Research Article

Christian Reuter*, Thomas Ludwig, Christoph Kotthaus, Marc-André Kaufhold,
Elmar von Radziewski, Volkmar Pipek

Big Data in a Crisis? Creating Social Media Datasets for Crisis Management Research

Abstract: A growing body of research in the area of information systems for crisis management is based on data from social media. After almost every larger disaster studies emerge with the focus on the specific use of social media. Much of this research is based on Twitter data, due to the ease of access of this (mainly public) data, compared to (more closed) data, such as Facebook or Google+. Based on the experience gained from a research project on social media in emergencies and our task to collect social media data sets for other partners, we present the design and evaluation of a graphical user interface that supports those stakeholders (such as emergency services or researchers) that are interested in creating social media datasets for further crisis management research. We do not specifically focus on the analysis of social media data. Rather we aim to support the gathering process and how actors without sophisticated technical skills can be supported to get what they want and especially need: relevant social media data. Within this article, we present a practice-oriented approach and implications for designing tools that support the collection of social media data as well as future work.

Keywords: Social Media, Big Data, Datasets, Crisis Management

*Corresponding author: Christian Reuter, University of Siegen, Kohlbettstr. 15, 57072 Siegen, Germany,
e-mail: christian.reuter@uni-siegen.de
Thomas Ludwig, University of Siegen, Kohlbettstr. 15, 57072 Siegen, Germany, e-mail: thomas.ludwig@uni-siegen.de
Christoph Kotthaus, University of Siegen, Institute for Information Systems, Germany, e-mail: christoph.kotthaus@uni-siegen.de
Marc-André Kaufhold, University of Siegen, Institute for Information Systems, Germany, e-mail: marc.kaufhold@uni-siegen.de
Elmar von Radziewski, University of Siegen, Institute for Information Systems, Germany, e-mail: elmar-online@gmx.de
Volkmar Pipek, University of Siegen, Institute for Information Systems, Germany, e-mail: volkmar.pipek@uni-siegen.de

1 Introduction

Social media are more and more common and thus have a high relevance for analysts from many different application fields. As van der Aalst [1] argues: "Society, organizations, and people are 'always on'" wherefore "data is collected about anything, at any time and at any place". This constant data collection will be further extended due to the independent trends of social, mobile, cloud and information computing [15]. This is why data scientists' work becomes a more important profession that deals with barriers of how to transform data into a value. Enormous amounts of messages and data are being created, many of them being publicly accessible by anyone on the internet. This 'Big Data' includes messages that became constantly relevant for safety-critical systems, especially crisis management. Police, fire fighters as well as private aid agencies have recognized the relevance and importance of social media content to gain knowledge about users' needs and therefore to improve situation assessment during emergencies [34].

Today, there is already a big amount of existing software tools that helps to yield the hoard of social media data. In terms of the EmerGent project (www.fp7-emergent.eu/), a cross-platform API has already been implemented to collect big data from different social media services. This API, to which we refer as "Cross-Platform Social Media API" (SMA), is a server-side application that gathers publicly available messages from five social media services to use their data for further processing: The two social networks Facebook and Google+, the photo and video sharing platform Instagram, the microblogging service Twitter and the video platform YouTube.

To understand the contribution of our paper, we want to give a brief excursion to issues that arose within our project work. Within this project, we were responsible for creating social media based datasets for the other partners that aim to further analyze those data sets. Each time, anything happened that could be of interest for crisis management (such as the terror attack of Paris), the

related partners asked us to perform short-term data collections based on our implemented SMA to get an appropriate data set. We, then, had to react immediately and start the data collection. To avoid the loss of data based on the short period of time between the occurrence of an event and the start of a collection, we decided to ease the process by designing a graphical user interface (GUI) that is available to the partners.

Within this paper, we describe how such GUI, namely the Social Data Collector, has been designed, implemented, evaluated and then enhanced to support emergency services, but also other stakeholders that could benefit from Big Data (such as marketing, sociology or journalism) to use the existing SMA. Besides, we describe the current state of the art in the area of Social Media Analytics and illustrate which problems exist while gathering and processing such data (section 2). With this theoretical foundation, we present the design and implementation of the SDC for collecting and partially analyzing data (section 3). In terms of the evaluation (section 4), specific and general findings about the use context of this social media analytics tool have been made with twelve participants who participated in cognitive walkthroughs. These findings will be discussed further based on the empirical work and literature as well as implications for design will be derived that should be considered while designing tools for big data gathering (section 5).

2 Related Work: Social Media Detection

The literature about social media analytics and bordering topics is versatile and thus can only be discussed roughly in this paper. At first, some areas of this application field will be presented (section 2.1). Afterwards, some problems with collecting and analyzing social media data will be described (section 2.2) and finally, a research gap will be identified (section 2.3).

2.1 Use Cases for Social Media Analytics in Research

Social media is already used as object and tool for research [20]. *Social media analytics* is, according to Stieglitz et al. [41], "an emerging interdisciplinary research field that aims on combining, extending and adapting methods for analysis of social media data" using text analysis, social network analysis and trend analysis. Besides the

support of the collection process, the SMA (which will be described in section 3) includes among others text analysis to calculate the sentiment of a message. Social media is "an invaluable source of time-critical information during a crisis" [11]. In this section, areas of application of social media analytics with a specific focus on *safety-critical systems* [14] will be presented.

Areas of application include the use inside companies [19], but also in the public. One important field includes crisis management, e. g. information gathering for crises and emergencies. Based on various studies the discipline of *"crisis informatics"* appeared. It "views emergency response as an expanded social system where information is disseminated within and between official and public channels and entities. Crisis informatics wrestles with methodological concerns as it strives to develop new theory and support sociologically informed development of both ICT and policy" [29]. This trend was predicted some years ago: "the role held by members of the public in disasters [...] is becoming more visible, active, and in possession of greater reach than ever seen before" [28]. Following this, our crisis management project collects data from social media to gain information before, during, or after an emergency.

In crisis management studies emerge how social media has been used there, e. g. during hurricane Sandy [16], the European floods [33], but also during smaller events. Recent findings indicate that different types of crises elicit different reactions from Twitter users [26]. Other studies suggest that "there is currently a lack of tools that enable civil protection agencies to easily make use of social media", and the authors suggest a prototype for the "real-time detection of emergency events, related information finding and credibility analysis" [22]. Further areas of application include the creation of datasets for early warning systems [41].

For social scientists, social media constitute a rich source of data for findings about human behaviour (Batrinca & Treleaven, 2014). Social science requires data that is as representative as possible – that becomes more difficult because of bots or spammers, among others. Thus it is difficult to say that social scientific hypotheses can be *verified* by data from social media. It would yet be possible to use data for *generating* hypotheses that can be verified by conventional methods subsequently. Examples are studies that demonstrate how divergent perspectives on crisis are collectively articulated in different Facebook groups on the same topic [6]. Likewise, journalists can use social media for their research (for this, it is important to determine how trustworthy the author of a message is [17]) or to do political opinion research with quantitative methods.

Another area of application of social media analytics is market research and public relations: Social media is used for enterprise-related crisis communication [40]. Information is being searched e. g. about how products are being perceived and how they could be improved. This is also of relevance in critical situations, e. g. if momentous product errors appear and as a consequence are being discussed in social media. Online firestorms, including a high speed and volume of communication [31] need sophisticated technical support to handle the situation and to pre-analyze the data, which sometimes is not possible by individuals. Social media already possess a high value for marketing as, according to Patino et al. [30], the increasing speed of social media made it necessary for the marketing sector to change over to social media: Communication channels like TV, print media, telephone and (letter) post lost much of their meaning and the potential customers spend more and more time online. Multi channel warning systems have to take this into consideration [18].

