Microblogging during the European Floods 2013:

What Twitter May Contribute in German Emergencies

Christian Reuter, Institute for Information Systems, University of Siegen, Siegen, Germany

Julian Schröter, Institute for Information Systems, University of Siegen, Siegen, Germany

ABSTRACT

Social media is becoming more and more important in crisis management. However its analysis by emergency services still bears unaddressed challenges and the majority of studies focus on the use of social media in the USA. In this paper German tweets of the European Flood 2013 are therefore captured and analyzed using descriptive statistics, qualitative data coding, and computational algorithms. Our work illustrates that this event provided sufficient German traffic and geo-locations as well as enough original data (not derivative). However, up-to-date Named Entity Recognizer (NER) with German classifier could not recognize German rivers and highways satisfactorily. Furthermore our analysis revealed pragmatic (linguistic) barriers resulting from irony, wordplay, and ambiguity, as well as in retweet-behavior. To ease the analysis of data we suggest a retweet ratio, which is illustrated to be higher with important tweets and may help selecting tweets for mining. We argue that existing software has to be adapted and improved for German language characteristics, also to detect markedness, seriousness and truth.

Keywords: Computer-mediated communication, disaster management, emergency, data mining, text mining, social media, microblogging, Twitter, crisis informatics, entity extraction, information retrieval, clustering, web-based services, RapidMiner.

1 INTRODUCTION

The usage of social media enables access to real-time data provided by citizens, the news, organizations and companies. Using Twitter communication during disasters is a major challenge because access to tweets is real-time and short-lived. This requires fast decisions on which information to select. This hidden implicit knowledge could add significant value to manage disasters. Many studies during the last decade covered the analysis of social media in disaster management mainly in the USA (starting with Murphy and Jennex (2006) on PeopleFinder and ShelterFinder following Katrina and Palen and Liu (2007), who were anticipating a future of ICT-supported public participation), but only a few case studies about countries such as Germany exist (Reuter et al., 2012). However, Twitter in Germany is used in a different way from that in the US, e.g. usage frequencies. In Germany 56% of internet-users are active on Facebook, whereas just 6% are active on Twitter (BITKOM, 2013). The question arises whether, in general, German tweets contain relevant information as compared to US disaster management studies (e.g. Vieweg, Hughes, Starbird, & Palen, 2010). Furthermore the applicability of existing mining methods to non-English tweets and the selection of appropriate technology is a challenge.

The availability of sources of data, a taxonomy and ontology for guiding search, retrieval and storage have been identified as some key points for organizations to focus on when considering social media (Jennex, 2010). Suggestions for dynamic quality assessment of citizen generated content (Ludwig et al., 2015), implemented as tailorable quality assessment services (Reuter, Ludwig, Ritzkatis, et al., 2015) can only be successful, if these requirements are fulfilled.

In order to address these points, our study (1) aims to first examine whether German emergency tweets contain *additional* and *relevant* information, useful for forecasting, prevention or crisis intervention. This objective is evaluated with retrospective Tweet analysis of the European Flood 2013 data in Germany. Following the structure suggested by Vieweg et al. (2010) this study also investigates (2) if existing computational data mining systems can be applied to German crisis Tweets. Furthermore, we examine (3) which methods (computational versus manual-supervised) are valuable and practical in producing trustworthiness and secure information.

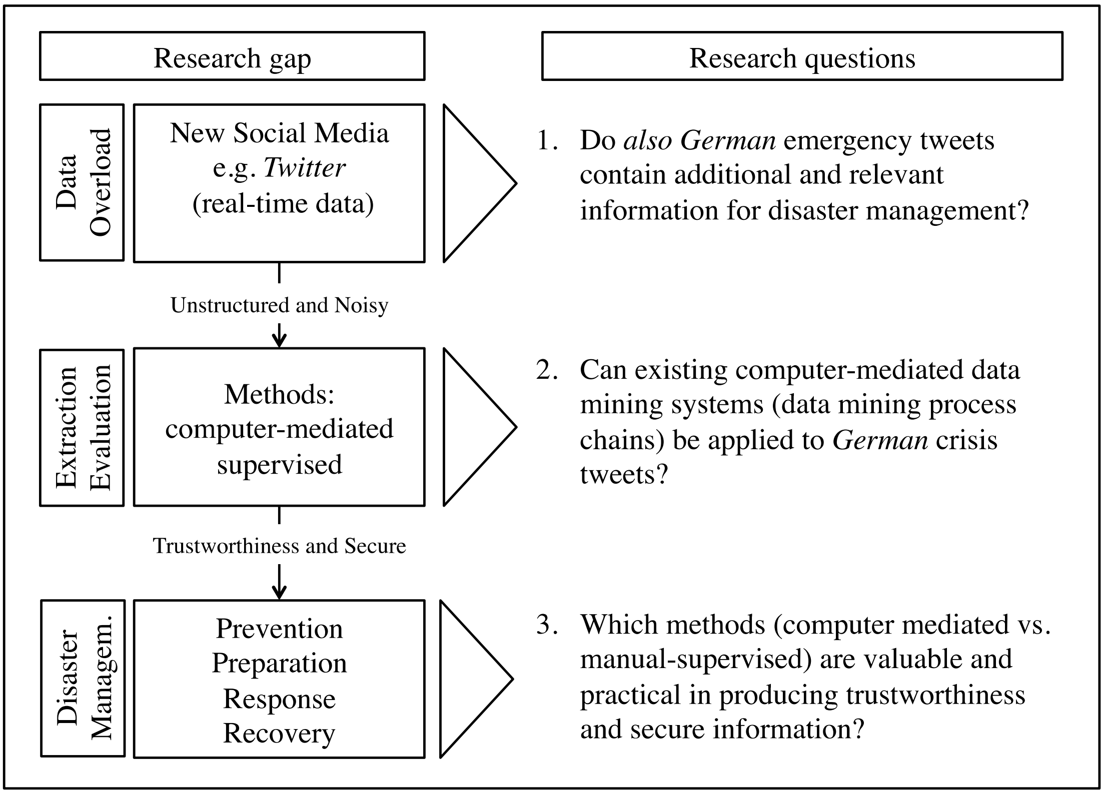

Figure 1: Research Gap and Research Questions

2 RELATED WORK

There has been much research about the use of social media in emergencies. For more than one decade, social media has been used by the public in crisis situations (Reuter et al., 2012): after the terrorist attacks of September 11th, 2001, wikis created by citizens were used to collect information on missing people (Palen & Liu, 2007). During the 2004 Indian Ocean tsunami (Liu et al., 2008) as well as the 2007 Southern California wildfires (Shklovski et al., 2008) photo repository sites were used by citizen to exchange information.

Many other published research papers focus on the use of Twitter during disasters, mainly in the USA since 2008 (Reuter, 2014): The use of *Twitter* has been analyzed scientifically in the context of various crises such as the 2008 hurricanes Gustav and Ike leading the observation of differences between the use of Twitter in crisis and the general use (Hughes & Palen, 2009), the 2008 Tennessee River technological failure, outlining the phenomena of broadcasting (Sutton, 2010), the 2009 Red River Floods, highlighting broadcasting by people on the ground as well as activities of directing, relaying, synthesizing, and redistributing (Vieweg et al., 2010).

