2012 11th International Conference on Machine Learning and Applications

Automatic Identification of Crisis-Related Sub-Events using Clustering

Daniela Pohl
Institute of Information Technology
Alpen-Adria-Universität Klagenfurt
Klagenfurt, Austria
daniela@itec.uni-klu.ac.at

Abdelhamid Bouchachia
Smart Technology Research Center
Bournemouth University
Bournemouth, UK
abouchachia@bournemouth.ac.uk

Hermann Hellwagner
Institute of Information Technology
Alpen-Adria-Universität Klagenfurt
Klagenfurt, Austria
hellwagn@itec.uni-klu.ac.at

***Abstract*—Social media are becoming an important instrument for supporting crisis management, due to their broad acceptance and the intensive usage of mobile devices for accessing them. Social platforms facilitate collaboration among the public during a crisis and also support after-the-fact analysis. Thus, social media are useful for the processes of understanding, learning, and decision making. In particular, having information from social networks in a suitable, ideally summarized, form can speed up such processes. The present study relies on Flickr and YouTube as social media and aims at automatically identifying individual sub-events within a crisis situation. The study applies a two-phase clustering approach to detect those sub-events. The first phase uses geo-referenced data to locate a sub-event, while the second phase uses the natural language descriptions of pictures and videos to further identify the ”what-about” of those sub-events. The results show high potential of this social media-based clustering approach for detecting crisis-related sub-events.**

***Keywords*-Clustering, Sub-Event Detection, Crisis Management**

I. Introduction

Crisis management has to deal with different scales of emergencies and a multiplicity of agencies and individuals involved. Usually, managing a crisis can be separated into three phases: preparedness, response, and recovery [1]. Preparedness focuses on actions to increase the awareness and readiness toward a possible disaster. Response focuses on actions to minimize the impact of a disaster. The task of the recovery phase is to improve the situation after the disaster and to reestablish “normal life”. Nevertheless, in all phases collaboration between professional first responders and the public is needed to gain an overview of the situation, organize response actions, and stabilize the situation at hand.

The widespread usage of mobile devices offers a new area of possibilities to support crisis management; e.g., mobile social media applications are such an important development. Studies show that social media is an important source of information for crisis-related tasks [2], [3]. People usually document the (alarming) situations they are involved in, especially if they are of public interest. This is in particular of importance for crisis management in case professional first responders cannot be on-site from the very beginning to gain an overview of the situation, to be informed on newly emerging situations, or for after-the-fact analysis. In addition, this information can be used in crisis management to plan and decide on response actions.

Based on the scale of an emergency and the number of involved parties, it is evident that a huge amount of information is gathered. Analyzing such information manually is a cumbersome or even impossible task. Therefore, in our work we focus on the automatic detection of sub-events (see Section III) based on social media data. As crisis management benefits from visual information, such as images and videos [4], we especially focus on crisis-related sub-event detection from Flickr and YouTube data.

Our current approaches [5], [6] apply clustering for identifying crisis-related sub-events. We focus on offline approaches, which are important for after-the-fact analysis (e.g., analyzing the situation for training purposes). These approaches currently do not take geo-referenced data into account [5]. Thus, in this paper we extend these approaches and explicitly include data containing geo-referenced information (i.e., longitude and latitude coordinates) to improve crisis management.

The present paper is structured as follows. Section 2 gives an overview of related research work. Section 3 describes the idea of sub-events and Section 4 presents our two-phase clustering approach for detecting sub-events (focusing on geo-referenced data items). Section 5 gives an overview of our results and findings. Section 6 summarizes the paper and shows future extensions and directions of the work.

II. Related Work

Social media have received increasing interest over the recent years, due to the utility and the acceptance of social platforms by the public. Most attention is paid to Twitter analysis of different varieties and purposes, e.g., detecting events from Twitter streams can be found in [7], [8], [9].

The application of Twitter to crisis management was studied by different research groups. The work of Terpstra et al. [10] shows a real-time analysis of Twitter data. They consider the geographical distribution of tweets and pre-defined keyword-based filters to conduct the analysis of tweets. In [11] an approach that relies on Twitter during emergencies is presented. The approach keeps track of

978-0-7695-4913-2/12 $26.00 © 2012 IEEE
DOI 10.1109/ICMLA.2012.170

"breaking news" and introduces a feedback loop back to the Twitter System.