In bioscience, social media analytics can contribute to gain findings about safety-critical situations, e. g. changes in behavior and their consequences, initiatives against smoking or obesity, or in order to surveil how diseases spread (Batrinca & Treleaven, 2014). Social media analytics are also being used as a business climate indicator for finance management, e. g. as an early warning system for financial crises. Bollen et al. [9] conducted a study according to which the values of the Dow Jones Industrial Average have accorded to a random sample from data from Twitter two or three days ago, with a probability of 87.6 %. A possible point of critique about this method is if manipulation by bots or strategically created accounts can be excluded.

2.2 Challenges Concerning the Detection and Evaluation of Social Media

The sheer mass of social media data makes its analysis a complex effort. Therefore, data mining techniques are used for information retrieval, statistical modelling and machine learning [45]. Hence, data can be pre-processed, analyzed and interpreted. Böck et al. [7] give an overview about current analysis methods of social media including in the "simplest case an overview of current frequently discussed topics, but also [...] more detailed and complex issues, like emotions, the interconnectedness between users". However, like in many of such articles, the focus is on the analysis and the creation of the dataset is not covered.

For the detection of relevant social media data, it is necessary, first to retrieve the data from social media platforms. Required data can be accessed via multiple sources. There are freely accessible sources such as Google Trends, commercial sources like Gnip to collect data in real-time via PowerTrack, or RSS feeds and so on [5]. In the field of marketing, Ullrich and Urbaniak [43] and Römer [38] point out that for social media data usage, it has to be defined which channels and sources should be crawled (e. g. by a URL-defined list or more complex search processes) and whether only publicly accessible social media sources should be accessed for the collection of text-based or also social analytics data. Further, Römer [38] emphasizes that it is important to define inclusion as well as exclusion criteria usually by keywords to collect relevant data only.

The free service Google Trends visualizes the relative frequency of different search keywords (or combinations) over the course of time, but it is necessary to monitor correlations between search keywords and the investigated topic ex ante (e. g. via Google Correlate) [39]. The problem with crawlers is that those services search the whole WWW for required keywords, but are only able to cover the Publicly Indexable Web (PIW) [3] and a lot of password-protected websites prevent automatical searches via crawlers with a "capture" [38]. The same is true for groups and pages on Facebook, which are usually restricted to the persons being member or follower of them. However, using tokens this data can be assessed, if the corresponding user has access to the respective groups [37].

Application programming interfaces (API) provide the opportunity to request social media data. APIs are, in contrast to crawlers, more rapid and precise, but include restrictions of access so that social media monitoring providers often use a combination of both approaches [38]. However, unrestricted access to all data via API is very expensive as providers of social media sites want to monetize their data. That is why Batrinca and Treleaven [5] fear that research within the field of social media will become exclusively for bigger companies, government authorities and a privileged amount of academic researchers so that their published results cannot be criticized or verified. There are more cost factors than the costs for the access of data (ibid.): The software for the acquisition and evaluation of data has to be developed or purchased. Furthermore, there is a need for sufficient computing and storage capacity as well as for the warranty of big data safety. Besides financial aspects, there are a lot more challenges to be considered for the acquisition and evaluation.

– *Mass*: Social media comprise a large number of "objects" (users), social connections and user-generated content [42]. Therefore, vast computing and storage capacity is required [5].

252 — C. Reuter et al., Big Data in a Crisis? Creating Social Media Datasets

DE GRUYTER OLDENBOURG

– *Structure*: Misspelling, shortcuts and ASCII emoticons impede the work of mining software [42]. The lack of structuredness can be corrected automatically (defined as data cleaning, cleansing or scrubbing) or it can be used as a basis for quality analyses [5]. According to our experience, the more common integration of videos, pictures, albums and articles in the news on social platforms contributes the lack of structuredness.

– *Context*: Symbols, ambiguous expressions, irony and many more aspects depend on context [41], metadata of users are often non-existent [42] and posts are linked because the creators are socially connected [42]. Conventional data mining does not regard this aspect, and "most research studies tweets in isolation as single statements without the monologic context, never mind the conversational context" [27].

– *Access:* There are diverse access methods to social platforms (e. g. different APIs) [41] with different technical and business model oriented restrictions [37], foreign languages and expressions that challenge the access [5]. Furthermore, Narang [23] defines missing, incorrect and inconsistent data as possible data problems. Within their study about social networking profiles, Alim et al. [3] retrieved varying profile structures because of different profile types and customized profiles.

– *Ethical consequences*: What consequences arise from collecting, processing, using and reporting of data, even if the data in principle are "public"? [41].

To analyze data, researchers require analytics dashboards, holistic data analysis to combine multiple social media sets and data visualization [5]. Agichtein et al. [2] investigated methods to automatically identify high quality content within Yahoo! answers and presented a general graph-based classification framework for quality estimation in social media. For the study of sentiment analysis of Facebook posts, Neri et al. [24] defined six logical components for the analysis and monitoring of social media: 1) crawler to gather documents or database sources, 2) semantic engine to identify relevant knowledge and therefore to detect semantic relations and facts, 3) search engine which enables natural language, semantic and semantic-role queries, 4) machine translation engine for an automatic translation of search results, 5) geo-referentiation engine for an interactive geographical representation of documents and 6) classification engine for (sub-)clustering results.

Several existing applications support the gathering and analysis of social media in general or for crisis informatics research. For instance, the *EPIC Analyze* platform supports researchers with the collection and analysis of social media data and promotes the core functionality of browsing, filtering, analyzing, and annotating Twitter data [4]. However, the downsides are that the platform is not publicly available and only integrates Twitter. The Java desktop application *Scatterblogs* allows to monitor and visually analyze data from multiple social media [12], but lacks the flexibility of web remote access and management of multiple gathering activities. Furthermore, commercial platforms like *Hootsuite, Sproutsocial, Brandwatch, Twiticent*, or *UberMetrics* each support the monitoring, filtering and analysis of various social media; however, their dashboard reports and visualizations focus on categories such as business performance, competitor benchmarking, and brand analytics [33].

2.3 Research Gap

Currently, there exists a research gap regarding how researchers compile datasets from social media: How do researchers decide which key words to use and which platforms to search? Although there is literature available on search engines, for example how GUIs can be configured for search engines and also on the application areas of social media analytics (see section 2.1). Furthermore, there are already various systems supporting the user to analyze data from social media, but few studies could be discovered focusing on the support of researchers, especially in crisis management, in the data collection processes.

3 Development of the Social Data Collector

This section briefly describes the conception of the Social Data Collector (SDC) (section 3.1), discusses its basic functionalities (section 3.2) and characterizes its implementation (section 3.3).

The SDC is a graphical user interface to create social media datasets for (crisis management) research. It therefore transforms both the data collected by the underlying Cross-Platform Social Media API (SMA) from social networks and the calculated metadata, which are available in JSON-format, into a visualized view. It further facilitates the operation of the SMA allowing new collections of data (so-called *crawljobs*) to be started, stopped, deleted etc. A search function is also offered, permitting social platforms to be

searched spontaneously. Both tools, SMA and SDC, support first and foremost the collection of data but also provide individual functions for a rough analysis of the data. Until now, the SMA has primarily been used for projects within the field of crisis management but further applications are also conceivable, for example to examine the impact of a product image within the field of market research. All the data collected could contribute to sociological studies, or could be used for journalistic opinion research. Although the design of the application was informed from requirements of our crisis management project, the aim was to ensure applicability for different thematic contexts. In addition, the SDC helps to test the functionality of the SMA.

3.1 Conception, Motivation and Related Approaches

The SDC aims to support users in gathering and managing large amounts of data from different social media providers. This primarily includes the day-to-day search for news on particular topics as well as the continuous gathering and archiving over a longer period of time, whereby each search initially constitutes a closed collection. For instance, in terms of crisis management, a researcher might be interested to monitor an actual emergency, but also certain events or locations where an emergency could or is likely to happen either to possibly capture the outbreak of an emergency or, in conjunction with an analysis module, to gather and process indicators of an upcoming emergency. Each of these collections is enriched with key figures (such as keyword, platforms, results, status) which – in connection with detailed views for particular posts – allows basic analysis and assessment operations as well as essential management operations like stopping, continuing and deleting search processes. The SDC itself offers four main functionalities: a) An overview of all collections; b) a detailed view of the resulting posts of a collection; c) the initiation of interval-based search operations (so-called crawl jobs, see section 3.2); and d) the initiation of one-time search operations.

