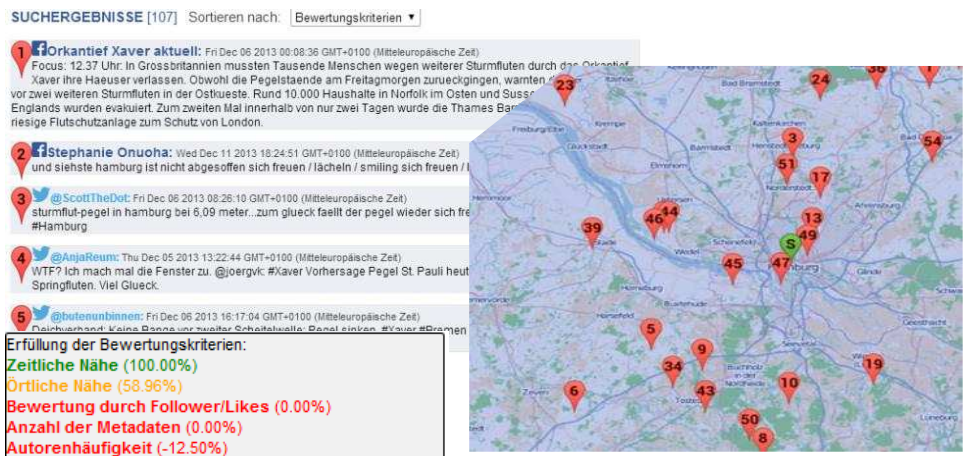


Abbildung 2: Suchergebnisse (links), Erfüllungsgrad (unten links) und Kartendarstellung (rechts)

Auswahl der Qualitätskriterien: Die Anzahl der Bewertungskriterien wirkte auf viele Nutzer im ersten Moment als zu hoch (E09, E07, E19). Dabei waren sich die Befragten darin einig, dass verschiedene Situationen verschiedene Bewertungskriterien erfordern (E12, E13, E08) und daher immer eine gewisse Anpassung der Kriterien an die Situation notwendig sei: „Also es sind schon einige Kriterien, aber ich denke, dass ist auch wichtig, weil eben verschiedene Fragen verschiedene Suchbegriffe erfordern in dem Sinne“ (E13). Daher wurde auch der Vorschlag geäußert, dass die Bewertungskriterien „ein bisschen schon so eingestellt werden können, dass man vielleicht schneller zum Ziel kommt“ (E12). Darüber hinaus wurden auch weitere Möglichkeiten, wie zum Beispiel die Schaffung einer Suche nach einem Unterbegriff (E19), als Ergänzung der Einstel-

lungsmöglichkeiten gewünscht. Die Evaluation hat gezeigt, dass die größte Herausforderung in der Identifikation passender Qualitätskriterien liegt. Während beispielsweise Aktualität von allen Testpersonen als wichtiges Kriterium verstanden wurde, konnten sich nur wenige Probanden unter der Koordinatenqualität etwas vorstellen. Jedoch konnte die Erklärung der Koordinatenqualität als Maß für die örtliche Nähe bereits Abhilfe schaffen. Ein Proband stellte zudem den Zusammenhang zwischen der Anzahl von Abonnenten eines Autors und seiner Reputation in Frage (E4).

Breite Informationsbasis und Informationsflut: Insbesondere die Möglichkeit, verschiedene Informationsquellen gleichzeitig mit einer Suche zu betrachten, wurde als Mehrwert dargestellt (E19, E15, E18): „Da wahnsinnig viel die öffentlichen Netzwerke genutzt werden, ist das also noch viel besser, die miteinander zu verknüpfen. Also es kann ja wahnsinnig gut helfen, den *Informationsbedarf zu decken*“ (E17). Die Anzahl der Quellen sollte fortlaufend um weitere sinnvolle Quellen ergänzt werden, wobei nicht nur soziale Netzwerke, sondern beispielsweise auch Nachrichtenseiten betrachtet werden sollten. Des Weiteren zeigten die Befragten bei einer Informationssuche in den sozialen Netzwerken Angst vor einer Informationsflut (E16, E19). Diese konnte aber durch die Sortierung der Ergebnisse im Bewertungsdienst schnell beseitigt werden. Die meisten Benutzer wollten nicht die ganze Liste der Suchergebnisse durchgehen, sondern nur die ersten Einträge betrachten (E10, E11). Trotz allem sollten auch die weiteren Suchergebnisse für den Benutzer einsehbar bleiben, denn einige Szenarien erfordern ggf. die Sichtung weiterer Ergebnisse.

