

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

Syed Attique Shah,^a Dursun Zafer Şeker,^b Hande Demirel^b

^a Istanbul Technical University, Faculty of Civil Engineering, Department of Geomatics, Maslak

^b Sarıyer, İstanbul, Turkey; shah@itu.edu.tr, seker@itu.edu.tr, hande.demirel@itu.edu.tr

Abstract: Social Media datasets are playing a vital role to provide information that can support decision making in nearly all domains of technology. It is due to the fact that social media is a quick and economical approach for data collection from public through methods like crowdsourcing. It is already proved by existing research that in case of any 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 components, the first part proposes a framework that provides updated and filtered real time input data for the disaster management system through social media and the second part consists of a designed web user API for a structured and defined real time data input process. This study contributes to the discipline of design science for the information systems domain. 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 to bring this retrieved data to the same level as that of taken through a structured and predefined mechanism of a web API. Both components are designed to a level such that they can potentially collaborate and produce updated information for a disaster management system to carry out accurate and effective

Keywords: Crowd Sourcing, Disaster Management, Design Science, Social Media, Real-time Data, API

1. Introduction

During disastrous incidents (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 the emergency services to act upon. 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 making a vast number of academic studies to research on many aspects of social media in Disaster Management (Veil et al. 2011). Social Media provides its user 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 et al. 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 of information in all of the disaster management life cycle phases and aid in developing a disaster resilient community (Chan and J.C 2014) (Ahmed, A 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 affected communities' requirements. In addition, authorities should be able to observe online com-

munities to detect mounting trends and possible hotspots that can substantiate as indicators for disaster (Chan and J.C 2014). 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; otherwise it can have serious consequences if the information is provided incorrect and late (Erskine et al. 2012) (Ostermann et al. 2011).

Real-time Geospatial Information Systems use social media as crowd sourcing virtual network to map social feeds using geotag metadata with longitude latitude coordinates. GIS systems with all its hardware based sensors can be combined with Social media as the basic theme is to create a compressive source of information to understand the disaster situation accurately so that the emergency responders and the general public can be added to improve overall awareness. (Adam et al. 2012) (Yin et al. 2012). However, there are some challenges also associated with social media data collection methods; such as variable quality of the data, intelligently managing the big volume of social media feeds, requirement of manual checking and verification of the data, accurately Geo-parsing map information and the need to find right balance between time wasting false positives and responsive alters.

Some related work already done shows the practicality of this research are the recognition of frameworks such as Twitcident (Abel et al. 2012) and SensePalce2

(MacEachren et al. 2011). Moreover, the concept of crowd souring is a corresponding research area that is aimed to work as a virtual sensor to collect data from every potential source possible and to support activities from basic mapping (e.g., Open Street Map) to be a source for providing data to the disaster management systems (e.g., CrisisMappers.org, Ushahidi).

A very good example of web API for collecting citizen reactions 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 the 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 et al. 2010).

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

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 an 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). Conducted research

has identified that the use of social media is more increased and even surpassed the use of other conventional communication methods such as fixed phones after a disastrous incident (Velve et al. 2012). Social Media sites like Facebook, Twitter and YouTube can be very handy when tsunamis, earthquakes, floods and other natural disasters strike to collect the real time data.

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 & 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., crowdsourcing)”
- 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, in applying the social media applications to traditional disaster management phases allows us to integrate literature 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, this research is able to 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 considering the method involved for data input. As seen in Figure 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 the 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 process to the disaster management system component. This model emphasis on 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.

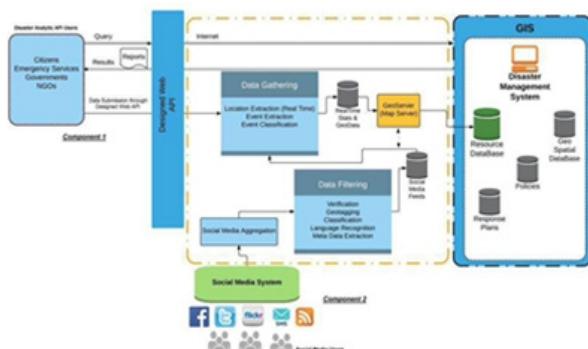


Figure 1. Proposed Research Model

3.1 Designed Web User API Component

This component named as “Disaster Analytic” is a proposed designed web API for receiving information in case of any disaster in a structured data entry template, which covers all the sections such as location, time, scale, description, attachments (photo or video) etc. for data to be actionable. Additionally, Disaster Analytic can also manage queries from the user requesting reports regarding the disaster situations. A rough outline for the interface is proposed in Figure 2. The potential users of this web API can be citizens/volunteers, governments’ officers, emergency services and NGOs who can provide data under defined format and query, search or request for a report through the disaster management system.

The main aim for the design and implementation of this web user 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 involved.

Every user need to be identified and 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 make more user friendly, 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 entry 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 for regarding the disaster.