From a design perspective, the focus was set on creating a comprehensive and responsive design. On the one hand, this was to support heterogeneous groups of users in the adoption of the artifact and on the other hand to optimize the web application to run on heterogeneous end-user devices. Apart from the classical and stationary work situation, this should also enable users to initiate data collections about emergent, time-critical events on mobile devices while in a mobile use context.

3.2 Basic Functionality of the Cross-Platform Social Media API (SMA)

The underlying SMA has been established as a fundamental technology for several scientific applications, including XHELP, a Facebook application to support volunteers during natural disasters [33], Social-QAS, an adaptable assessment service for data from social media [36] and CrowdMonitor, a platform for recognizing physical and virtual civil initiatives [21]. The relative core functionality of the SMA consists in configurable data gathering from the social networks Facebook and Google+, as well as the microblogging service Twitter and the multimedia platforms Instagram and YouTube. Behind the unified interface, the complexity and limitations [37] of the different platforms are reduced to a standardized structure, and the data collected is saved persistently to allow standardized processing [44]. In an extended class, additional meta data is calculated by algorithms or saved, even if it has no equivalent in the ActivityStreams specification: This contains the language, sentiment (positive or negative) and popularity (likes, dislikes, shares, retweets, views) of a message, as well as some text statistics (e. g. number of words). The general structure is visualized in Figure 1.

The SDC primarily uses the interfaces to manage crawl jobs and search requests. In terms of the SMA, a *crawl job* is a process which collects data from social media according to the configured search criteria and
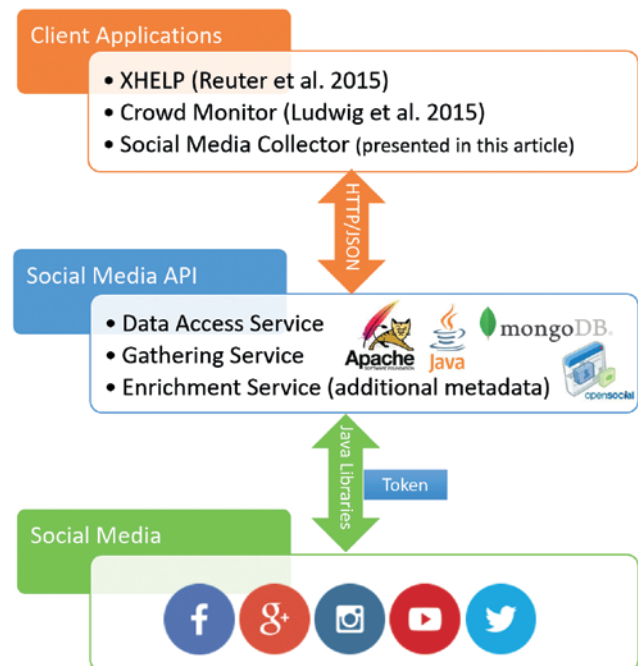

Figure 1: Basic Architecture.

repeats this search request periodically to include recent results. A *search request* on the other hand specifies a one-time search operation. The configuration of the endpoints allows the results to be reduced to certain keywords, social platforms (Facebook, Google+, Instagram, Twitter, YouTube), a period of time (start and end time) or a geographical area (coordinates and a radius). Additional auxiliary help interfaces support the intended data management of SDC.

3.3 Implementation of the Social Data Collector (SDC)

The SDC has been implemented as a web application based on *HTML5*, *CSS* and *JavaScript*, *jQuery* and *jQuery Mobile*. Until now, the SDC has undergone three iterations of development of which the third iteration realizes the essential results of the evaluation (section 4). While the second iteration provided bugfixes and minor changes of functionality with minimal visual impact, the results

of the third iteration were primarily additions to existing functionality which is why we present the state of the third iteration in this section.

The application displays an overview of all current crawl jobs (Figure 2). For each crawl job, keywords, location (global or in a radius around pairs of coordinates), platforms, number of results, the duration of the search interval, and some symbols with the following functionality are displayed: The magnifier icon displays further details about the crawl job, the eye icon allows the results to be viewed. The alternating start / stop icon allows a crawl job to be continued or paused and the recycle bin enables the user to delete a crawl job. A click on the pencil icon opens the 'Edit Crawl job' view, in which a crawl job's keyword etc. can be changed.

On the results page (Figure 3), a compromise had to be found to provide a presentation of very heterogeneous posts which is rich in detail and yet still compact. In this overview, the most important meta-data (if available: number of retweets or shares, positive or negative sentiment, supposed language and so on) is represented by

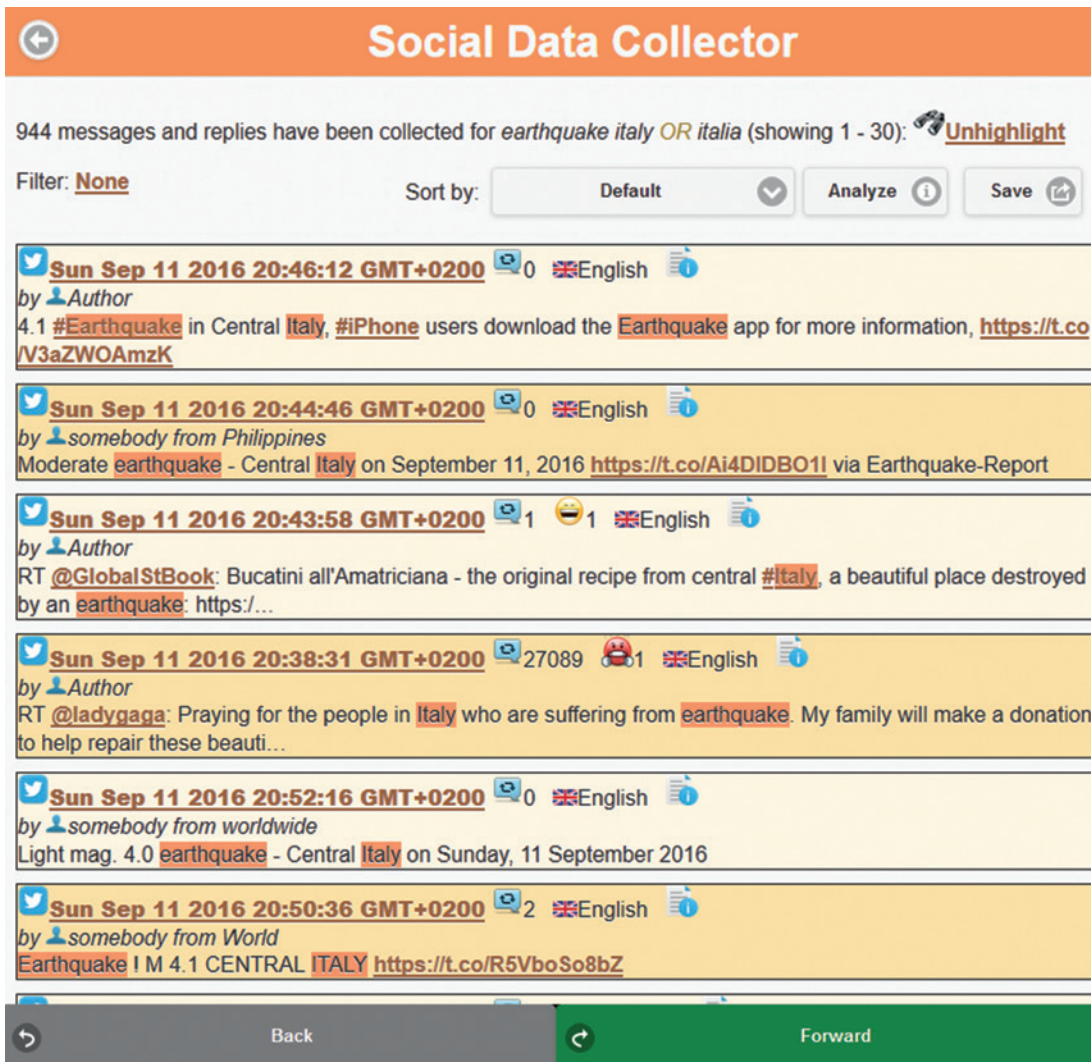**Figure 2:** Overview of the crawl jobs.