A study of Public Information Officers of the Los Angeles Fire Department in 2009 highlights the importance of the information evangelist within organizations (Latonero & Shklovski, 2011). Other studies on the 2009 Oklahoma Fires highlight the role of retweeting (Starbird & Palen, 2010); the 2009 attack on four police officers in Lakewood, Washington shows the ability of Twitter to organize and disseminate crisis-related information (Heverin & Zach, 2010), the 2010 San Bruno Californian gas explosion and fire disaster illustrates that sentiment analysis with emotions performed 27% better than Bayesian Networks alone (Nagy & Stamberger, 2012).

The analysis of 2011 Super Outbreak compared real “emergent groups” (as defined by Stallings & Quarantelli, 1985) and virtual “digital volunteers” (coined by Starbird & Palen, 2011) and distinguished groups of twitterers, such as helpers, reporters, retweeters, and repeaters (Reuter et al., 2013). A journey about the availability of social media during the 2011 San Diego/Southwest Blackout illustrates that “contrary to expectations, the cell phone system did not have the expected availability, and as a result, users had a difficult time using social media to status/contact family and friends” (Jennex, 2012). A study on the 2011 Vancouver Riots revealed the “unintended do-it-yourself-justice”: citizens overruling authorities and enforcing justice on their own terms and by their own means through social media (Rizza et al., 2014). Concerning the organizational use a study on 2012 hurricane Sandy shows that few departments used online channels in their response efforts and that communications differed between fire and police departments and across media type (Hughes et al., 2014). Another study on 2012 hurricane Isaac leads to knowledge which classification algorithms work best in each phase of emergency (Yang et al., 2013). Lang and Benbuan-Fich (2010) provide a first step to formalize the use of social media by proposing a framework based on four modules: (1) selection, (2) facilitation, (3) deliberation, and (4) aggregation.

Besides these studies about English tweets during crises in the USA, other international studies enhance this knowledge: The analysis of the 2008 Sichuan earthquake outlines that people gather and synthesize information (Qu et al., 2009). The 2010 Haiti earthquake was analyzed with the help of translators and reveals the phenomenon of “digital volunteers” (Starbird & Palen, 2011). The case of 2010 Yushu earthquake in China shows that people use microblogging to seek information about the status of people (Qu et al., 2011). The 2010 mass panic at the Love Parade music festival in Germany as well as the 2010 volcano Eyjafjallajökull in Iceland outline the need for duplex communication (Reuter et al., 2012).

In a study about the 2011 Norway attacks the notion of peripheral response has been developed in relation to emergent forms of agile and dialogic emergency response (Perng et al., 2013). A study on the 2011 Egyptian uprising shows how the crowd expresses solidarity and does the work of information processing through recommendation and filtering (Starbird & Palen, 2012), a paper about 2011 Tunisian revolution that social media linked the young activists with actors in other cities (Wulf et al., 2013).

Recent contributions studied the 2013 European Flood in Germany: Fuchs et al. (2013) confirms the potential of Twitter as a distributed 'social sensor' but at the same time highlights some caveats in interpreting immediate results. Reuter, Ludwig, Kaufhold et al. (2015) highlight the self-organization of volunteers which appeared to be cross-platform and cross-media; they propose a tool to support this. Backfried et al. (2015) build on this study and "aim to take the cross-media approach one step further by not only including data from multiple social media platforms, but also by combining information from traditional and social media into a joint system for analysis".

While comparing all these approaches one aspect becomes apparent: Methodologically different approaches are distinguishable - one area primarily conveys qualitative and supervised text analyses. Tweets are viewed, tagged, and categorized by hand: content and information are focal points. The other area conveys technical, algorithm-based analysis, in which the process and the technical information retrieval architecture are applied.

2.1 SUPERVISED TEXT ANALYSIS

Supervised text analysis has been used in various contributions, e.g. (Palen et al., 2010; Reuter et al., 2013; Starbird et al., 2010; Starbird & Palen, 2012; Vieweg et al., 2010).

Vieweg et al. (2010) analyzed tweets during the Oklahoma Grassfire and the Red River Floods in 2009. The authors manually parsed, visualized, and coded data using in-house software *E-Data viewer*[1] allowing "to see the data from close up and far away, and to code each message quickly - and in context. Tools allow a user to zoom in on sections of the data." Each tweet was read and analyzed to gain a fundamental understanding of each event. Upon this analysis, coding categories were created (animal management, place name, evacuation, highway, city name). According to the authors, real-time mining processes are semi-automatable, but crisis information retrieval with structural information extraction requires manual data analysis.

Likewise Starbird et al. (2010) or Starbird & Palen (2012) used qualitative coding to assess information diffusion. Reuter et al. (2013) analyzed real and virtual volunteers. Due to the amount of posts they selected and analyzed 41 Twitterers with the most tweets and 51 Twitterers who were retweeted the most and used open coding (Strauss, 1987) to analyze the material and to uncover interesting phenomena. In summary these methods are adequate to analyze a subset of the data, but not to deal with all messages.

2.2 ALGORITHM-BASED TEXT ANALYSIS

Algorithm-based text analysis is an additional approach:

Fuchs et al. (2013) analyzed German tweets of the European Flood 2013 under the assumption of appropriate geo-referenced messages. The authors applied visual analytics to harvest the 'social sensor' Twitter and revealed caveats in interpreting those immediate results.

---

[1] http://www.cs.colorado.edu/~starbird/e-dataviewer.html

Zheng et al. (2010) applied advanced mining technologies on weakly structured texts. They designed and implemented a web-based application that is primarily based on information extraction algorithms to initially extract entities and relations from original data. It is a text-based process to integrate input data from different sources. Within entity extraction process, *Part-of-Speech (POS)* tagging marks both entities: time and expression. Appropriate labels were identified with the *Viterbi-like* dynamic programming algorithm. The researchers had to train linear *Conditional Random Fields (CRF)* for automation.

Nagy et al. (2012) conducted sentiment detection on the 2010 San Bruno Californian gas explosion and fire disaster. The authors used a *SentiWordNet 3.0*[2] framework for sentiment detection and added emoticons, a sentiment dictionary and short language words (lol, wow) to the process. The researchers proved that annotated list added to *Bayesian Networks* significantly improved sentiment detection by about 27%. The authors recommend a combination of Bayesian Networks with annotated word lists to generate better results.

Yang et al. (2013) studied how to classify tweets about Hurricane Isaac in 2012 into the four categories of the *Four Phase Model* of Emergency Management: mitigation, preparedness, response, and recovery. The authors designed *PhaseVis*, a visualization tool to determine tweet categories. Therefore, three classification algorithms were trained to conduct categorization. *Multiclass Support Vector Machines (mSVM)* are more precise than *Naïve Bayesian Multinomial* classifiers in analyzing tweets for emergencies, according to the *Four-Phase Model*.

