A Framework for Enhancing Real-Time Social Media Data to Improve the Disaster Management Process

Syed Attique Shah, Dursun Zafer Şeker and Hande Demirel

Abstract Social Media datasets are playing a vital role providing information that can support decision-making in nearly all domains. This is due to the fact that social media is a quick and economical approach for collecting data. It has already been proved that in case of disaster (natural or man-made) the information extracted from Social Media sites is very critical to Disaster Management Systems for response and reconstruction. This study comprises of two parts: The first proposes a framework that provides updated and filtered real time input data for the disaster management system through social media, and the second consists of a designed web user API for a structured and defined real time data input process. The aim of this study is to propose a framework that can filter and organize data from the unstructured social media sources through recognized methods and bring this retrieved data to the same level as that acquired through structured and predefined mechanisms, such as a web API. Both components are designed such that they can potentially collaborate and produce updated information for a disaster management system to carry out accurate and effective decision-making.

Keywords Crowd-sourcing · GIS · Disaster management · Design science · Social media · Real-time data · API

S.A. Shah (✉) · D.Z. Şeker · H. Demirel
Faculty of Civil Engineering, Department of Geomatics, Istanbul Technical University, Maslak Sarıyer, Istanbul, Turkey
e-mail: shah@itu.edu.tr

D.Z. Şeker
e-mail: seker@itu.edu.tr

H. Demirel
e-mail: hande.demirel@itu.edu.tr

M.P. Peterson (ed.), *Advances in Cartography and GIScience*, Lecture Notes in Geoinformation and Cartography, DOI 10.1007/978-3-319-57336-6_6

1 Introduction

During disasters (e.g., floods, earthquake, storms, large fire, etc.), people tend to report and share their observations, findings and suggestions on various social media sites. However, it is still a challenge for researchers as how to automatically filter out useful information and make that information searchable and accessible for emergency services.

Social Media applications are considered very useful to collect information in case of any disaster because it is the fastest and the cheapest source to provide effective, updated and relevant information for decision-making. The practical use of such applications is resulting in a large number of academic studies to research many aspects of social media in Disaster Management (Veil et al. 2011). Social Media provides the opportunity to contribute and disseminate valuable information be it in the shape of text, pictures, audio and video that is necessary for disaster management processes and communications (Turel and Serenko 2012).

Current research states that the communication services such as Short Message Services (SMS) or social media (Facebook, Twitter) have the ability to improve the regular and updated transmission of valuable information and provide the effective resources in all of the disaster management life-cycle phases and aid in developing a disaster resilient community (Chan 2013; Ahmed 2011). In case of any disaster, the emergency service authorities should be able to access the social media networks and blogs to identify the source and scale of the disaster and develop the recovery plans according to the requirements of affected community. In addition, authorities should be able to observe online communities to detect mounting trends and possible hotspots that can substantiate a disaster (Chan 2013). When it comes to managing disasters efficiently, the main thing for government and emergency agencies is to be provided with accurate, updated and complete information. There may be serious consequences if the information is incorrect and or not delivered in a timely manner (Erskine and Gregg 2012; Ostermann and Spinsanti 2011).

Real-time Geospatial Information Systems use crowd-sourcing to map social feeds using geotag metadata with longitude/latitude coordinates. With hardware-based sensors, GIS can be combined with social media to create a compressive source of information to understand the disaster situation accurately, as general public can be added in the process to improve the overall awareness (Adam et al. 2012; Yin et al. 2012). However, there are numerous challenges associated with social media data collection methods, such as variable quality of the data, intelligently managing the big volume of social media feeds, checking and verification of the data, accurately geo-parsing map information and the need to find the right balance between verifying and responding to new information.

Prior research shows the practicality of this research using frameworks such as Twitcident (Abel et al. 2012) and SensePalce2 (MacEachren et al. 2011). Moreover, crowd-souring is emerging as a corresponding research area that works as a virtual sensor to collect data from every possible potential source, and to support activities

from basic mapping (e.g., Open-Street Map) to providing data for a disaster management system (e.g., CrisisMappers.org, Ushahidi).

A very good example of web API for collecting social media data in case of an earthquake and then model earthquake activity according to that accessed information is a website named as 'Did You Feel It?' by U.S Geological Survey (USGS) (http://earthquake.usgs.gov/earthquakes/dyfi/).

2 Social Media in Relation with Disaster Management

Social Media are applications that are totally depended on user-generated content or applications in which the user-generated contents and activities play an important role in increasing the overall value of that application or service (Kaplan and Haenlein 2010).