Figure 3: Contributions to a crawl job.

symbols, whereas other meta-data can be displayed by clicking on the info icon where necessary. A click on the "author" icon opens a popup with some additional information about the author. If a video is attached to the post, it can be watched by an embedded YouTube player inside the SDC; photos are resized.

Furthermore, functions for analysis, filtering and sorting were added in the third iteration of development (Figure 4, Figure 5). The analysis function determines the distribution of positive and negative sentiment, counts the total of messages per social platform and the distribution of languages as well as the most frequent hashtags and mentions inside the collection. The filter function allows the existing data set to be filtered for desired and undesired words, social platforms and languages. The sorting function provides predefined criteria to sort the data e. g. by length, date, popularity, relevance and sentiment of the posts. Furthermore, the export function allows the collections to be exported into JSON format as ActivityStreams.

Figure 6 depicts the form for creating a new crawl job. The user defines a suitable keyword and search interval as well as selects the social media to be searched. Optionally, the geographical coordinates and their radius as well as the start and end date of the search request can be specified in order to narrow down the set of results. After creating a crawl job, the user is redirected to the overview of crawl jobs. The form for initiating one-time search processes has been designed analogically but excludes the interval definition, and also it redirects the user to the results.

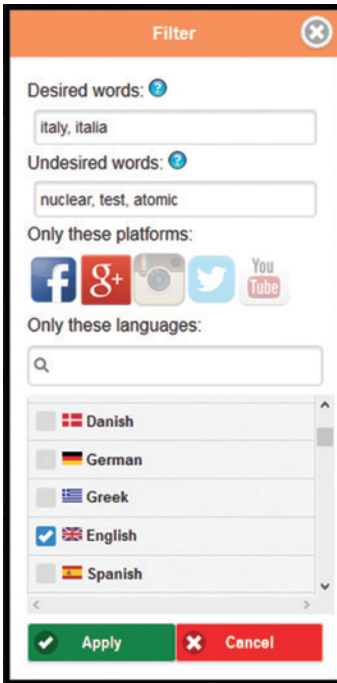Figure 7 presents the FAQ answering many important questions.

Figure 4: Implemented filter words, undesired words platforms, and languages.

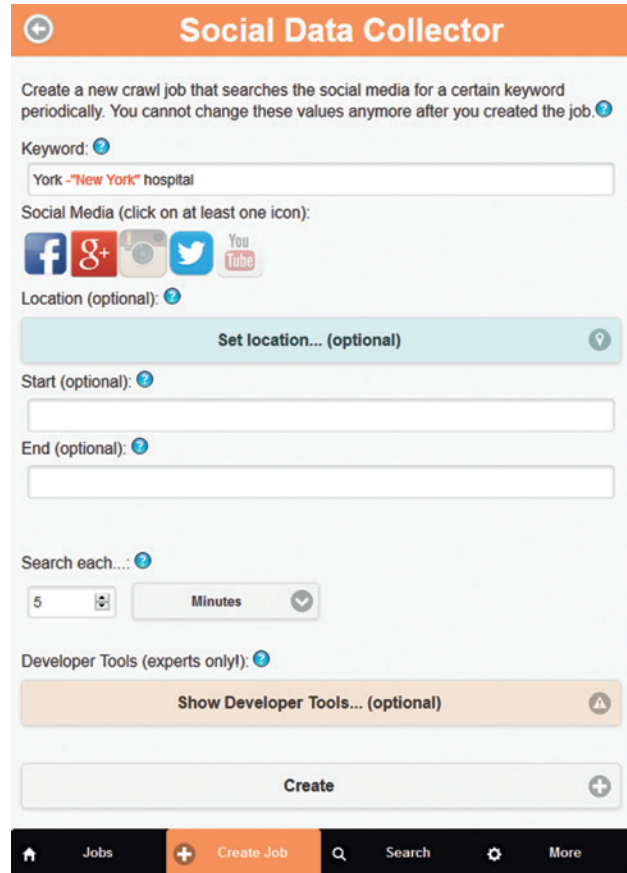**Figure 5:** Implemented filter platform, languages, hashtags and mentions.
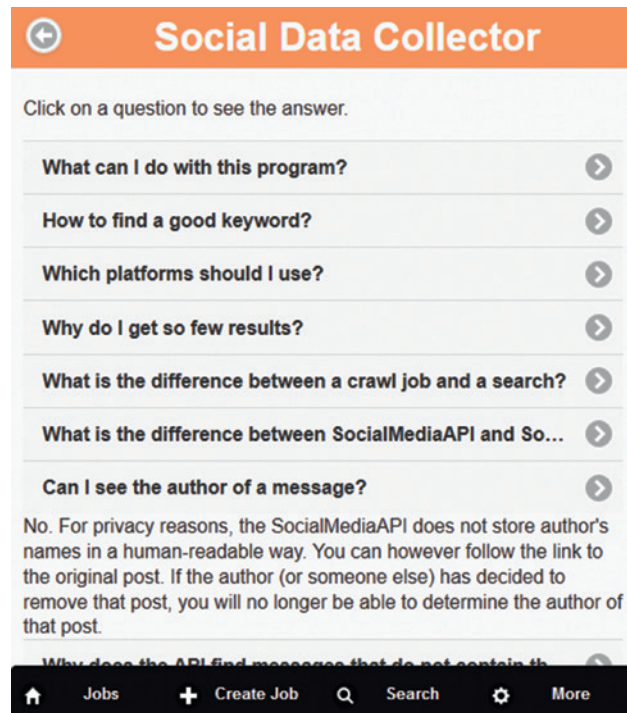
Figure 6: Creating crawljobs.

Figure 7: FAQ.

4 Evaluation: Towards the Ease of Data Collection

This section presents the methodology of the evaluation (section 4.1) followed by the results (section 4.2) and finally the derived requirements for design (section 4.3).

4.1 Methodology

The interest of research in this evaluation was primarily to obtain statements concerning users and context of utilization. The software is fully implemented and the running software was tested by users. Thus the usability of the software was also evaluated. Five (two senior researchers with extensive experience, three junior researchers) of the twelve participants of the evaluation were working in the field of crisis management. The evaluations were conducted in terms of a cognitive walkthrough [32]. Both researchers and participants were able to ask questions and receive answers. The participants were free to create datasets on their topic of interest. Furthermore, exemplary scenarios on market research (impact of a new product), in social science (monitoring if people feel supervised), or journalism (reputation of scandals or rumors about a food company), all with a relationship to crisis management, were given. Additionally, one of the participants chose the scenario of blizzard Jonas that occurred in the USA shortly before the evaluation. Each evaluation took 30–60 minutes, beginning with a short interview. This was followed by the participants conducting a task whilst "thinking aloud" [25]. Unless the conversation partner wished otherwise, the conversations were recorded.

It is assumed that an empirical researcher can directly or indirectly influence the observed person and therefore also the result. For example, people interviewed tend to exaggerate [13]. Therefore, the observation of the users while using the software was also part of the evaluation. Among other things, there were many factors in the use context such as time pressure, colleagues, project context, which were not possible to consider.

It should be noted that after the first three evaluations, a small patch was made on the software. Some bugs were fixed and some texts were changed. As opposed to benchmark tests, comparability was not in focus; it was simply useful to clear those barriers or the bugs instead of leaving them in the software. Due to this, the design of the following interview (from B4) was more fluent and the focus was enabled to be more on the process rather than on difficulties due to misunderstandings caused by usability issues.