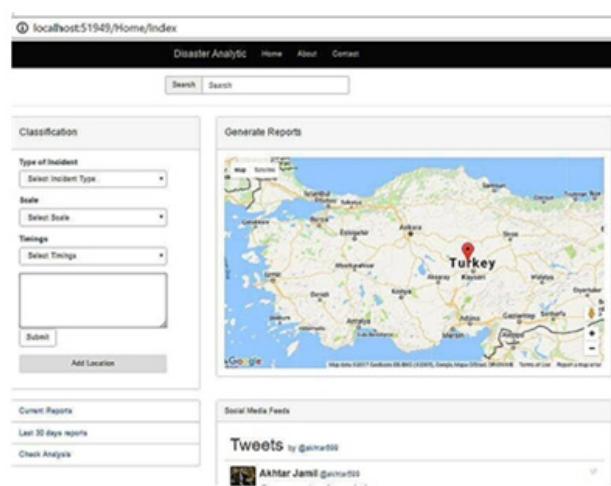


Figure 2. Designed Web API (Disaster Analytic)

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 even if it's regarding a specific disaster. The aim of this Social Media component is to filter and extract the required data gained through crowd sourcing, and make it processable and beneficial for disaster management systems.

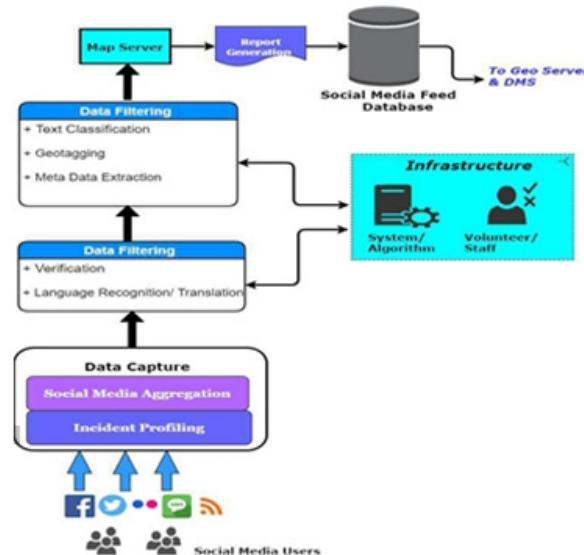


Figure 3. Proposed Social Media Component

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 literature review some notable filtering methods are selected and their details are presented in sequence in the following sections.

3.2.1 Data Capturing

Whenever an incident is reported the system is triggered to get the 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 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 occurring. 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 per-

formed 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.

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

Table 1. Incident Profiling Attributes

3.2.1.1 Verification

In this phase of data filtration the authenticity and reliability of the crowd-sourced data are checked because of the anonymous source 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 concern is also an issue and need to be addressed with proper procedures. A lot of social media sites have already taken their users in satisfaction regarding location and feeds in their user agreements. Verification can also be performed manually by volunteers or stuff, but can be time consuming and laborious.

3.2.1.2 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.

3.2.1.3 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, and furthermore 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 and R.J.P 2011).

3.2.1.4 Geotagging

Through a designed API we can get map source data that provide street and building level locations using coordinates, but on the other hand social media feeds generally describe locations on regional levels if satellite navigation system is not activated. As we know that 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.

3.2.1.5 Text Classification

This phase is the core part of feature extraction as it indicates the extracted contents and provides the required information for reporting or to be used in a disaster management system. It may contain reports regarding casualties, 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 view the filtered social media feeds for classification can be,

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

Table 2. Text Classification Attributes

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 systems. This design-science research demonstrates that the concept of social media crowdsourcing 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 social media crowd sourced data processable so that it can be compared and merged with a structured data sources such as a web API getting first hand data with designed parameters. The effectiveness of real-time crowd sourced

disaster management systems has been proved by new studies and but still this research domain has to cover a lot of research gaps and challenges. The design science to model integrating frameworks plays a key role for providing the basis for interdisciplinary research to be carried out.

5. References

- Abel, F., Hauff, C., Houben, G.J., Stronkman, R. and 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. and Staffin, R., (2012). Spatial computing and social media in the context of disaster management. IEEE Intelligent Systems, 27(6), pp.90-96.
- Ahmed, A., (2011). Use of social media in disaster management. ICIS 2011 Proceedings. Paper 16.
- Chan, J.C., (2014). 1 THE ROLE OF SOCIAL MEDIA IN CRISIS PREPAREDNESS, RESPONSE AND RECOVERY By.
- 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". Web. <http://crystalwashington.com/the-role-of-social-media-during-natural-disasters/>. Accessed 13 Oct. 2016.
- Erskine, M. and Gregg, D., (2012). Utilizing volunteered geographic information to develop a real-time disaster mapping tool: A prototype and research framework. CONFIRM 2012 Proceedings.