Pak & Paroubek (2010) give instructions how to automatically generate a text corpus for sentiment detection for opinion mining. They conducted a linguistic analysis of tweets and created a sentiment classifier that distinguishes between positive, negative, and neutral sentiments. Their classifier is based on the English language but can be adapted to other languages. Dynamic dashboards cluster integrated and classified data. Text data is transformed into vectors. Vectors are checked for similarity. The *term frequency–inverse document frequency (tf–idf)* algorithm measures importance of a word in a text. *K-medoids* algorithms perform data partitioning. A dynamic dashboard ranks partitioned data by relevance. A *Top-K* query is an optimal aggregation algorithm for middleware. Methods for summarizing previous processed and aggregated data using sentence selection and *Minimum Redundancy and Maximum Relevance (mRMR)* algorithms are revealed. Bayesian Networks extract information. Tweets lack of context and contain jargon. A combination of Bayesian Networks with annotated word lists generates better results (Nagy & Stamberger, 2012). *Support Vector Machines (SVM)* are superior to *naïve Bayesian multinomial* (Yang et al., 2013).

Bügel and Zielinski (2013) studied the challenges of multilingual twitter feed analysis. In particular they investigated ten earthquakes and defined four language-specific classifiers. Their preliminary results indicate that such a filter has the potential to confirm forecast parameters.

Text mining is a semi-automated process, requiring supervision. If data is geo-tagged, *geo-spatial patterns* are applied to gather information for emergency management (Bhat et al., 2011). Spatial algorithms are applied before information extraction. General-purpose information retrieval has to cope with social media. The *Firehose* captures the entire *Twitter* universe in real time and serves as input. Classical relational databases, such as *MySQL*, are inappropriate to cope with the amount of data of unstructured tweets. Distributed file system in conjunction with a database management system guarantee efficient and effective operation.

[2] http://sentiwordnet.isti.cnr.it

Spatial clustering defragments data if regions are densely populated or fragmented (Zheng et al., 2010). Optimizing and search methods, such as genetic algorithms, are applied to further enhance inductive quality of outputs.

2.3 SUMMARY AND RESEARCH GAP

Disaster management relies on real-time data analysis for prevention and evaluation. Hence, social media mining seems to make sense. But mining unstructured data is not trivial. Recent research has followed two methodological directions: computational and manual-supervised analyses (Table 1).

Recent research concludes: Supervised mining is more accurate than solely computational mining; however computational mining is faster that supervised mining – a dilemma:

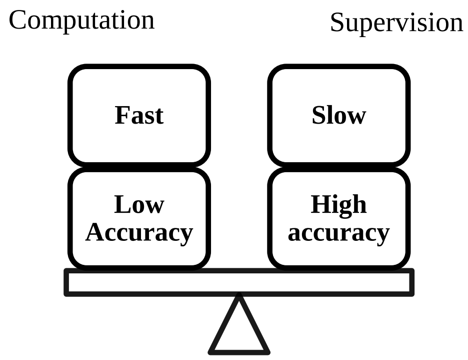

Figure 2: Tradeoff: computation versus supervision

It is necessary to evaluate contributions of both methods, including advantages and disadvantages for reliability and security of information in disaster management. Furthermore none of the studies presented explicitly focuses on the possibilities of application in Germany. However, Twitter in Germany is used in a different way from that in the US. Hence, the question arises whether German tweets in general contain relevant information as compared to US disaster management studies. Another problem is software made for application in English speaking countries. It is crucial to evaluate the utility of extraction systems for German data.

Type	**Exemplary Papers**	**Methods**	**Tools**	**Algorithms**	**Mediums**
Manual	(Reuter et al., 2013) (Starbird & Palen, 2012) (Vieweg et al., 2010) (Starbird et al., 2010) (Palen et al., 2010) ...	Qualitative coding, descriptive statistics	*E-Data viewer, Excel etc.*	-	*Twitter*
Computational	(Zheng et al., 2010) (Barbosa & Feng, 2010) (Paul & Dredze, 2011) (Pak & Paroubek, 2010) (Yang et al., 2013) (Fuchs et al., 2013) (Nagy & Stamberger, 2012) (Klein & Laiseca, 2012) …	Sentiment detection, classification	*NER, SentiWordNet 3.0, Weka, DKPro, PhaseVis*	*POS, (m)SVM, Naïve Bayes (Binary) Multinomial, Viterbi-like, CRF, tf-idf, k-medoids, top-K query, greedy search, mRMR, DBSCAN*	*End-of-Course (EOC) reports, Twitter*

Table 1: Tools and Methods of Social Software Retrieval

3 THE STUDY EVENT

The 2013 European Floods in Central Europe is analyzed to find out if German emergency tweets provide additional information worth mining and are sufficiently minable with standard mining operations. Days of heavy rainfall in the spring of 2013 led to heavy inundations in seven central European countries. This event has been rated as the century`s most extreme event where some of the water gages measured reached record levels.

Fifty-five districts of Germany had to declare this disaster a high-emergency situation, in particular the Federal States of Bavaria, Saxony and Saxony-Anhalt. The town of Passau on the Danube river saw the heaviest floods of all time, during which the drinking water supply as well as teaching at schools and lecturing at universities had to be temporarily suspended. An overflowing lake in Saxony forced several villages to be evacuated; trains could temporarily not circulate on some railroad lines in the area of Dresden town. Heavy thunderstorms caused more devastating floods on June 8 and 9.

The floods of the Saale River in Saxony-Anhalt entailed infrastructural restrictions: the industrial freight traffic had to be stopped as a result of heavy damage to a railroad bridge between the towns of Zeitz and Altenburg, and the Federal Highway number 181 had to be closed for a while. On the one hand, 23,000 people had to be temporarily evacuated in the town of Magdeburg, east of the Elbe River; on the other hand, a transformer station was threatened by water breaking in, which would have jeopardized the power supply to 30,000 households for several months. On the whole, this disaster caused the deaths of eight people in Germany, and the damage reported by the states summed up to about 6.7 billion Euros. To get an overview of the flood, affected areas of Germany are highlighted in a map (Figure 3). The darker the areas in the map are, the higher the corresponding alert-state is.

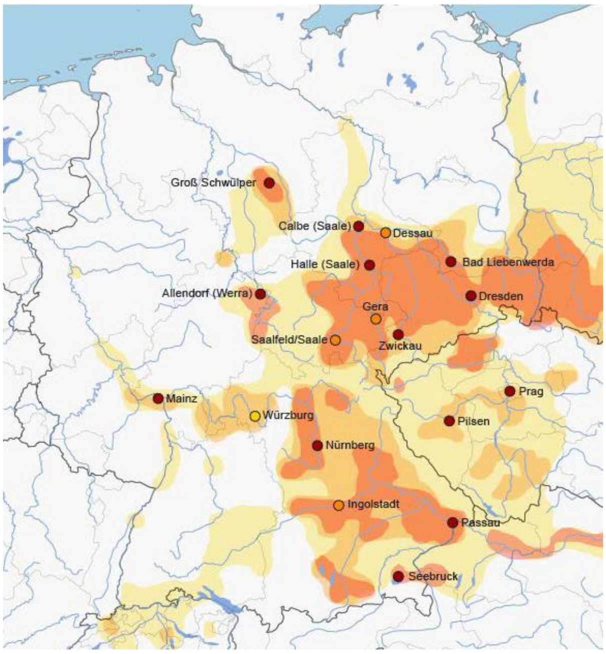

Figure 3: Alert Stages Map (Source: eHyd, VODA, mhz.de, IMG, map by NaturalEarth, June 4th, 2013 from ZEIT ONLINE. http://www.zeit.de/gesellschaft/zeitgeschehen/2013-06/grimma-flut-altstadt-aufraeumarbeiten)

4 METHODOLOGY AND DATA COLLECTION

We overall followed an exploratory approach using descriptive statistics.