4.2 Results I: Findings on the Use

Exploration during dataset generation: A central recognition gained by observation is that simply filling out the form is not enough to create an appropriate collection of data. It is recognized that users immediately waited for the initial results for examination. In many cases new crawljobs were created with more detailed and general search parameters. Apparently evaluating the first results and honing the search parameters is part of the creation process.

Lack of syntax knowledge: Another important observation was that the users did not know how to formulate a keyword in the dialogue. There is a large amount of variety, also among the different social media providers, which are combined in our SMA and finally the SDC: Keywords can be separated by commas, enclosed by quotes, provided with minus (to include), and more. Thus the formulation of a keyword for B9 was not clear and so the person began to experiment: "*Otherwise it can be noticed that the thing apparently searches for words instead of a complete expression… could be possible to change this by placing it in quotation marks.*" *(B9)*. Another ambiguity was the difference between a crawljob (search during a period) and a search (search at a specific time) in the SMA.

Location-dependent searches: Another frequent problem was that many participants were unaware that determining place and period of time restricted their search results as much. Firstly, the SMA has no access to historical data due to technical or business oriented limitations of the individual platforms (at least while using a free account), which means that usually messages cannot be older than two days. Moreover, it is not clear from which coordinates the messages were transmitted. A location-dependent search prohibits most contributions as they are not provided with coordinates (see incompleteness). For example, only 10.3 % of all Twitter users worldwide have the geo-location enabled and therefore around 90 % of all tweets do not contain location information within the metadata (http://www.beevolve.com/twitter-statistics/). As a consequence, many participants did not receive their desired results but did not realize the reason. The integration of messages containing words of the location or from users who have the location in their profile was a mentioned idea to increase the number of messages available.

Big data records: The navigation of big data records was also difficult: To find the first occurrence of a certain platform (e. g. Facebook), you had to click the button "forward" frequently. B5 searched for a filter function to filter the results by platform: "*Up to now I have almost only seen Twitter everywhere; this one seems to be a record of pure Twitter-data. Let's see if I can click on the symbol, to see the relevant Facebook or Google+ entries for the respective search. But this does not work*" (B5).

Presentation: The SDC received a lot of praise for the way they presented the contributions found: Pictures were shown, videos could be watched immediately in SDC, keywords were highlighted and by clicking on a symbol, a popup appeared saying: "*Well, this is a very good overview of how to gain first insights. And then there is also a positive feeling – I don't know what it is about, lets click on it... [...] Ah, they've highlighted the keyword for me. [...]*" (B8). The presentation of the results (even if the system does not aim at analysis) was perceived to be an important aspect.

4.3 Results II: Suggestions for Improvement

Filter: Many interviewees requested functions to filter or sort the messages with regard to platform, sentiment, hashtags, languages and many more. "*Do you have filters? [...] I only want to see the positives, only the negatives, only the ones in English or the ones in Spanish*" (B1).

Export: Furthermore, many persons requested the possibility of exporting the messages to other software (e. g. Atlas, a software used mostly in qualitative research or qualitative data analysis): "*Yeah, basically what I would do to complete the task is like search the content, analyze with Atlas maybe, or with any software that can analyze the content itself*" (B3). "*Of course you have to specify more precisely what a suitable output format could be, and if it belongs to Activity Streams or thinking any other format. CSV, which could rather be read from another program*" (B11).

Statistics: Although the intension of the application is to gather the data, not to analyze it, many people also requested visual statistics (for a quick overview of the data records): "*The first thing I see shouldn't be the post itself, it should be like this big picture result: This many positives, this many negatives, in this language [...] before I see the raw data. [...] I'm missing some sort of visualization before getting the raw data [the messages]*" (B1). This indicates that gathering and analyzing (or at least pre-analyzing) go hand in hand.

Transparency: In many cases, it was unclear to the users how the SMA works. For example, why there was

nothing found while searching with a keyword, why certain messages were referred to search results and how the sentiment is calculated: "*A user should be able to get a statement on why nothing is displayed – whether, for example, it is because no results were to be found, or because the search ran out of capacity*" (B5). "*Who determines these negative or positive sentiments, is it done just by an algorithm? [...] Well, actually it says it is calculated with the help of positive keywords in the text [...], but who determines the positive keywords*" (B6).

Mobile version: In order to always be able to create a crawljob when an interesting incident occurs (for example, a natural disaster) also when not in the office at the computer and thus probably miss important messages from the early stages, there ought to be a mobile version of SDC: "*Mobile Version – as already mentioned – would be good to execute the work well and fast. Wherever I am, wherever I get the idea – getting work done fast and easily*" (B12).

4.4 Results III: Implications for Design

The following Table 1 offers an overview of the most important empirical results and design requirements. Some usability problems of former SDC versions have not been included. However, other usability problems result from the general design challenges of software could be useful for the design of similar software.

5 Discussion and Conclusion: Big Data in a Crisis or Big Data Crisis?

This section comprises a summary of the outcomes of this study followed by a presentation of the findings and finally the prospect of further possible developments of the GUIs (and partly even of the APIs).

5.1 Summary: One Step towards Easier Data Set Creation

Social media data plays an important role in crisis management, such as during major emergencies, in public relations of companies, for social scientists, journalists, or in bioscience (see section 2). We found it worthwhile to study how researchers create data collections for messages from social media and how they are used further: How can the keywords and platforms which

Table 1: Overview of the empirical results and their requirements on design.

Aspect	Empirical Result	General Design Requirements
Syntax across media	When formulating keywords, it was not clear if quotation marks, commas, plus- or minus signs could be used.	The formulation rules should be made very transparent to the user.
Punctual search and crawljobs with a time frame	The difference between a search and a crawljob was not clear.	The introductory sentence of the search page, for instance, should explain the difference between different types of search, e. g. onetime search or crawljobs with a time frame.
Dataset creation and analysis are related	The crawljob view suffers from a bad navigability of the results (e. g. to find the first Facebook-result, you had to click "Forward" several times). The meaning of flags and icons was not clear.	Functions for sorting and filtering are needed to find relevant contributions more quickly; also the intention of the software is to just create the datasets. By clicking on flags and icons, a popup explaining their meaning should open.
Highlighting functionality and user experience	The coloured highlighting of keywords, the presentation of pictures and videos belonging to a message, the pop-up messages and the possibility to immediately play videos in the message overview were praised. Support through help pages and pop-up messages were welcomed by some people while others prefer to experiment.	The presentation format of messages (including the coloured highlighting of keywords) should be maintained. Offering information by pop-up messages should be complemented with a FAQ-designed help page.
Export, visualization and smart data collection	Request for filter and sorting functions, an export function, visual statistics, a sortable crawljob list and a history of keywords and crawljobs. Crawljobs should accept countries in addition to coordinates. A full-text search was also requested.	Filter and sorting functions in particular as well as statistics (not graphical if appropriate) and sortable crawljob lists should be implemented. Additionally, there should be an export function, a history, a full-text search and a search for countries in addition to coordinate-based search.
More transparency	During the evaluation it was often unclear why the SMA did not find any results, why some posts were displayed (although they did not contain the keyword) or how positive and negative sentiment is calculated.	To enable users to better understand the way the SMA works, a FAQ with the most important questions should be implemented at the very least. Furthermore, a glass box approach [10], where the underlying functionality becomes visible, also showing the bottleneck, might be interesting.
Mobile version	Crawljobs are generated spontaneously, while an event or disaster is taking place. A version that can also be used on a smartphone is therefore necessary.	Besides the responsive design, pictures and texts from messages should not be displayed next to each other. As a consequence, an app for small displays is required for spontaneous work.
Naming and personal annotations	Description texts for crawljobs were requested as well as the function of annotating or marking interesting posts or metadata to have them at hand in the future.	Description texts for crawljobs, marks for posts and metadata and a message annotation feature should be added, e. g. to support scientific coding.

have to be searched through be determined? How is the collected data used? We were able to build on literature about how to design GUIs for search engines (e. g. how search results should be presented). Furthermore, one can find literature on social media analytics regarding challenges and fields of application, with a specific focus on crises and emergencies. For instance, market researchers search for information about how products are perceived and how they could be improved. Moreover, social media is a rich data source for social scientists regarding findings about human behavior. However, they need valuable data for their purposes which are often hampered by bots or spammers. Journalists can use social media for investigations on the one hand or for political opinion polling by using quantitative methods on the other hand.