Ireson [12] presents an approach for analyzing local forums (belonging to a specific city) to find topics related to a current emergency. Therefore, topic clusters based on the most-frequent single terms are created via hierarchical clustering for subsequent identification of related forum messages (with the same topic).

On the other hand, geo-referenced information is used in various research areas. For instance, in [13] it is used to map tags to specific events/locations. Yin et al. [14] cluster similar Twitter messages; they use geo-data/tagging only for visualization. In contrast, we want to consider geo-data directly during clustering. The work of Jaffe et al. [15] suggests a hierarchical clustering approach for geo-data of pictures, which is the basis for a photo ranking and visualization approach. Within the work of Zhou et al. [16], a density-and-join-based algorithm based on geo-referenced data is introduced to discover personal relevant locations.

In contrast to the state-of-the art approaches, we consider the application of geo-data directly during clustering, where non-geo-referenced feeds in Flickr and YouTube are clustered as well.

Moreover, we are not just focusing on approaches for (keyword-based) filtering a data set given an emergency situation, but we aim at automatically detecting sub-events by applying clustering algorithms using the existing (rare) geo-data information in order to support crisis management. The goal is, therefore, to uncover the "what-about" of the identified sub-event locations.

III. Sub-Event Detection

Crisis situations contain different sub-events (hotspots or mini-crises [17]) on which crisis management has to focus. For example, in different locations, e.g., states, cities, or districts, a crisis may have different severity and consequences, hence a tailored management is needed.

In our case an event is specified by time and location [18] and is defined as the context in which sub-events occur. Hence, the event describes the disaster context (e.g., UK riots 2011), whereas sub-events break down the event in more refined parts, e.g., looting in London. Therefore, sub-event detection aims at identifying potential and dominant threats of a crisis.

Detecting sub-events as soon as possible helps in efficiently managing a crisis situation. Collaboration with those people that have information from the very beginning is vital. Citizens, increasingly, react on crisis situations by gathering, sharing, and updating information about their current environment in social platforms. Although social platforms can contain rumors, it is indeed important to extract information from those platforms, especially for raising awareness and taking corrective actions when needed.

In a previous work [5], we applied clustering algorithms for detecting crisis-related sub-events using social media platforms. We used clustering since this is an unsupervised learning method which does not need labeling and hence additional effort for preparing data. In [5] we do not include geo-referenced data. Instead, a pure textual analysis is performed. However, if there are enough data items with geo-referenced data, this data is a valuable source of information to consider.

Based on the availability of geo-referenced data, a choice can be made between taking an existing approach [5] or applying the presently suggested two-phase clustering approach (see Section IV) relying on geo-referenced data. A short discussion on how to decide which approach should be used is given in Section V.

IV. Two-Phase Clustering

Geo-referenced data - when available - is an important source of information for crisis management. Due to the fact that sub-events are related to specific locations (given by the event), we decided to apply a two-phase clustering approach that relies on longitude and latitude coordinates of existing data items for sub-event detection. In our case, one data item (picture or video) is represented by two parts (see Figure 1): the coordinates and the terms. The coordinates are represented by longitude and latitude values. For the extraction of the terms, we use textual metadata fields (title, description, and tags) belonging to a specific item. The corresponding *term frequency-inverse document frequency (tf-idf)* values [19] are computed. This results in a vector containing term-value pairs for each item which can be used for clustering.

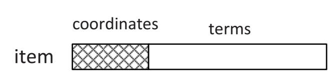

Figure 1. Item (e.g. video or image) representation for clustering

Both parts are important for the clustering approach, but in different stages. The suggested approach places high-priority on geo-references contained in the data (i.e., coordinates). The approach can be summarized as follows:

- **Pre-processing:**
  - Tf-idf analysis of the metadata (title, description, and tags) of data items (i.e., pictures and videos).
  - Localization of data containing geo-referenced information.
- **Two-phase Clustering:**
  - Phase 1: Use of geo-referenced data and calculation of *term-based* centroids with a *Self-Organizing Map (SOM)* [20].
  - Phase 2: Assignment of best fitting data points to the calculated centroids using reassignment and the *cosine distance measure* [19].

- **Visualization:**
  - Visualization of geo-referenced data, clusters, and composite labels of the clusters in a map.
  - Presentation of the five most important items (e.g., pictures or videos) based on the cluster distance.

The two-phase clustering approach is the core of our implementation. In the first phase, all items containing geo-referenced data serve to create cluster centroids which are used in the next step (see Figure 2). Within this step, only the coordinates parts are of importance and, hence, fed into the *SOM* clustering algorithm. This algorithm clusters closely related items into similar clusters [20]. A cluster is represented via an attribute-value vector, in our case containing geo-referenced data.