46,176 tweets with hashtag #hochwasser (#flood) are captured from 6/7/2013 2:12:33 PM to 6/15/2013 2:24:22 PM, using *Tweet Archivist* (http://de.tweetarchivist.com/) which uses the Twitter Search API. The file is mined with standard software *IBM® SPSS® 22*, *Microsoft Excel® 2011* and *RapidMiner 5.3.008*. In addition, *Text mining* and *Data mining extension 5.3.0* is installed from the website of the *RapidMiner* Marketplace.

Descriptive statistics are computed with *Excel* using function AVERAGE and TREND for the creation of timelines. Hashtags and selected word frequencies are computed with function COUNTIF. Timestamp variables are extracted and accumulated with COUNTIF. Computational keyword mining replicates manual and supervised geo-location tagging by Vieweg et al. (2010). *Geo-location* is extracted with COUNTIF. Besides the initial use of geo-location extraction with *Stanford NER*, own dictionaries for German cities, rivers, and highways are created from German Wikipedia entries, as described in the following. Redundancy exploration is done with *SPSS* function *Identify Duplicate Cases*. Redundancy is measured in all strings to know how this redundancy is evolved. Retweet behavior measures inbuilt redundancy. A process for document similarity is created with RapidMiner. Text mining methods are applied in a certain order: *Information extraction, document clustering, ranking, data summary*, and *spatial clustering*. Tokenization algorithms are applied on the input file: *Non-letter, space, linguistic sentences, and linguistic tokens*. Different tokenization algorithms are tested with same inputs.

The information extraction process is twofold: a standard preprocessing chain for word lists and frequencies, and complex algorithms for document similarity and further processing. A standard chain is set up before complex algorithms. Complex algorithms check redundancy and applicability. A standard preprocessing mining chain for text is: *Tokenization, token filtering, case transformation, stopword filtering, stemming*. A preprocessing chain breaks down unstructured tweets into similar entities that uniforms semantic and syntax. *Token filtering, case transformation,* and *stemming* reduce information. An entity and relation extraction chain is: *Tokenization, point of sentence (POS)-tagging, targeted output*. Document similarity measures redundancy.

5 DATA DESCRIPTION

5.1 TRAFFIC INFORMATION

Tweet creation follows a symmetrical curve (Figure 4). Its climax is at midday. The graph includes tweets and retweets. A polynomial trend (dotted curve) fits to the actual data.

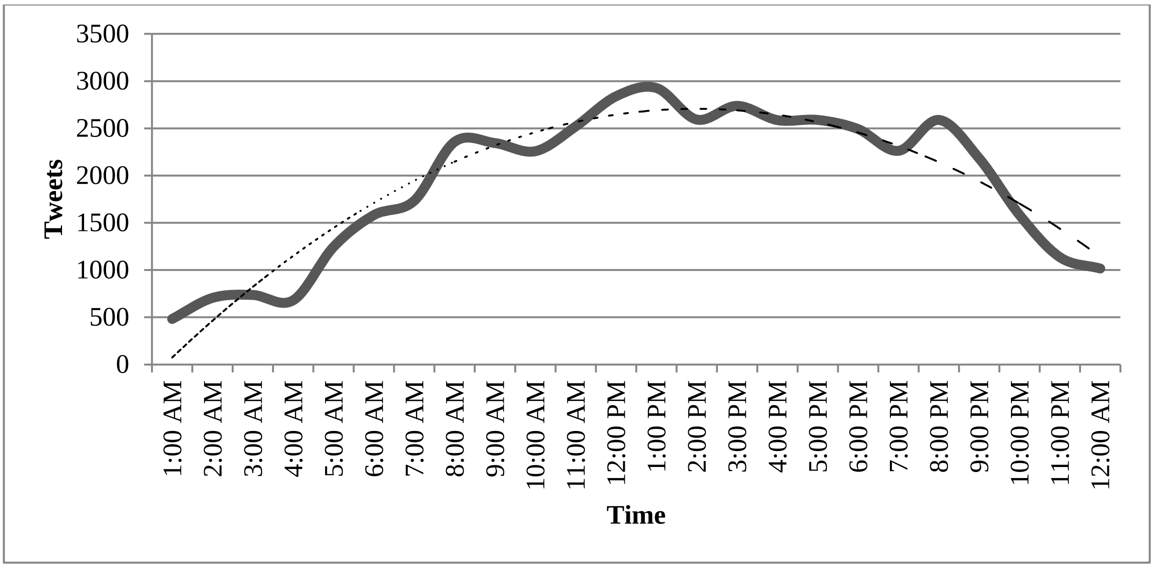

Figure 4: Daily Tweets Timeline with Polynomial Trend

5.2 GEO-LOCATION EXTRACTION

German tweets are insufficiently geo-tagged (7.59% of the tweets of the tweet data set). Geo-location is information that includes city names, river names, and highway names. Vieweg et al. (2010) manually extracted geo-location from tweets as a non-computational mining process. Natural Language Processing (NLP) tools, a subset of NER, aim to extract semantic information from unstructured texts using linguistic concepts such as part-of-speech and grammatical structures (Kao & Poteet, 2007). They are able to recognize locations (LOC), organizations (ORG), people (PER) and miscellaneous information (MISC) from unstructured data. Geocoding services, such as Google Geocoder, can then deliver coordinates for city names or addresses. Different NERs are available: The *Stanford Natural Language Processing Group* [3] provides a Java-based Open-Source NER. Faruqui & Pad (2010) developed two German language models for Stanford NER: *DeWac* uses the web as corpus, but considers just top .de domains (Baroni et al., 2009); *HGC*[4] (huge German corpus) is based on the Stuttgart University Newspaper corpus, consisting of newspapers and legal texts. However, both classifiers insufficiently tagged the hashtagged-words (“#”) and routing prefixes (“@”), (e.g. “#Magdeburg” and “@Magdeburg”). These symbols are ex ante replaced by a space sign to circumvent recognition problems. Both classifiers, based on deWaC and HGC, insufficiently tagged inputs[5] (Table 2):

[3] http://www-nlp.stanford.edu/ner/

[4] http://ims.uni-stuttgart.de

[5] Tagged and untagged Stanford NER files for cities, rivers and highways are provided on the disk: /NER.

Geo-location	**Name**
Cities	Wahlitz, Schemmerhofen, Rockenstuhl, ...
Rivers	Elbe, Oder, Aller, Nahe, Bist, Datteln, Sieg, Elde, …
Highways	A2, A8, A9, B170, B171, …

Table 2: Examples of Non-Tagged Entities with NER's deWaC and HGC Classifier

Emergency management depends on accurate geo-location. Since disasters do not only affect single cities, collateral damage spreads out to peripheral cities. Hence, *NER* with German classifiers sufficiently covers German city name tagging. Just a few exceptions are left out. River detection is insufficient: Although most of the German rivers are tagged, some important rivers are not recognized. Highway tagging is not supported at all. Cities are the predominant geo-information items (Figure 5).