The Cross-Platform Social Media API (SMA), on which this work is based, enables the creation of crawljobs that, in a specified interval and according to a defined keyword, search for new messages in the social media platforms Facebook, Google+, Instagram, Twitter and YouTube. It is further possible to specify a certain place and period of time to collect messages, and additionally, onetime searches can be started to spontaneously display current

results from the target platforms within a few minutes. In the EmerGent project on social media in crisis management we wanted to acquire more flexibility in enabling our project partners with varying technical knowledge from various disciplines to create datasets. Thus we decided to ease the process by designing a user interface based on our technical interfaces. Accordingly, the Social Data Collector (SDC) was designed and implemented to cover the central functions of the SMA.

As the evaluation of SDC revealed, the interface offers the possibility of creating, starting, stopping, deleting and showing crawljobs without prior programming skills as well as executing searches. The collected messages are visualized and therefore more lucid; pictures and videos are shown; meta-data is presented with icons; URLs, hashtags and mentions are highlighted; navigating through result pages is feasible, and much more. The forms for the creation of crawljobs and searches contain help bubbles that display a help text with only one click. Within this, the usage context of the software was investigated, e. g. for what purposes the SDC can be used and how users handle the software, what difficulties exist and what aspects need to be improved.

However, it was also pointed out that – even though only a few details were required for the creation of crawljobs or searches – it was commonly unclear how the relevant forms should be filled in: Too restrictive details regarding keywords, social platforms, coordinates and period of time lead to few or even no results. Vague searches lead to irrelevant results, e. g. by using too general keywords. In case of a (nearly) empty result list, the underlying reason needs to be obvious. If the result list is long but the results barely fit the topic, the functions of filtering the results and perhaps finding hints for better search keywords are necessary and have to be improved. This is why the interviewees especially requested filtering and sorting functions, a statistical preparation of the data, export functionality for the messages (to analyze them with other applications), more transparency regarding the SMA (e. g. why crawljobs do not find results) and a mobile version of the application and / or an improved responsive design.

During the further development of the SDC many of these aspects were implemented: Filtering, sorting and analyzing. Moreover, export and import functionalities in JSON format were added; a help page contains information even on the functionality of the SMA and the responsive design of the application was improved for mobile use. The pop-up messages and description texts were revised and by clicking on an icon, an explanation pops up to improve the self-descriptiveness of the software.

5.2 Contributions: How Should a Tool Support this Process?

In this work findings on the required characteristics of a tool to collect and evaluate data from social media were gathered. In the evaluation it became apparent that insufficient comprehensibility and transparency constitute a severe usability obstacle: In general, it is important to explain the terms "crawljob", "keyword", "sentiment", "interval" etc. in the application because the understanding of such terms cannot be presumed. Furthermore, it should be explained which criteria can restrict the sample space of a crawljob or search and how. Concerning the functionalities, it should be evident (a) why a crawljob did not find any (or only a few results); (b) according to which criteria a sentiment of a message is computed; (c) what meaning the computed value has, e. g. "information content" in the example of the SMA; and (d) why a message is listed as a result even though the keyword does not occur in the message etc. Moreover, the evaluation indicated that users want to see first results directly after the start of a crawljob. For this, the requirement for a filtering and sorting functionality or for a (visual) statistic about the collected messages respectively has arisen and was implemented. Software which primarily serves for collection and less for analysis should provide such functionalities. Anyway, users desire an additional function for exporting the collected data into other software for analysis.

Another interesting focus was on a mobile version of the application, e. g. to spontaneously create crawljobs from a smartphone or tablet. This is especially true as long as the occurrence or starting point of crises and emergencies is not necessarily known before (e. g. terroristic attacks, political upheavals, natural disasters) and therefore ad hoc decisions are necessary [35]. While [4] discuss the requirements of reliable, scalable, extensible, and efficient environments for data gathering and analysis in crisis informatics, we argue that flexibility in terms of access, interfaces and use is a critical factor to achieve more complete crisis-related datasets and to provide suitable input for, e. g., event detection. Further findings were gathered regarding how users create datasets (they especially like to combine search keywords and search across all available platforms at the same time).

It was confirmed that difficulties within the analysis of data from social media, as pointed out in the literature, especially lie in the aspects of a lack of structure, mass, incompleteness [42] and context dependence [41]. The "context dependence" of messages in social media reaches as far as the collected messages can often hardly be interpreted not only by algorithms but also by humans, as it turned out

during this work. For the aspect "lack of structure" it was shown that text messages in social media are often complemented or even substituted by pictures, albums, articles or videos (possibly even several at a time). To evaluate these messages correctly, software in the field of social media analytics has to be able to analyze the meaning of picture and video material, too. In addition, the problem has to be mentioned that even the meta data is subject to the problems mentioned above (especially lack of structure, incompleteness and context dependence): Therefore, a "like" on Facebook might have another meaning compared to a like on YouTube or Google+ (e. g. because Facebook has more members or because one can partially obtain more rewards for a "like" of a specific page). Another example is the lack of information like "views" (in terms of number of views) for the comparison of popularity of messages on many other platforms besides YouTube.

Nevertheless, the evaluation revealed that the software (SMA and SDC together) is not equally suitable for the accomplishment of each task. The SMA and SDC especially are suitable when messages are required for daily, current events or messages have to be collected over a specific period of time. For non-daily and current tasks, searches by web search engines will lead to much better results because high-quality contributions from online newspapers or portals feed into the searches that cannot be collected by the SMA.

Advantages and disadvantages of the evaluation method were detected and confirmed: They are qualified for the evaluation of software that is at an early development state like the Social Data Collector. They enable more insight into the thoughts, experiences and expectations of the interviewees than only observational methods. A disadvantage is that those methods are time-consuming and users need more time for the processing of a task because of dialogues and questions.

Another important aspect is that many of the requirements are also relevant for non-crisis situations, however in time-critical situations data collections have to be started spontaneously, sometimes outside office hours, and without sophisticated staff-support. Therefore, technical possibilities have to be provided to researchers in an easy and lightweight way.

5.3 Outlook: Still a Lot to do

The range of functions of the SDC as well as of the SMA could be expanded in a few ways. For instance, the software could be developed in such way that it does not only provide the collection but also the analysis of the data. The qualitative analysis could be achieved by the additional function of annotating messages, e. g. with comments, evaluations or keywords. To make this possible when several persons work simultaneously with the data, collaborative functionalities tools could be implemented for the arousal of attention, the conflict resolution and further communication.

For a quantitative analysis, implemented improved algorithms that use linguistic tools could be used to make a point about a text. So instead of evaluating messages one-dimensionally as "positive" or "negative", the six GPOMS dimensions "calm", "alert", "sure", "vital", "kind" and "happy" [8] could be applied. Furthermore, the current algorithm only takes into account words of a positive and negative word list but does not recognize negations. A formulation like "not good" would be rated positively at the moment of the SMA because a positive word exists within the phrase. Also interesting for the quantitative analysis would be mechanisms for the recognition of duplicates (respectively retweets) as well as of bots, spammers or special accounts that were created strategically to post comments and feedbacks in favor of individual persons or brands.

Including a visualization of the data into the software could be interesting as well. Therefore, charts could be used to make the composition of the collection of messages visible at a glance regarding speech, platforms, sentiment etc. A timeline could illustrate in which periods of time particularly many messages occurred to a certain keyword and moreover facilitate the navigation through the search results temporally such as to jump to messages of a specific calendar week by a few clicks.