Based on the resulting clusters, centroids are calculated. All items related to one specific cluster are used to calculate the term and geo-referenced centroid. For the calculation, two measures are checked for their suitability (see Section V-B):

- **Average:** As in state-of-the-art approaches, the arithmetic mean across all values corresponding to the cluster is computed. The resulting vector is used for the next step.
- **Median:** The median of the geographical data is determined, and the terms part of this specific item is used for the next step.

This step results in a term-based representation for each cluster which describes the "what-about" of the identified sub-event represented by this cluster. In addition, the resulting geo-referenced centroid is used for result visualization.

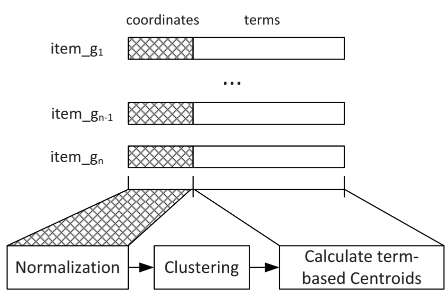

Figure 2. Phase 1 - Clustering with n geo-referenced (g) data items

In the second phase (see Figure 3), the previously calculated term-based centroid or determined median for each cluster is used to assign the remaining data items - without geo-referenced data - to the centroids based on their terms parts. This is performed via the *cosine distance measure* [19], which is very often used in traditional text clustering. The idea behind this step is that items corresponding to the same sub-event are described by people in a similar way, using similar terms for the description (including also textual-based location information correlating with identified clusters in the first phase).

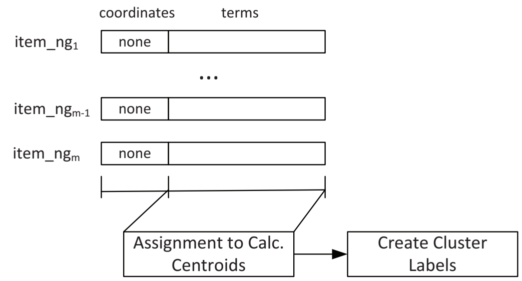

Figure 3. Phase 2 - Assignment of m data items without coordinates (ng)

Based on the smallest cosine distance, the item is assigned to the closest cluster. The original term-based representation of the cluster is not updated after an assignment. Hence, the representation does not change and consequently there is no influence on following assignments. The terms of the assigned items are used to calculate the labels of the clusters found. The resulting labels, the cluster centroids, and the data used to calculate the geo-referenced centroids are displayed in a map (see Figures 4 and 6 and explained in Section V-A).

V. Experiments

The approach was applied to four data sets (see Table I) created by a user-defined keyword-based search, e.g. "UK riots 2011". This is based on a suggested Media Exploration Framework described in [5]. The data sets were chosen such as to represent different crises with different scales, impacts, and origins. Consequently, each of them considers another number of sub-events. The data sets contain metadata information from pictures and videos (from YouTube and Flickr) related to specific crisis situations of the last year.

Table I
DATA SETS

Title	**Period in 2011**	**No. of Images, Videos**	**Geo. Data**
Mississippi Flood	04-19 May	2039, 442	573
Oslo Bombing	22 July	31, 222	13
UK Riots	06-10 Aug.	178, 274	18
Hurricane Irene	23-29 Aug.	455, 700	200

The data of the Mississippi Flood in 2011 shows the impact on several states in the USA when the Mississippi river burst its banks. The Hurricane Irene data shows impacts of the tropical storm on several states of the US east coast as well. The Oslo Bombing data contains information on the terror attack in Norway, with the bomb explosion in Oslo and the shooting at the Utøya island. The UK data gives insights into the UK riots in summer 2011 affecting different cities.

From Table I it can be seen that two data sets, Oslo Bombing and UK Riots, contain less geo-referenced data compared to the others. Independently of the amount of geo-referenced data, we applied the two-phase clustering approach on all four data sets. We used the data representation (coordinates and terms parts) described previously for the two-phase clustering algorithm.

We are describing the major aspects of the applied clustering algorithm for our suggested Media Exploration Framework [5]. For sake of clarity, we are first focusing on the visualization of the results in the next section; the computation is explained in Section V-B.