On the other hand, a disaster is a sudden event that seriously affects the normal routine of a community or society. It could be natural calamities such as earthquakes, tornadoes or hurricanes, but also man-made destructive activities such as terrorist attacks or industrial accidents. It has not only an economic and environmental impact, but also an important humanitarian component.

Disaster management can be modeled into four phases, namely mitigation, preparedness, response, and reconstruction. Having a good strategy for each of the phases is essential for efficient disaster management. In order to accomplish this, managers need proper information about the different activities within each of the four disaster management phases (Jayaraman et al. 1997). With social media, information is now accessible in real-time, so those activities can be planned more accurately. With disaster management models and its phases a lot of research work needs to be done to map where and how social media information can be used to improve the decision-making (Cozzolino 2012). Research has identified that the use of social media after a disastrous incident increase and even surpasses the use of other conventional communication methods such as fixed phones (Velve and Zlateva 2012). Social Media sites like Facebook, Twitter and YouTube can be very useful for real time data collection when tsunamis, earthquakes, floods and other natural disasters strike.

According to Crystal Washington (2016) "Social media is the application that;

- Provides valuable information to those in a disaster area pre- and post-disaster (via the Internet, if available, or SMS updates).
- Drives awareness to those outside the affected areas, generating volunteers and/or donors.
- Connects displaced family and friends.
- Provides information about unclaimed property and in worst case, scenarios, bodies.
- Offers information about aid centers and other resources available to those affected."

Five discrete uses for social media in disaster management are identified by (MacEachren et al. 2011) as,

(a) "to disseminate information to the public (e.g., for alerts)"
(b) "to gather information from the public (e.g., crowd-sourcing)"
(c) "to coordinate with crisis management professionals"
(d) "to monitor activities of crisis management professionals"
(e) "as input to situational assessment for crisis management".

Instead of categorizing the existing research into disaster management phases directly, we added the social media application because they represent a more fine-grained perspective of social media within a disaster. Moreover, applying the social media applications to traditional disaster management phases allows us to integrate pervious research about potential social media activities into the disaster management model. This will act as a theoretical lens to classify the existing research into disaster management phases.

3 Proposed Research Framework

This research work contributes to the discipline of design science for the information systems domain. Based on reviewed literature and detailed perception of related technologies, we propose a model that can provide updated and essential input data for the disaster management system through two different potential crowd-sourced data platforms; the social media component and a designed web user API component.

The proposed model is divided in two sections based on the method for data input. As seen in Fig. 1, the first section is the web user API, named as 'Disaster Analytic API' which is a designed web template to take the data according to some structured parameters already defined and to somewhat enforce users to follow the designed format for feeding data. This second section is the 'social media system component' which is taking in the unstructured social media data and applying selective recognized methods to filter the data and present it for further processing to the disaster management system component. This model emphasizes the quality of real time data gained through crowd-sourcing, which is normally considered noisy and unfit to use for accurate decision-making processes. This model can be a good example to compare and measure data quality gained through a designed API for structured data and real time data gained through social media. There is no such comparison currently found in the literature and this area needs to be focused to make the coward-sourcing data gathering mechanism more effective and open to both structured (web data entry API) and unstructured (social media) for new actionable insights.