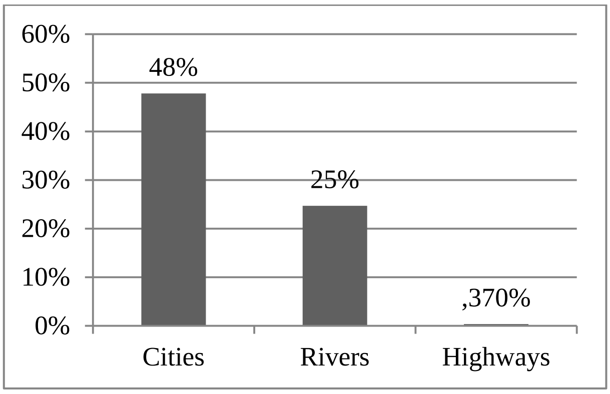

Figure 5: Relative Frequencies of Geo-Location

City names are mentioned twice as often as river names. A small fraction of tweets consists of highway names. Half of the tweets contain relevant geo-location. Geo-locations are further analyzed (Figure 6).

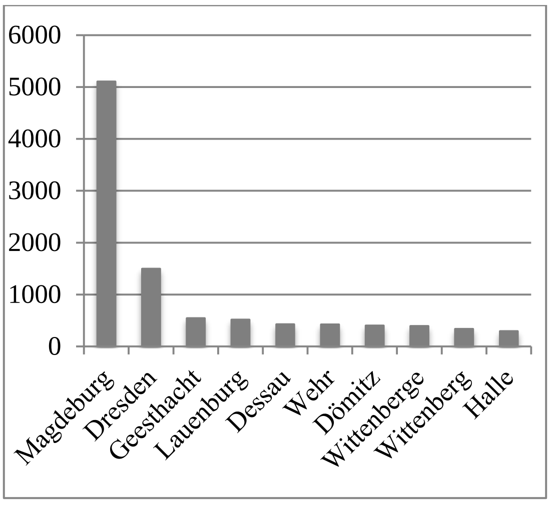

Figure 6: Top-Ten City Names

5.3 HIGH YIELD TWITTERERS

The top three high yield twitterers are broadcast users (Figure 7). The average non-broadcast user created three tweets within the captured period. The top broadcaster, a radio station, produced over seven thousand tweets. High frequency users apply automated publishing[6]. High yield broadcasters are unlikely to publish original and urgent tweets. The following tweets include specific information regarding condition and geo-location:

```
RT @verf_gbar: WE URGENTLY NEED HELPERS AT GÜNTHERKLOTZ PARK ON THE ALB RIVER, BRING SHOVELS! #FLOOD #Karlsruhe #OHNO http:/…

#Flood #GNEVSDORF EP @ #LOWER PART OF HAVEL RIVER ---> #waterlevel increased by about 8 cm from 604 cm to 612 cm within 1 hour. [As of: June 7, 2013

RT @szonline: Dresden buses go again to the suburb of Kleinzschachwitz. More information from Dresden transport services to be had under: http://t.co/zPhe6Sjdhc #flood …
```

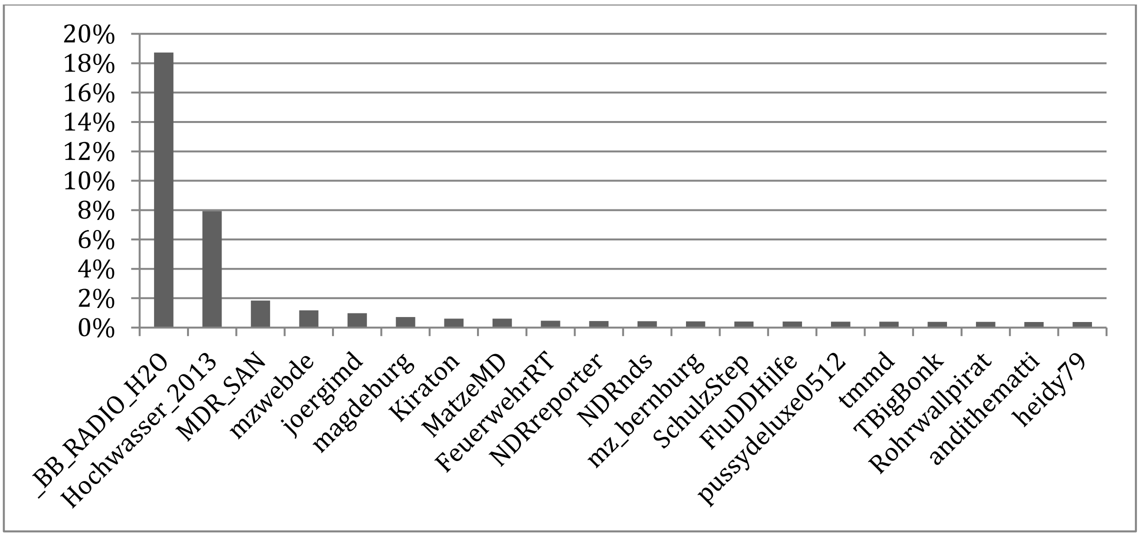

Figure 7: Top-20 Users

5.4 RETWEET INFORMATION

Forwarded messages (retweets) "are deemed especially interesting or noteworthy" (Vieweg et al., 2010). In the data set, we calculated the percentage of certain situational update and non-situational update tweets that are retweets in order to measure the importance of a message. The retweet is marked with the string "RT" within a tweet text. Corresponding strings count related retweets ("RT*dringend*" and "RT*Helfer*gesucht" ("RT*urgent*" and "RT*helpers*wanted)) (Table 3).

These tweets contain water level information. This information is redundant, derived, latent, and distributed by broadcast users (radio or other media). The first row is a benchmark (100%) containing original tweets without retweets ($\wedge \neg$ RT*). The second row contains retweets for all tweets with string `*pegel*` (*water level*). The retweet-ratio is 0.3. One example for unimportant and non-retweeted-message:

[6] http://support.twitter.com/entries/76915

```
#Flood #Ratzdorf @ #ODER ---> #waterlevel dropped by 1 cm from 511 cm to
510 cm within 3 hours. [As of: June 7, 2013 13:20]

#Flood "As for me you may drown": Hackers crack Facebook account...
http://t.co/oWUmaMzjA4
```

Approximately seventy tweets contain the string urgent. Those tweets are original (Table 4) and qualify for high-priority analysis with a retweet-ratio of 3.71.