Additional platforms could be integrated into the SMA, e. g. the possibility to search through forums and weblogs or websites as well. Moreover, uniform search syntax oriented towards the syntax of the search engine Google meanwhile have been implemented by us. Thus, cross-platform signs as "+" or "−" could be used to mark certain keywords as compulsorily necessary or undesirable words. Quotation marks mark words as related. Instead of exact searches as before maybe one could enable vague searches as well where e. g. spelling mistakes would be tolerated or not all referred keywords would have to occur in the same message to be found.

However, this study has some limitations: It is uncertain whether or how the filtering and search functions will help the users to evaluate if the crawljob works according to the given criteria. Furthermore, it has not been investigated whether the new help site will support the users to handle the application. Some wishes for improvement, like the integration of export functionality for special analyzers, could not have been implemented yet. In the empirical part of this work it has not been possible to investigate

how the users utilize the software over a longer period of time in a real working context, potentially even while teamworking. In this context, some important questions arise: a) Which implications for design have to be taken into consideration for domains that use the SDC frequently (e. g. journalists) and for rare use, e. g. in crisis management. b) In crisis management in particular, which personnel or unit will actually use the software (control room, section control, operatives, etc.)? c) referring to b), depending on the actual users in crisis management, is a mobile version necessary for conducting spontaneous crawljobs? In addition, it might be necessary in the future to implement new functions of the underlying social media providers as well on the SDC to test and comfortably use them.

References

[1] van der Aalst, Wil M P: Data Scientist: The Engineer of the Future. In: Mertins, K.; Bénaben, F.; Poler, R.; Bourrières, J.-P. (Hrsg.): *Enterprise Interoperability VI: Interoperability for Agility, Resilience and Plasticity of Collaborations*. Cham: Springer International Publishing, 2014 — ISBN 978-3-319-04948-9, S. 13–26.

[2] Agichtein, Eugene; Castillo, Carlos; Donato, Debora: Finding High-Quality Content in Social Media. In: *Proceedings of the International Conference on Web Search and Data Mining*, 2008, S. 183–193.

[3] Alim, S.; Abdul-Rahman, R.; Neagu, D.; Ridley, M.: Data retrieval from online social network profiles for social engineering applications. In: *2009 International Conference for Internet Technology and Secured Transactions, (ICITST)*: IEEE, 2009 — ISBN 978-1-4244-5648-2, S. 1–5.

[4] Anderson, Kenneth M; Aydin, Ahmet Arif; Barrenechea, Mario; Cardenas, Adam; Hakeem, Mazin; Jambi, Sahar: Design Challenges / Solutions for Environments Supporting the Analysis of Social Media Data in Crisis Informatics Research. In: *2015 48th Hawaii International Conference on System Sciences*, 2015 — ISBN 9781479973675, S. 163–172.

[5] Batrinca, Bogdan; Treleaven, Philip C.: Social media analytics: a survey of techniques, tools and platforms. In: *AI & SOCIETY* Bd. 30, Springer-Verlag (2014), Nr. 1, S. 89–116.

[6] Birkbak, Andreas: Crystallizations in the Blizzard: Contrasting Informal Emergency Collaboration In Facebook Groups. In: *Proceedings of the Nordic Conference on Human-Computer Interaction (NordiCHI)*. Copenhagen, Denmark: ACM, 2012 — ISBN 9781450314824, S. 428–437.

[7] Böck, Matthias; Köbler, Felix; Anderl, Eva; Le, Linda: Social Media-Analyse – Mehr als nur eine Wordcloud? In: *HMD Praxis der Wirtschaftsinformatik* (2016).

[8] Bollen, Johan L T M; Mao, Hulna: Predicting economic trends via network communication mood tracking | US 8380607 B2.

[9] Bollen, Johan; Mao, Huina; Zeng, Xiaojun: Twitter mood predicts the stock market. In: *Journal of Computational Science* Bd. 2 (2011), Nr. 1, S. 1–8.

[10] du Boulay, Benedict; O'Shea, Tim; Monk, John: The black box inside the glass box: presenting computing concepts to novices. In: *International Journal of Man-Machine Studies* Bd. 14 (1981), Nr. 3, S. 237–249.

[11] Castillo, Carlos: *Big Crisis Data – Social Media in Disasters and Time-Critical Situations*: Cambridge University Press, 2016.

[12] Chae, Junghoon; Thom, Dennis; Bosch, Harald; Jang, Yun; Maciejewski, Ross; Ebert, David S.; Ertl, Thomas: Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition. In: *IEEE Conference on Visual Analytics Science and Technology 2012, VAST 2012 – Proceedings* (2012), S. 143–152 — ISBN 9781467347532.

[13] Heinecke, Andreas M.: *Mensch-Computer-Interaktion: Basiswissen für Entwickler und Gestalter*, *X.media.press*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012 — ISBN 978-3-642-13506-4.

[14] Herczeg, Michael: *Prozessführungssysteme: Sicherheitskritische Mensch-Maschine-Systeme und interaktive Medien zur Überwachung und Steuerung von Prozessen in Echtzeit*: De Gruyter, 2014.

[15] Howard, C.; Plummer, D. C.; Genovese, Y.; Mann, J.; Willis, D. A.; Smith, D. M.: *The nexus of forces: Social, mobile, cloud and information*, 2012.

[16] Hughes, Amanda Lee; Denis, Lise A St; Palen, Leysia; Anderson, Kenneth M: Online Public Communications by Police & Fire Services during the 2012 Hurricane Sandy. In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. Toronto, Canada: ACM, 2014, S. 1505–1514.

[17] Hughes, Amanda Lee; Tapia, Andrea H.: Social Media in Crisis: When Professional Responders Meet Digital Volunteers. In: *Journal of Homeland Security and Emergency Management* Bd. 12 (2015), Nr. 3, S. 679–706.

[18] Klafft, M: Diffusion of emergency warnings via multi-channel communication systems an empirical analysis. In: *Autonomous Decentralized Systems (ISADS), 2013 IEEE Eleventh International Symposium on*, 2013, S. 1–5.

[19] Koch, Michael; Richter, Alexander: *Enterprise 2.0 – Planung, Einführung und erfolgreicher Einsatz von Social Software in Unternehmen*: Oldenbourg-Verlag, 2009 — ISBN 9783486590548.

[20] König, Christian; Stahl, Matthias; Wiegand, Erich: *Soziale Medien: Gegenstand und Instrument der Forschung*: Springer, 2014.

[21] Ludwig, Thomas; Reuter, Christian; Siebigteroth, Tim; Pipek, Volkmar: CrowdMonitor: Mobile Crowd Sensing for Assessing Physical and Digital Activities of Citizens during Emergencies. In: *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*. Seoul, Korea: ACM Press, 2015.

[22] Mccreadie, Richard; Macdonald, Craig; Ounis, Iadh: EAIMS : Emergency Analysis Identification and Management System. In: ACM (Hrsg.): *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval (SIGIR '16)*. New York, 2016 — ISBN 9781450340694, S. 1101–1104.

[23] Narang, Rishi K.: *Inside the Black Box: The Simple Truth About Quantitative Trading*. Hoboken: Wiley, 2009.

[24] Neri, F.; Aliprandi, C.; Capeci, F.; Cuadros, M.; By, T.: Sentiment Analysis on Social Media. In: *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*: IEEE, 2012 — ISBN 978-1-4673-2497-7, S. 919–926.

[25] Nielsen, Jakob: *Usability Engineering*. San Francisco, USA: Morgan Kaufmann, 1993.

[26] Olteanu, Alexandra; Vieweg, Sarah; Castillo, Carlos: What to Expect When the Unexpected Happens: Social Media Communications Across Crises. In: *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15. New York, NY, USA: ACM, 2015 — ISBN 978-1-4503-2922-4, S. 994–1009.