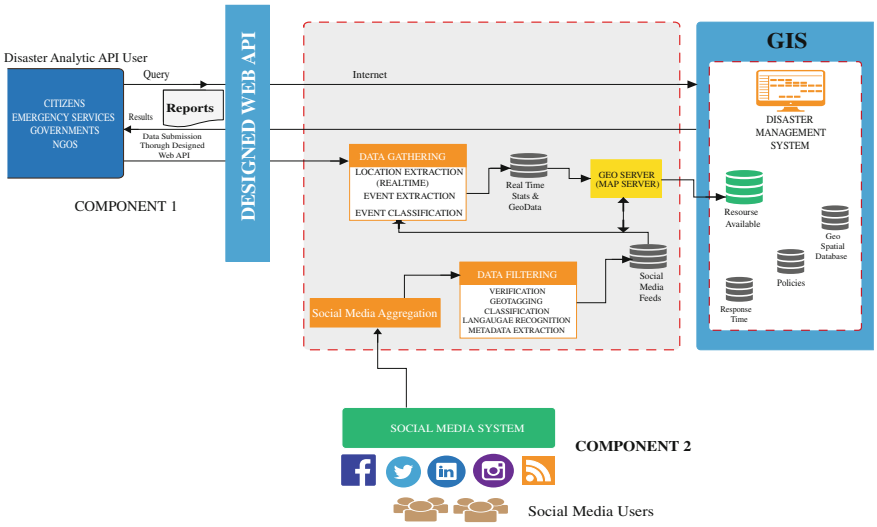

Fig. 1 Proposed research model

3.1 Designed Web User API Component

This "Disaster Analytic" component is a proposed web API for receiving information for any disaster in a structured data entry template. It includes, all potentially required sections such as location, time, scale, description, attachments (photo or video) etc. Additionally, the Disaster Analytic API can also manage queries from the user requesting reports regarding the disaster situation. A rough outline for the interface is proposed in Fig. 2. The potential users of this web API can be citizens/volunteers, governments' officers, emergency services and NGOs who can provide data with a defined format and query. Reports can be searched and acquired through the disaster management system.

The main aim for the design and implementation of this web API is making sure that accurate and machine-readable information is received. It also should be user friendly, and can be operated without much setup or training. Every user needs to be identified and an authorized login should be created. There are many different types of format constraints to be followed in order to get the structured and machine readable information. The notable constraints can be user's log in ID (who), type of incident (what), photo of the identified incident (what), scale (what), date and time (when), location of the incident (where), classification and description (how). To increase the user-friendliness, constraints such as date and time need to be provided automatically so that user should be able to select the options rather than make manual entries in the form. The user can request the forms directly from the disaster management system whenever they want and are able to see the recent verified social media feedbacks and trends being shared regarding the disaster.

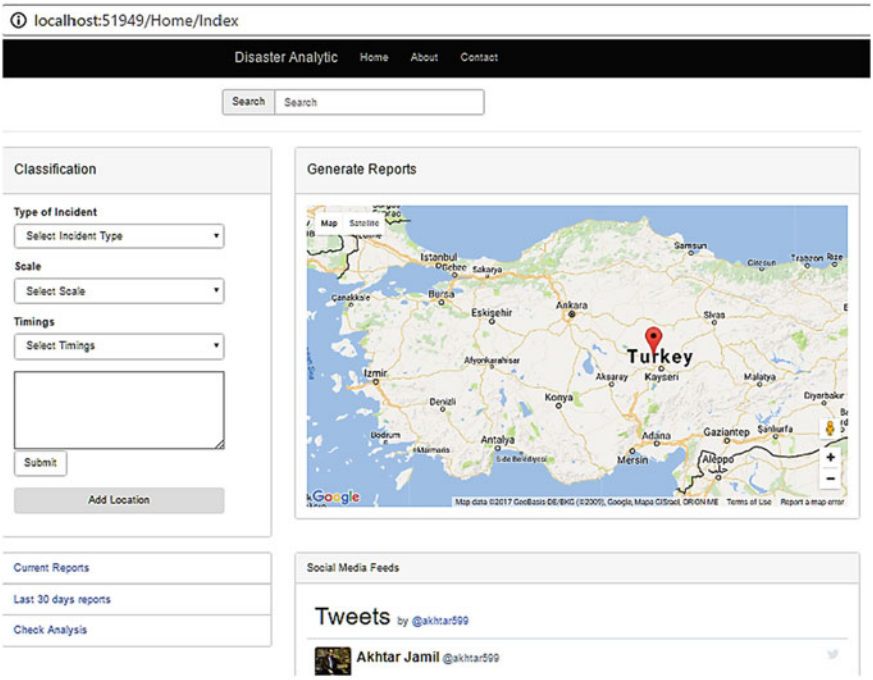

Fig. 2 Designed web API (Disaster analytic)

3.2 Social Media System Component

We are living in an information age where people tend to report, discuss and share the ground facts, observations, and their experiences on different social media forums. The aim of this social media component is to filter and extract the required data gained through crowd-sourcing, and make it beneficial for disaster management (Fig. 3).

The data filtering process is following the semantic methods as it supports filtering and extraction of social media feeds to recognize only the suitable feeds relevant to the incident and provides means to organize information about the incidents for performing real-time analysis. After intensive review of the literature some notable filtering methods are selected and their details are presented in sequence in the following sections.