The following tweets are examples for important tweets:

```
Residential area on the Elbe in Magdeburg town #Buckau completely flooded.

"@Pegel50: #Attention: Assistants in #Doemitz town urgently need food and
drinks! Immediate HELP!!!!!!! "#Elbe #flood #floods

#Followerpower: Helpers are urgently needed at Rennbahn junction. Please
#RT. @HalleSpektrum @HalleON_de @stadt_halle #flood #halle

Cattle up to their necks in the #flood - it is a life and death question
in the wet meadow of Wulfener Bruch http://t.co/E3ZY8NgA2X
```

Critical tweets approximately quadrupled. Each tweet containing the term “`dringend`” (urgend) is retweeted about 3.71, whereas each tweet containing the term “`pegel`” (water level) is just retweeted 0.31. The retweet-ratio states if a subset of the tweets is important and if it qualifies for forwarding. This intuitive concept of this retweet-ratio supports computational mining processes to learn how to important (urgent) and unimportant information.

	String	**Notes**	**Absolute**	**Relative**
1	*pegel* ∧ ¬ RT* (*water level* ∧ ¬ RT*)	Original messages	8,440	100.00%
2	RT*pegel* (RT *water level*)	Retweets	2,599	30.79%
3	*pegel* (*water level*)	Original messages and retweets (sum)	11,039	130.79%

Table 3: Unimportant Retweet Information: “pegel” (water level)

	String	**Notes**	**Absolute**	**Relative**
1	*dringend* ∧ ¬ RT* (*urgent* ∧ ¬ RT*)	Original messages	69	100.00%
2	RT*dringend* (RT*urgent*)	Retweets	256	371.01%
3	*dringend* (*urgent*)	Original messages and retweets (sum)	325	471.01%

Table 4: Important Retweet Information: “dringend” (urgend)

5.5 SERIOUSNESS AND TRUTH

Tweets are not filtered before publishing. During emergencies, tweets, informal discussions, opinions, political and philosophical statements find their way to the masses. All these tweets of the data set that are not exclusively disaster-relevant also contain disaster-specific keywords. In this study, the following keywords are determined: “`#Hochwasser`” (flood), “`ersaufe`” (drown), “`dringend`” (urgent) and “`erwarte`” (expect). Twitterers might use these keywords in other contexts. Or they use them within an emergency context but without serious and true implication. One example for ambiguous information is:

```
#Flood #Barby I drown in groundwater, no postman can reach me - I expect
urgent mail. Where is it?
```

Another example is:

```
Does the #Bundeswehr actually rescue pacifists from the #flood or does it
let them drown consistently?
```

The emergency keyword density in these tweets is high although the information is possibly not serious. Humor, offhand speech, or culture-specific wordplay makes evaluation fuzzy and difficult. Irony is subtle, requiring well-trained and supervised machine-learning patterns. This is a major challenge to future extraction systems.

5.6 MARKEDNESS

The following example also shows the ambiguity of meaning:

```
#Flood #Ratzdorf @ #ODER ---> #waterlevel dropped by 1 cm from 511 cm to
510 cm within 3 hours. [As of: June 7, 2013 13:20]
```

Is the German River “Oder” or the conjunction “oder” (or) meant? Semantic misinterpretation may be the result. Computational mining algorithms extract syntax and entities and have to determine semantic meaning. Markedness occurs if “each of two or more words having the same spelling or pronunciation but different meanings and origins“[7]. Standard mining lacks markedness detection in terms of homonyms in German flood data 2013. It was major problem that the river “Oder” could not be identified as a river because the system classified it as the German conjunction “oder” (or). Hence, systems have to be improved to detect homonyms.

6 DISCUSSION

This paper aimed to answer the questions: (1) Do German emergency tweets contain additional and relevant information for disaster management? (2) Can existing computer-mediated-data mining systems be applied to German crisis tweets? (3) Which methods are valuable and practical in producing trustworthiness and secure information?

6.1 RELEVANT INFORMATION

Q1: Do German emergency tweets contain additional and relevant information for disaster management?

The following paragraph discusses different criteria for answering the question whether German emergency tweets contain additional and relevant information for disaster management (Table 5).

Criterion	**Research subject**	**Method**	**Result**
Traffic (section 5.1)	Adoption and content by Germans	Descriptive statistics	Sufficient traffic
Geo-location (section 5.2)	Geo-information	Supervised geo-location extraction and semantic analysis	Sufficient geo-locations
High yield users (section 5.3)	User base	Frequency analysis	Warnings, evaluations, conditions

[7] http://oxforddictionaries.com/definition/english/homonym

Original data / redundancy (section 5.4)	Duplicate data	Identify duplicates	Three fourth original

Table 5: German Disaster Tweets

Traffic and geo-location information: In this case, users created more than three thousand tweets on average per day. The results indicate that Twitter is used on a large scale in German disasters. Important information used to construct a current disaster situation was found in tweets. Data comprises e.g. highways, cities, rivers, and individual situational updates.

High yield twitterers are professional users within the Twitter domain. They compose tweets similar to reports. These tweets are more structured than autobiographical tweets. The flood analysis revealed structured water level information. Structured tweets are automatically generated and have a special syntax. Although these tweets contain already known information, they are new for the German twitterers.

Redundancy: Emergency tweets contain redundant information: one fourth of data is original. High redundancy characterizes Twitter as broadcasting medium where users act as relays or content multipliers. Retweets forward information to specific users that the relaying users think is helpful.

Hence we conclude that German emergency tweets contain additional and relevant information for disaster management.

6.2 EXISTING MINING SYSTEMS

Q2: Can existing computer-mediated-data mining systems be applied to German crisis tweets?

Results showed that geo-locations and homonyms ("markedness") could not accurately be extracted. Important information may remain undiscovered or unimportant information is filtered as significant. Using keyword-specific entity extraction, situational updates and process tracking during the flood data was more accurate (Table 6).

Criterion	**Research subject**	**Method**	**Result**
Geo-location (section 5.2)	Reliability of NER in extracting German geo-information	NER vs. own dictionaries	NER with German classifier cannot recognize German rivers and highways

Table 6: Software Transfer

There are computational general-purpose extraction systems, trained for the English language. Li et al. (2012) proved NER efficiency as it has been adopted in recent studies. Ritter et al. (2011) showed that standard NLP tools perform inferior on tweets. They are not entirely adoptable for the German language although they are supposed to be. Hence, there is no reliable German application because of their insufficiently trained classifiers.

6.3 COMPUTER VERSUS SUPERVISION

Q3: Which methods are valuable and practical in producing trustworthiness and secure information?

Finally, computational vs. manual-supervised methods are evaluated for producing trustworthiness and secure information (Table 7).

Criterion	Research subject	Method	Result
Retweet information (section 5.4)	Harvests tacit knowledge	Retweet ratio / social filter	“Retweet ratio” is higher in important tweets
Seriousness and truth (section 5.5)	Irony and wordplay	Qualitative analysis	Special evaluation systems
Markedness (section 5.6)	Manual supervision detects more information	Standard mining chain	More precise computation

Table 7: Extraction Methods

A mixture of methods (computer mediated vs. manual-supervised) is crucial for a valuable and practical procedure in obtaining trustworthiness and secure information. On the one side, the retweet ratio (social filter) describes a manually-supervised approach to select subsets of tweets that may be of importance. On the other side, German classifiers lack to detect markedness, truth, and seriousness of information.