Automatische und manuelle Qualitätsbewertung notwendig: Die Qualitätsbewertung der Informationen findet automatisch statt. Dieser Automatismus wird durch die Benutzer dadurch akzeptiert, dass sie eine Kontrollmöglichkeit der Bewertung haben und zugleich nachvollziehen können, warum was wie bewertet wurde (E08; E11). Denn „*wie immer, wenn Sachen bewertet werden, ne Bewertung von anderen ersetzt nicht die eigene Meinung*“ (E10). Dabei wurde die generelle Möglichkeit der Kombination der Kriterien als Vorteil dargestellt: „Die Standardeinstellung ist egal. Das heißt, wenn ich nichts *mache, verändert das mein Suchergebnis nicht*“ (E13). „Dadurch ist natürlich eine *vielfältige Kombination möglich, die ich auch für hilfreich erachte*“ (E07). Um negativen Auswirkungen auf Handlungen entgegenzuwirken, sollte eine manuelle Nachbewertung umgesetzt werden, die die Korrektur bzw. Identifizierung fehlerhafter Einträge ermöglicht.

Verbesserung des Informationsflusses: Mit den aufgezeigten Möglichkeiten und Verbesserungsvorschlägen besteht das Potenzial, den Informationsfluss in Schadenslagen zu verbessern (E13, E19, E14). Dieses könnte insbesondere durch eine Klassifizierung der Schadenslagen und darauf aufbauend durch eine Voreinstellung der Gewichtungen realisiert werden. Wesentliche zeitliche und organisatorische Engpässe können vermieden werden, indem Informationen von Personen vor Ort (realer Helfer) oder im Netz (virtueller Helfer) [HeRe13] eingeholt werden (E07, E16): „Die Vorteile sind tatsächlich, dass ich Sachen schnell und zügig finden kann, [...] da man auch gezielt nach etwas suchen kann und es wirklich in den verschiedensten Plattformen ausgespuckt wird, so wie ich das möchte. Und ich kann anhand der Bewertungskriterien eben sehr gut *gewichten*“ (E18; 00:19:23-5).

7 Fazit

Schadenslagen stellen besondere Herausforderungen, wie situative Ungewissheit oder zeitliche sowie organisatorische Engpässe für Einsatzkräfte und ungebundene Helfer, dar. Um diese Herausforderungen bewältigen zu können, werden möglichst viele relevante Informationen, auch aus sozialen Medien, benötigt.

Dieser Beitrag hat gezeigt, wie die Qualitätsbewertung bürgergenerierter Informationen aus sozialen Medien unterstützt werden kann. Dabei liefert die Designfallstudie drei Beiträge zum Forschungsbereich: (1) eine Analyse des Umgangs mit bürgergenerierten Inhalten in Schadenslagen durch eine empirische Studie [ReRi14] (Kap. 3), (2) einen funktionsfähigen SOA-basierten und anpassbaren Bewertungsdienstes, der flexibel in verschiedene Anwendungen integriert werden kann (Kap. 4 und 5), sowie (3) die Ergebnisse dessen qualitativer Evaluation innerhalb einer Webanwendung (Kap. 6). In diesem Beitrag sollte insbesondere untersucht werden wie Nutzer die Qualitätskriterien angemessen artikulieren können; unser Ansatz hierzu wurde mithilfe der Evaluation überprüft. Zusammenfassend lässt sich sagen, dass die Flexibilität in Form der Anpassungsmöglichkeit von Sortierung, Quellenauswahl, Qualitätsbewertung hilfreich ist, da die Informationsbewertung immer subjektiv ist und somit auch diese Faktoren vom persönlichen Empfinden, der Erfahrung und der Lage abhängen.

Der durch diese Arbeit entwickelte Bewertungsdienst weist verschiedene Vorteile auf: Im Gegensatz zu [ViBo12] können so nicht nur Einsatzkräfte diese Informationen erhalten, sondern auch ungebundene Helfer können in ihrer Arbeit durch eine flexible Bewertung und freie Zugänglichkeit unterstützt werden. Eine sehr breite Informationsbasis wird durch die Verwendung verschiedener sozialer Netzwerke als Informationsquellen realisiert [TVSP12]. Im Vergleich zu [StPa11], die durch die Benutzung einer festen Syntax die Informationsmenge eingrenzen, werden hier aus den verschiedenen Quellen sämtliche Informationen betrachtet. Schließlich ermöglicht im Vergleich zu [VVCP11] erst die Verwendung von mehreren Bewertungskriterien, die miteinander kombiniert werden können, eine adäquate Qualitätsbewertung bürgergenerierter Informationen.

Gefördert im Rahmen des EU-FP7-Projekts „EmerGent“ (Fördernummer 608352).

Literaturverzeichnis

- [BZAZ09] Bharosa, N., Zanten, B. Van, Appelman, J., & Zuurmond, A.: Identifying and confirming information and system quality requirements for multi-agency disaster management. In: Proc. ISCRAM. Gothenburg, Sweden, 2009
- [CACI13] Chowdhury, S., Amer-Yahia, S., Castillo, C., Imran, M., & Asghar, M. R.: Tweet4act: Using Incident-Specific Profiles for Classifying Crisis-Related Messages. In: Proc. ISCRAM. Baden-Baden, Germany, 2013, pp. 834–839
- [ChRe13] Christofzik, D. & Reuter, C. The Aggregation of Information Qualities in Collaborative Software. In: IJEV, vol. 5 (2013), Nr. 3, pp. 257–271
- [DARO13] Díaz, P., Aedo, I., Romano, M., & Onorati, T. A framework to integrate large-scale participation in Disaster and Emergency Management. In: Workshop on Large-Scale Ideation and Deliberation Systems. Proc. C&T. München, 2013