3.2.1 Data Capture

Whenever an incident is reported the system is triggered to get real time data through social media. Through this framework, the data is captured from different social media sources and translated into an incident profile. The main aim of this

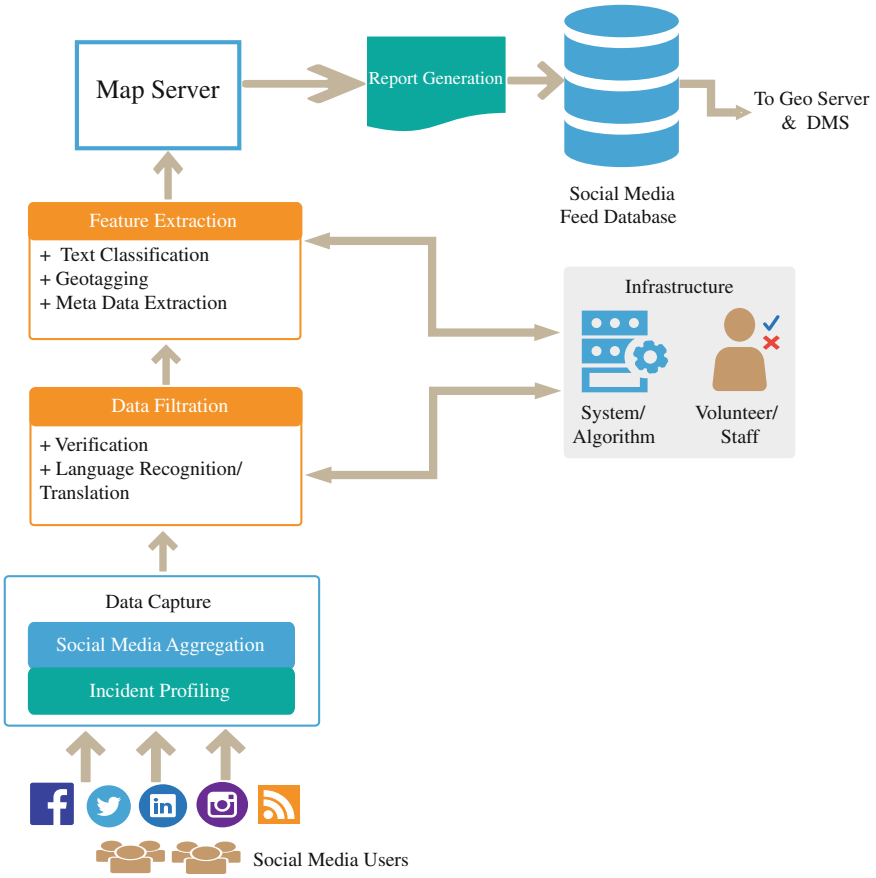

Fig. 3 Proposed social media component

incident profiling is to produce a profile that can provide the valid raw data for data filtering and extraction. "An incident profile is a set of weighted attribute-value pairs that describe the characteristics of the incident" (Abel et al. 2012). These attributes contain data regarding the incidents and might have specific weights to highlight the importance of each attribute according to the type of incident. Incident profiles are kept open to change as it needs to be updated according to the modifications that may occur during the incident. Keeping in view the incident profile designed for the disaster, social media aggregation is performed to capture any type of feed that fulfills the criteria. These feeds are then reported and processed for extracting the useful information for the system. Some common attributes for an incident profiling are mentioned in Table 1.

Table 1 Incident profiling attributes

Incident profiling attributes	Description
Classification	Earthquake, fire, flood, power failure, accident
Address	State, city, street, zip code, coordinates (if extractable)
Descriptive location	Any location name used
Time of occurrence	Time stamp
Scaling	Size or value reported
Image reporting	Any image or video reported
Observations	Comments regarding the incident

3.2.2 Verification

In this phase of data filtration the authenticity and reliability of the crowd-sourced data are checked because of the anonymous way through which it is collected. This remains an issue with social media data and a lot of algorithms are designed to cope with this concern. Privacy concerns are also an issue and need to be addressed with proper procedures. A lot of social media sites have already addressed their users in satisfaction regarding location and feeds in their user agreements. Verification can also be performed manually by volunteers or stuff, but this can be time-consuming and laborious.

3.2.3 Language Recognition

Regions having multiple languages can get social media feeds in different languages and hence affect the source data by adding noise. This issue highlights the need for improved filtering techniques to translate feeds in a common language (i.e., English). Neuro-Linguistic Programming (NLP) techniques can be used to manage multilingual situations.

3.2.4 Metadata Extraction

Metadata Extraction is a vital part of these social feeds as it provides additional information on each feed. The metadata table can contain information about the originator of the feed. To verify the source it may contain the profile picture, number of followers and feeds, location and timestamp when sharing the feed. Such type of Metadata can strengthen the reliability and accuracy of the data provided for processing and decision-making (Stronkman 2011).