7 CONCLUSION

This work investigates (1) whether also German emergency tweets contain additional and relevant information, (2) if existing computational data mining systems can be applied to German crisis Tweets and (3) which methods (computational versus manual-supervised) are valuable and practical in producing trustworthiness and secure information.

Results suggest that German flood tweets contain important and relevant information for disaster management. Tweets contain important information such as geo-locations, warnings, appeals for help, and support requests. Twitter is an appropriate medium to assist German disaster management research efforts. However, Twitter is not a supervised medium such as Wikipedia or Amazon, which support user validation, message ranking, and content editing. Tweet-trustworthiness and security of information are not provided (Starbird et al., 2010). Tweets stand for themselves with additional information, such as the here suggested *retweet ratio*, a social filter that cuts off the distribution of not-important information, as well as user profiles, user activity, and number of followers. The retweet ratio describes a non-computational mining process that is outsourced to users. Users add their tacit knowledge to computational mining processes that might further adjust or calibrate computational algorithms (e.g. machine learning).

This work emphasizes challenges in disaster management as for how to mine social media. The study analyzed whether existing computational mining processes are applicable to German crisis tweets. Results indicate that English language-based systems can not entirely be transferred to the German language. Intelligent systems have to be trained with adequate German classifiers that provide an additional mining module with semantic homonym identifier. Disaster management relies on secure and trustful information that is accurate. Supervision procedures have to be reduced to a minimum by appropriate software. Mining systems have to be flexible for adaption to other disaster types. Word lists have to be permanently updated and adapted to according disaster types. A clustered mining approach is adequate.

For information to be understood and interpreted properly, it is necessary to improve disaster management using computational methods. This analysis of the flood data 2013 showed that computer-mediated systems could not completely map real situations for several reasons: Recent research revealed pragmatic (linguistic) barriers resulting from irony, wordplay, and ambiguity. The importance of tweets can be ex-post derived from the retweet-behavior. Real-time evaluation and extraction are still difficult, also because of those pragmatic problems.

Current approaches are still time- and resource consuming. Preparation of big data will be a big challenge for future predicting or preventing life-threatening situations. Discussed aspects emphasize the major challenge for developing reliable and practical methods. This study also supports a *qualitative* and *quantitative* approach, which combines both methods. According to Palen et al. (2010) it is crucial for research and development to consider and combine the quantity and quality of information.

In future work of our research project EmerGent we aim is to design technology to gather and analyze social media data (Reuter, Ludwig, Ritzkatis, et al., 2015) allowing dynamic quality assessment of citizen generated content (Ludwig et al., 2015), and to detect relevant events and generate alerts. Therefore a combination of different sources – not just Twitter – is necessary as long as social media is used cross-platform (Reuter, Ludwig, Kaufhold, et al., 2015).

ACKNOWLEDGEMENTS

The research project EmerGent' was funded by a grant of the European Union (FP7 No. 608352).

REFERENCES

Backfried, G., Schmidt, C., & Quirchmayr, G. (2015). Cross-Media Linking in Times of Disaster. In *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)*. Kristiansand, Norway.

Barbosa, L., & Feng, J. (2010). Robust Sentiment Detection on Twitter from Biased and Noisy Data. In *Proceedings of the International Conference on Computational Linguistics* (pp. 36–44). Beijing, China, China.

Baroni, M., Bernardini, S., Ferraresi, A., & Zanchetta, E. (2009). The WaCky Wide Web : A Collection of Very Large Linguistically Processed Web-Crawled Corpora. *Language Resources and Evaluation*, *43*(3), 209–226.

Bhat, F., Oussalah, M., Challis, K., & Schnier, T. (2011). A Software System for Data Mining with Twitter. In *2011 IEEE 10th International Conference on Cybernetic Intelligent Systems (CIS)* (pp. 139–144). London, United Kingdom: IEEE. doi:10.1109/CIS.2011.6169149

BITKOM. (2013). *Soziale Netzwerke 2013 - Eine repräsentative Untersuchung zur Nutzung sozialer Netzwerke im Internet*. Berlin, Germany.

Bügel, U., & Zielinski, A. (2013). Multilingual Analysis of Twitter News in Support of Mass Emergency Events. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, *5*(1).

Faruqui, M., & Pad, S. (2010). Training and Evaluating a German Named Entity Recognizer with Semantic Generalization at Stuttgart. In *Proceedings of Conference on Natural Language Processing (KONVENS)*. Saarbrücken, Germany.

Fuchs, G., Andrienko, N., Andrienko, G., Bothe, S., & Stange, H. (2013). Tracing the German Centennial Flood in the Stream of Tweets: First Lessons Learned. In *SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information* (pp. 2–10). Orlando, USA.

Heverin, T., & Zach, L. (2010). Microblogging for Crisis Communication: Examination of Twitter Use in Response to a 2009 Violent Crisis in the Seattle-Tacoma, Washington Area. In *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)* (pp. 1–5). Seattle, USA.

Hughes, A. L., Denis, L. A. S., Palen, L., & Anderson, K. M. (2014). Online Public Communications by Police & Fire Services during the 2012 Hurricane Sandy. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)* (pp. 1505–1514). Toronto, Canada: ACM.

Hughes, A. L., & Palen, L. (2009). Twitter Adoption and Use in Mass Convergence and Emergency Events. In J. Landgren & S. Jul (Eds.), *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)*. Gothenburg, Sweden.

Jennex, M. E. (2010). Implementing Social Media in Crisis Response Using Knowledge Management. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, *2*(4), 20–32.

Jennex, M. E. (2012). Social Media – Viable for Crisis Response? Experience from the Great San Diego/Southwest Blackout. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, *4*(2), 54–68.

Kao, A., & Poteet, S. R. (2007). *Natural Language Processing and Text Mining*. London: Springer.

Klein, B., & Laiseca, X. (2012). Detection and extracting of emergency knowledge from twitter streams. In *Proceedings of the 6th international conference on Ubiquitous Computing and Ambient Intelligence* (Vol. 2). Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-35377-2_64

Lang, G., & Benbunan-Fich, R. (2010). The Use of Social Media in Disaster Situations: Framework and Cases. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, *2*(1), 11–23.

Latonero, M., & Shklovski, I. (2011). Emergency Management, Twitter, and Social Media Evangelism. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, *3*(4), 1–16.

Li, C., Weng, J., He, Q., Yao, Y., & Datta, A. (2012). Twiner: named entity recognition in targeted twitter stream. In *Proceedings of the International Conference on Research and Development in Information Retrieval*. New York, USA. Retrieved from http://dl.acm.org/citation.cfm?id=2348380

Liu, S., Palen, L., & Sutton, J. (2008). In search of the bigger picture: The emergent role of on-line photo sharing in times of disaster. In *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)*. Washington D.C., USA. Retrieved from http://works.bepress.com/cgi/viewcontent.cgi?article=1021&context=vieweg

Ludwig, T., Reuter, C., & Pipek, V. (2015). Social Haystack: Dynamic Quality Assessment of Citizen-Generated Content in Social Media during Emergencies. *Transactions on Human Computer Interaction (ToCHI)*, *21*(4). Retrieved from http://dl.acm.org/citation.cfm?id=2798442.2749461

Murphy, T., & Jennex, M. E. (2006). Knowledge Management , Emergency Response , and Hurricane Katrina. *International Journal of Intelligent Control Systems*, *11*(4), 199–208.