- [FrPK10] Friberg, T., Prödel, S., & Koch, R.: Analysis of information quality criteria in crisis situation as a characteristic of complex situations. In: Proc. IQ. Little Rock, 2010
- [HeRe13] Heger, O. & Reuter, C.: IT-basierte Unterstützung virtueller und realer Selbsthilfegemeinschaften in Katastrophenlagen. In: Proc. WI. Leipzig, 2013, pp. 1861–1875
- [KaRe14] Kaufhold, M.-A., & Reuter, C.: Vernetzte Selbsthilfe in Sozialen Medien beim Hochwasser 2013 in Deutschland. In: i-com vol. 13 (2014), Nr. 1, pp. 20–28
- [NaSt12] Nagy, A. & Stamberger, J.: Crowd Sentiment Detection during Disasters and Crises. In: Proc. ISCRAM. Vancouver, Canada, Canada, 2012, pp. 1–9
- [Niel93] Nielsen, J.: Usability Engineering. San Francisco: Morgan Kaufmann, 1993
- [Orga07] Organisation for Economic Co-operation and Development: Participative Web: User-Created Content. URL <http://www.oecd.org/internet/ieconomy/38393115.pdf>
- [ReMP12] Reuter, C.; Marx, A. & Pipek, V.: Crisis Management 2.0: Towards a Systematization of Social Software Use in Crisis Situations. In: IJISCRAM vol. 4 (2012), Nr. 1, pp. 1–16
- [ReRi14] Reuter, C. & Ritzkatis, M.: Adaptierbare Bewertung bürgergenerierter Inhalte aus sozialen Medien. In: Mensch & Computer 2014. München: Oldenbourg, 2014
- [ReSc14] Reuter, C. & Scholl, S.: Technical Limitations for Designing Applications for Social Media. In: Workshop-Proceedings Mensch & Computer 2014. München: Oldenbourg, 2014.
- [StMP12] Starbird, K., Muzny, G., & Palen, L.: Learning from the Crowd: Collaborative Filtering Techniques for Identifying On-the-Ground Twitterers during Mass Disruptions. In: Proc. ISCRAM. Vancouver, Canada, 2012
- [StPa11] Starbird, K., & Palen, L.: Voluntweeters: Self-Organizing by Digital Volunteers in Times of Crisis. In: Proc. CHI. Vancouver, Canada: ACM-Press, 2011
- [Stra04] Strauss, A. L.: Methodologische Grundlagen der Grounded Theory. In: Strübing, J.; Schettler, B. (eds.): Methodologie interpretativer Sozialforschung: Klassische Grundlagentexte. Konstanz, Germany: UVK, 2004, pp. 427–453
- [TVSP12] Terpstra, T., Vries, A. de, Stronkman, R., & Paradies, G. L.: Towards a realtime Twitter analysis during crises for operational crisis management. In: Proc. ISCRAM. Vancouver, 2012
- [ViBo12] Vivacqua, A. S., & Borges, M. R. S.: Taking advantage of collective knowledge in emergency response systems. In: Journal of Network and Computer Applications vol. 35, Elsevier (2012), Nr. 1, pp. 189–198
- [VVCPI1] Verma, S., Vieweg, S., Corvey, W., & Palen, L.: Natural Language Processing to the Rescue? Extracting “Situational Awareness” Tweets During Mass Emergency. In: Proc. ICWSM. Barcelona, 2011, pp. 385–392
- [WaSG96] Wang, R., Strong, D., & Guarascio, L.: Beyond accuracy: What data quality means to data consumers. In: Journal of Management Information Systems (1996)
- [Wolf99] Wolf, P.: Konzept eines TQM-basierten Regelkreismodells für ein “Information Quality Management” (IQM). Verlag Praxiswissen. Dortmund, 1999
- [Wulf09] Wulf, V.: Theorien sozialer Praktiken zur Fundierung der Wirtschaftsinformatik: Eine forschungsprogrammatische Perspektive. In: Becker, J.; Krcmar, H.; Niehaves, B. (eds.): Wissenschaftstheorie und Gestaltungsorientierte Wirtschaftsinformatik. Berlin: Springer/Physika, 2009, pp. 211–224
- [ZeHe96] Zeist, R. H. J., & Hendriks, P. R. H. Specifying software quality with the extended ISO model. In: Software Quality Journal vol. 5 (1996), Nr. 4, pp. 273–284
- [ZiBü12] Zielinski, A., & Bügel, U.: Multilingual Analysis of Twitter News in Support of Mass Emergency Events. In: Proc. ISCRAM. Vancouver, Canada, 2012, pp. 1–5
- [ZMTW13] Zielinski, A., Middleton, S., Tokarchuk, L., & Wang, X.: Social Media Text Mining and Network Analysis for Decision Support in Natural Crisis Management. In: Proc. ISCRAM. Baden-Baden, Germany, 2013, pp. 840–845