Table 2 Text classification attributes

Classification type	Description
Impact area	Map highlighting the potential affected areas
Status	Current status of the incident
Threats	Possible future threats
Related news	Ongoing news about the situation
Casualties and injured	Number of casualties and injuries reported
Image or video reporting	Any images of the incident received
Respond time	Possible time to respond to rescue
Instructions	Any precautions needed

3.2.5 Geotagging

Through a designed API, we can integrate map source data that provides street and building level locations using coordinates. Social media feeds generally describe locations at regional levels if GPS is not activated. As we know, the precise location of the incident plays an important role in disaster management. Therefore, to enable spatial exploration of social media feeds, a geotagging model should be used to display the narrative of the feeds at its exact location on the map. The latitude/longitude coordinates of the user profile can be used to get the desired location.

3.2.6 Text Classification

This phase is the core part of feature extraction as it indicates the extracted contents and provides the required information for reporting. It may contain reports regarding casualties and possible threats and damages about the incident. The classification can further be categorized according to the feeds if the publisher was witness, hearing the news, observing or smelling something. Handcrafted rules are used for the classification that can work in both the attribute-value pairs and the plain words stated in the social media feeds (Abel et al. 2012).

Some common attributes, keeping in mind the filtered social media feeds for classification, are described in Table 2.

4 Conclusion

This study presents a design model for the development of an integrated system consisting of social media crowd-sourced component and a designed web API component through which organized and reliable data can be provided for real-time disaster management. This design-science research demonstrates that the concept of social media crowd-sourcing can effectively be used for real-time disaster

management and tries to aid the theory of making crowd-sourced data as trustworthy as other data sources. The basic theme of this design is to make the unstructured crowd-sourced data processable so that it can be compared and merged with a structured data sources such as a web API. The effectiveness of real-time crowd-sourced disaster management systems has been proven but there are many gaps and challenges in this research domain. The design science to model integrating frameworks plays a key role for providing the basis for interdisciplinary research to be carried out.

References

Abel, F., Hauff, C., Houben, G.J., Stronkman, R. & Tao, K. (2012, April). Twitcident: fighting fire with information from social web streams. In *Proceedings of the 21st International Conference on World Wide Web* (pp. 305–308). ACM.

Adam, N. R., Shafiq, B., & Staffin, R. (2012). Spatial computing and social media in the context of disaster management. *IEEE Intelligent Systems, 27*(6), 90–96.

Ahmed, A. (2011). Use of social media in disaster management. *ICIS 2011 Proceedings.* Paper 16.

Chan, J. C. (2013). 1 The role of social media in crisis preparedness, Response and Recovery.

Cozzolino, A. (2012). *Humanitarian logistics: Cross-sector cooperation in disaster relief management*. Springer Science & Business Media.

Crystal Washington, N. P. (2016). The role of social media during natural disasters. http://crystalwashington.com/the-role-of-social-media-during-natural-disasters/. Retrieved October 13, 2016.

Erskine, M., & Gregg, D. (2012). Utilizing volunteered geographic information to develop a real-time disaster mapping tool: A prototype and research framework. In *CONFIRM 2012 Proceedings*.

Jayaraman, V., Chandrasekhar, M. G., & Rao, U. R. (1997). Managing the natural disasters from space technology inputs. *Acta Astronautica, 40*(2), 291–325.

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59–68.

MacEachren, A. M., Jaiswal, A., Robinson, A. C., Pezanowski, S., Savelyev, A., Mitra, P., et al. (2011, October). Senseplace2: Geotwitter analytics support for situational awareness. In *2011 IEEE Conference on Visual Analytics Science and Technology (VAST),* (pp. 181–190). IEEE.

Ostermann, F. O., & Spinsanti, L. (2011, April). A conceptual workflow for automatically assessing the quality of volunteered geographic information for crisis management. In *Proceedings of AGILE* (pp. 10–14).

Stronkman, R. J. P. (2011). Exploiting Twitter to fulfill information needs during incidents. (Doctoral dissertation, TU Delft, Delft University of Technology).

Turel, O., & Serenko, A. (2012). The benefits and dangers of enjoyment with social networking websites. *European Journal of Information Systems, 21*(5), 512–528.

Veil, S. R., Buehner, T., & Palenchar, M. J. (2011). A work-in-process literature review: Incorporating social media in risk and crisis communication. *Journal of Contingencies and Crisis Management, 19*(2), 110–122.

Velve, D., & Zlateva, P. (2012). Use of social media in natural disaster management. *International Proceedings of Economic Development and Research*, *39*, 41–45.

Yin, J., Lampert, A., Cameron, M., Robinson, B., & Power, R. (2012). Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems, 27*(6), 52–59.