A. Visualization

For visualizing our data and results we employ OpenStreetMap[1](OSM). This representation of the clustering results helps in conveying the big picture of the crisis and in evaluating the identified sub-events. It offers us the possibility to explore the data and its location information.

Figure 4. Representation of a cluster based on the UK Riots 2011

Figure 4 shows the representation of an example cluster of our implementation. The first line shows the name of the cluster (e.g., Cluster 1). The first number within the square brackets describes how many geo-referenced data items were used to create the cluster in the first phase; the second number represents how many data items without geo-referenced data are assigned to it in the second phase. This line also contains the coordinates (place of the cluster on the map). In addition, geo-referenced data items are displayed as dots in the map, indicating where the items were captured.

The second line shows the labels of the cluster separated by spaces before any of the non-geo-referenced data points are assigned. Labels are represented in their principal form. In addition, similar words (e.g., car and automobile) are grouped together and treated as single words, using the WordNet database [21]. Additionally, the third line shows labels (or changes of labels) when closely related non-geo-referenced data points are assigned to the cluster.

B. Computation of Sub-Events (Centroids)

In our experiments, we consider two different measures for calculating the clusters' centroids for the second phase (see Section IV): median and average. First, we used the median for the term-based centroid due to the fact that we wanted to create a picture of what happens at a specific location. So, the most central data point was used to represent the sub-event at this place best. This turned out to be not fully appropriate because people describe things from different perspectives, use different words, or describe other details as well.

[1]http://www.openstreetmap.org/

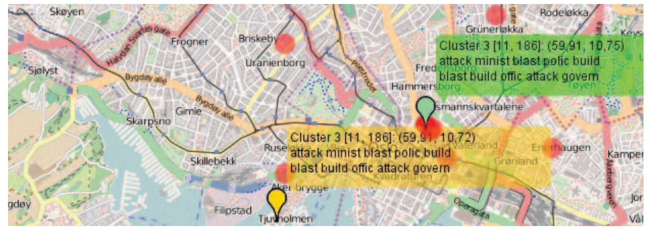

Figure 5. Visualization of median (green/dark) and average (orange/bright) for cluster representation in the map

Hence, the best way to create the starting term-based centroid from the data sets is the traditional average measure, representing an average understanding of all data points - describing the facts of this location - used to create this cluster. We only use the median to represent the cluster (geo-referenced centroid) in the map, as it shows best the location of the concentration of used data items located in this area. In addition, the median is less sensitive to location-based outliers. Figure 5 shows the application of both measures. It can be seen that the median is more suitable for marking clusters on the map and less sensible to outliers.

C. Analysis

The experiments with our data sets show promising results for geo-referenced sub-event detection in emergency cases. We compared the resulting sub-events of the approach with manually extracted sub-events from online documentations of the crises [22], [23], [24] and [25]. Due to space limitations, we describe only two data sets in more detail: Hurricane Irene and Mississippi Flood 2011.

Through the visualization of the clusters and geo-referenced data points in Figure 6 (left), it is possible to follow the course of the Hurricane Irene from the beginning till the end of the threat. Using [22] and [26], the following sub-event locations can be identified: Puerto Rico, Bahamas, South Carolina (Charleston, Georgetown etc.), North Carolina, Virginia (Beach), New Jersey, New York (City, Brooklyn), New Hampshire, North-Eastern Canada. It can be seen that New York was one of the major sub-events during this crisis. When looking in more detail on the map in the area of New York and New Jersey, it can be seen that there is a clear distinction between the areas of New Jersey and New York.

Most problems are related to flooding and power outages see, e.g., Cluster 6 in the map Cluster 12 identifies the flooding in the area of New Jersey. A closer look of the map shows identified sub-events related to Virginia and Carolina.

There are also clusters where it seems to make sense to have a more detailed view of the terms, considering only