[27] Palen, Leysia; Anderson, Kenneth M: Crisis informatics: New data for extraordinary times. In: *Science* Bd. 353, American Association for the Advancement of Science (2016), Nr. 6296, S. 224–225.

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

[29] Palen, Leysia; Vieweg, Sarah; Liu, Sophia B.; Hughes, Amanda Lee: Crisis in a Networked World: Features of Computer-Mediated Communication in the April 16, 2007, Virginia Tech Event. In: *Social Science Computer Review* Bd. 27 (2009), Nr. 4, S. 467–480.

[30] Patino, Anthony; Pitta, Dennis A.; Quinones, Ralph: Social media's emerging importance in market research. In: *Journal of Consumer Marketing* Bd. 29, Emerald Group Publishing Limited (2013), Nr. 3, S. 233–237.

[31] Pfeffer, Jürgen; Zorbach, Thomas; Carley, Kathleen M.: Understanding online firestorms: Negative word-of-mouth dynamics in social media networks. In: *Journal of Marketing Communications* Bd. 20 (2104), Nr. 14, S. 1–2.

[32] Polson, P.; Lewis, C.; Rieman, J.; Wharton, C.: Cognitive walkthroughs: A method for theory-based evaluation of user interfaces. In: *International Journal of Man–Machine Studies* Bd. 36 (1992), S. 741–73.

[33] Reuter, Christian; Ludwig, Thomas; Kaufhold, Marc-André; Pipek, Volkmar: 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, 2015.

[34] Reuter, Christian; Ludwig, Thomas; Kaufhold, Marc-André; Spielhofer, Thomas: Emergency Services Attitudes towards Social Media: A Quantitative and Qualitative Survey across Europe. In: *International Journal on Human-Computer Studies (IJHCS)* Bd. 95 (2016), S. 96–111.

[35] Reuter, Christian; Ludwig, Thomas; Pipek, Volkmar: Ad Hoc Participation in Situation Assessment: Supporting Mobile Collaboration in Emergencies. In: *ACM Transactions on Computer-Human Interaction (ToCHI)* Bd. 21, ACM (2014), Nr. 5.

[36] Reuter, Christian; Ludwig, Thomas; Ritzkatis, Michael; Pipek, Volkmar: 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, 2015.

[37] Reuter, Christian; Scholl, Simon: Technical Limitations for Designing Applications for Social Media. In: M.~Koch,~A.~Butz, ~&~J.~Schlichter (Hrsg.): *Mensch & Computer: Workshopband*. München, Germany, Germany: Oldenbourg-Verlag, 2014, S. 131–140.

[38] Römer, Stephan: Erhebungsmethoden und Tools. In: Bundesverband Digitale Wirtschaft (BVDW) e. V. (Hrsg.): *Social Medie Kompass 2015/20162*. Düsseldorf, 2015, S. 14–16.

[39] Steffen, Dirk: Verknüpfung von Daten aus Sozialen Medien mit klassischen Erhebungsmethoden. In: König, C.; Stahl, M.; Wiegand, E. (Hrsg.): *Soziale Medien – Gegenstand und Instrument der Forschung*. Wiesbaden: Springer Fachmedien Wiesbaden, 2014 — ISBN 978-3-658-05326-0, S. 97–110.

[40] Stieglitz, Stefan; Bruns, Axel; Krüger, Nina: Enterprise-Related Crisis Communication on Twitter. In: *Wirtschaftsinformatik Proceedings*, 2015.

[41] Stieglitz, Stefan; Dang-Xuan, Linh; Bruns, Axel; Neuberger, Christoph: Social media analytics – An Interdisciplinary Approach and Its Implications for Information Systems. In: *Business and Information Systems Engineering* Bd. 6 (2014), Nr. 2, S. 89–96.

[42] Tang, Jiliang; Chang, Yi; Liu, Huan: Mining social media with social theories: a survey. In: *ACM SIGKDD Explorations Newsletter* Bd. 15, ACM (2014), Nr. 2, S. 20–29.

[43] Ullrich, Susanne; Urbaniak, Mathias: Datenquellen für die Social-Media-Datenerhebung. In: Bundesverband Digitale Wirtschaft (BVDW) e. V. (Hrsg.): *Social Medie Kompass 2015/2016*. Düsseldorf, 2015, S. 11–13.

[44] World Wide Web Consortium: *ActivityStreams Vocabulary*. — W3C Working Draft.

[45] Zuber, Mohammed: A Survey of Data Mining Techniques for Social Network Analysis. In: *International Journal of Research in Computer Engineering and Electronics* Bd. 3 (2014), Nr. 6, S. 1–8.

Bionotes

Christian Reuter
University of Siegen, Germany
christian.reuter@uni-siegen.de

Christian Reuter, PhD, studied Information Systems at the University of Siegen and the École Supérieure de Commerce de Dijon, France and received a PhD for his work on (inter-)organizational collaboration technology design for crisis management. Before his scientific engagement he was full time consultant for a telecommunication corporation. He has acquired, conducted and managed national and international consultancy and research projects and has published scientific articles in information systems, human-computer interaction, crisis management and social media. He is divisional director for crisis information systems at the University of Siegen and voluntary founding chairman of the section "HCI in safety-critical systems" of the German Informatics Society.

Thomas Ludwig
University of Siegen, Germany
thomas.ludwig@uni-siegen.de

Thomas Ludwig, PhD, studied Information Systems at the University of Siegen and the University of Newcastle (Australia). He received his PhD for his work on the design of ICT Tools for researching complex information infrastructures. He holds the divisional direction of Cyber-Physical Systems at the University of Siegen, where he published lots of research articles in the fields of computer-supported cooperative work, human-computer interaction, crisis management as well as internet-of-things.

Elmar von Radziewski
University of Siegen, Germany
elmar-online@gmx.de

Elmar von Radziewski, B.Sc. in Information Systems, is a student of Information Systems at the University of Siegen. He graduated as a B.Sc. in 2016 and now works as a student research assistant while pursuing his M.Sc. degree. His primary research interest is user-centered software development based upon the combination of interview-based empirical work with agile development approaches. In his work, he assumes the roles of both a software designer and a programmer.

Christoph Kotthaus
University of Siegen, Germany
christoph.kotthaus@uni-siegen.de

Christoph Kotthaus, Dipl. Wirt.-Inf., studied Information Systems at the University of Siegen and the University of Newcastle, Australia. He did his degree at a rail technology company about mobile device management. After three years of experience as a project manager in this company he returned to the University of Siegen to do his PhD. He continues the work he did during his studies as a research assistant in the domain of crisis management and also works at a research project related to Cyber-Physical Systems as a project leader. His research is focused on computer-supported cooperative work and human-computer interaction in both these domains.
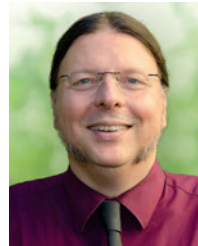
Volkmar Pipek
University of Siegen, Germany
volkmar.pipek@uni-siegen.de

Volkmar Pipek, PhD., studied Computer Science and Economics at the University of Kaiserslautern and received a PhD degree in Information Processing Science from the Laboratory of HCI and Group Technology at the University of Oulu, Finland. He is Professor for Computer Supported Cooperative Work and Social Media at the Institute for Information Systems at the University of Siegen, Germany.

Marc-André Kaufhold
University of Siegen, Germany
marc.kaufhold@uni-siegen.de

Marc-André Kaufhold, M.Sc., studied Information Systems at the University of Siegen. In his master thesis, he investigated the applicability and potentials of Flow theory in human-computer interaction. During his study he primarily assisted in the research projects InfoStrom (2010–2013, BMBF) and EmerGent (2014–2017, EU). He now works as a research assistant at the SME Graduate School and Chair of Computer-Supported Cooperative Work and Social Media while pursuing his PhD. His research is focused on IT-supported crisis management, volunteerism in social media, and Flow theory.