Nagy, A., & Stamberger, J. (2012). Crowd Sentiment Detection during Disasters and Crises. In *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)* (pp. 1–9). Vancouver, Canada, Canada.

Pak, A., & Paroubek, P. (2010). Twitter as a Corpus for Sentiment Analysis and Opinion Mining. In *Proceedings of Seventh International Conference on Language Resources and Evaluation LREC* (pp. 1320–1326). Malta.

Palen, L., Anderson, K. M., Mark, G., Martin, J., Sicker, D., Palmer, M., & Grunwald, D. (2010). A Vision for Technology-Mediated Support for Public Participation & Assistance in Mass Emergencies & Disasters. In *Proceedings of the Visions of Computer Science* (pp. 1–12).

Palen, L., & Liu, S. B. (2007). Citizen communications in crisis: anticipating a future of ICT-supported public participation. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)* (pp. 727–736). San Jose, USA: ACM Press.

Paul, M., & Dredze, M. (2011). You Are What You Tweet: Analyzing Twitter for Public Health. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media* (pp. 265–272). Barcelona, Catalonia, Spain.

Perng, S.-Y., Büscher, M., Wood, L., Halvorsrud, R., Stiso, M., Ramirez, L., & Al-Akkad, A. (2013). Peripheral Response: Microblogging During the 22/7/2011 Norway Attacks. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, *5*(1).

Qu, Y., Huang, C., Zhang, P., & Zhang, J. (2011). Microblogging after a Major Disaster in China: A Case Study of the 2010 Yushu Earthquake. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)* (pp. 25–34). Hangzhou, China: ACM.

Qu, Y., Wu, P. F., & Wang, X. (2009). Online Community Response to Major Disaster : A Study of Tianya Forum in the 2008 Sichuan Earthquake. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS)*. Waikoloa, USA.

Reuter, C. (2014). *Emergent Collaboration Infrastructures: Technology Design for Inter-Organizational Crisis Management (Ph.D. Thesis)*. Siegen, Germany: Springer Gabler. Retrieved from http://www.springer.com/springer+gabler/bwl/wirtschaftsinformatik/book/978-3-658-08585-8

Reuter, C., Heger, O., & Pipek, V. (2013). Combining Real and Virtual Volunteers through Social Media. In T. Comes, F. Fiedrich, S. Fortier, J. Geldermann, & T. Müller (Eds.), *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)* (pp. 780–790). Baden-Baden, Germany.

Reuter, C., Ludwig, T., Kaufhold, M.-A., & Pipek, V. (2015). XHELP: Design of a Cross-Platform Social-Media Application to Support Volunteer Moderators in Disasters. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. Seoul, Korea: ACM Press.

Reuter, C., Ludwig, T., Ritzkatis, M., & Pipek, V. (2015). Social-QAS: Tailorable Quality Assessment Service for Social Media Content. In *Proceedings of the International Symposium on End-User Development (IS-EUD). Lecture Notes in Computer Science*.

Reuter, C., Marx, A., & Pipek, V. (2012). Crisis Management 2.0: Towards a Systematization of Social Software Use in Crisis Situations. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, *4*(1), 1–16.

Ritter, A., Clark, S., & Etzioni, O. (2011). Named Entity Recognition in Tweets : An Experimental Study. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1524–1534). Stroudsburg, USA.

Rizza, C., Pereira, Â. G., & Curvelo, P. (2014). “Do-it-Yourself Justice”: Considerations of Social Media use in a Crisis Situation: The Case of the 2011 Vancouver Riots. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, *6*(4), 42–59.

Shklovski, I., Palen, L., & Sutton, J. (2008). Finding Community Through Information and Communication Technology During Disaster Events. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)*. San Diego, USA: ACM-Press.

Stallings, R. A., & Quarantelli, E. L. (1985). Emergent Citizen Groups and Emergency Management. *Public Administration Review*, *45*(Special Issue), 93–100.

Starbird, K., & Palen, L. (2010). Pass It On?: Retweeting in Mass Emergency. In *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)* (pp. 1–10). Seattle, USA.

Starbird, K., & Palen, L. (2011). Voluntweeters: Self-Organizing by Digital Volunteers in Times of Crisis. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)* (pp. 1071–1080). Vancouver, Canada: ACM-Press.

Starbird, K., & Palen, L. (2012). (How) will the revolution be retweeted?: information diffusion and the 2011 Egyptian uprising. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)*. Bellevue, WA, USA: ACM Press. Retrieved from http://dl.acm.org/citation.cfm?id=2145212

Starbird, K., Palen, L., Hughes, A. L., & Vieweg, S. (2010). Chatter on The Red: What Hazards Thret Reveals about the Social Life of Microblogges Information. In K. I. Quinn, C. Gutwin, & J. C. Tang (Eds.), *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)* (pp. 241–250). Savannah, USA: ACM Press.

Strauss, A. L. (1987). *Qualitative Analysis for Social Scientists*. Cambridge Press.

Sutton, J. (2010). Twittering Tennessee: Distributed networks and Collaboration Following a Technological Disaster. In S. French, B. Tomaszewski, & C. Zobel (Eds.), *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)*. Seattle, USA.

Vieweg, S., Hughes, A. L., Starbird, K., & Palen, L. (2010). Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI)* (pp. 1079–1088). Atlanta, USA: ACM.

Wulf, V., Misaki, K., Atam, M., Randall, D., & Rohde, M. (2013). “On the Ground” in Sidi Bouzid: Investigating Social Media Use during the Tunisian Revolution. In *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW)*. San Antonio, USA: ACM.

Yang, S., Chung, H., Lin, X., Lee, S., & Chen, L. (2013). PhaseVis: What, When, Where, and Who in Visualizing the Four Phases of Emergency Management Through the Lens of Social Media. In *Proceedings of the Information Systems for Crisis Response and Management (ISCRAM)* (pp. 912–917). Baden-Baden, Germany.

Zheng, L., Shen, C., Tang, L., Li, T., Luis, S., Chen, S.-C., & Hristidis, V. (2010). Using Data Mining Techniques to Address Critical Information Exchange Needs in Disaster Affected Public-Private Networks. In *Proceedings of the International Conference on Knowledge Discovery and Data Mining* (p. 125). Washington DC, DC, USA: ACM Press. doi:10.1145/1835804.1835823

CV

Dr. Christian Reuter studied Information Systems at the University of Siegen, Germany and the École Supérieure de Commerce de Dijon, France (Dipl.-Wirt.Inf.; M.Sc.) and received a PhD for his work on (inter-)organizational collaboration technology design for crisis management (Dr. rer. pol.) with summa cum laude. He has worked as a web developer, consultant and researcher and has published more than 60 scientific articles. He is voluntary founding spokesman of the section "human computer interaction in security relevant systems" of the German Informatics Society.

Julian Schröter studied Economics and Business Administration at the University of Siegen, Germany (M.Sc.). His research topic is data analysis of unstructured data. He has worked as a consultant and researcher.