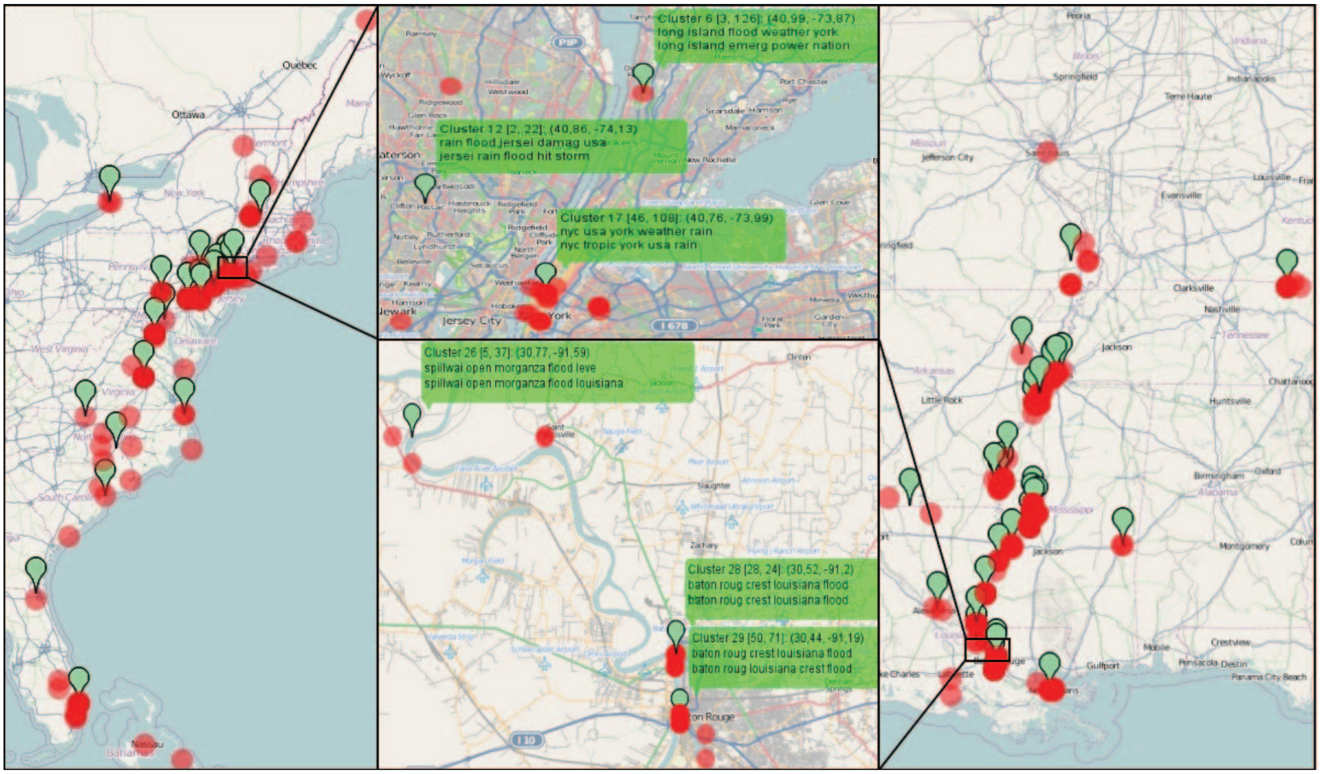

Figure 6. Left: Hurricane Irene with sub-event samples (area: New York, New Jersey); Right: Mississippi flood with sub-event samples (Baton Rouge)

data items of this cluster. Hence, a more detailed text-based analysis is necessary to strengthen the results. Despite the limited amount of geo-referenced data for the Hurricane Irene data set, the two-phase clustering approach gives some valuable hints of major causes and where to look next.

For the Mississippi flood, it is also possible to identify the major sub-events due to flooding. Based on [23], the following sub-events can be identified: Arkansas (along White River), Tennessee (Memphis), Mississippi (Vicksburg, Greenville, Yazoo), Louisiana (Morganza Spillway opened, Baton Rouge, St Martin, Morgan City, New Orleans). Via a closer look of the southern part of the affected area, the approach clearly distinguishes areas of Baton Rouge and New Orleans (see Figure 6 - right). There is also a clear picture that the Morganza Spillway was opened to flood the Atchafalaya Basin to prevent New Orleans and Baton Rouge from major flooding (cluster annotations: spillwai, open, morganza, etc.). Additionally, it can also be seen that Memphis and towns like Greenville and Vicksburg are impacted by the Mississippi river flood.

Note that for the Mississippi Flood much more geo-referenced data is available compared to, e.g., UK riots. Applying the approach to data sets with less geo-referenced information shows surprisingly good performance. For example, for the UK Riots data set (see [25]), sub-events in London, Manchester, and Birmingham are identified. Detailed labels on fire, looting, presence of police, or related information were found.

We also apply the approach to the Oslo Bombing data set. It shows all major sub-events (see [24]): the bomb explosion in Oslo and the shooting at Utøya island. A very interesting insight can be demonstrated in Figure 5. It shows the cluster related to the bomb explosion in Oslo. The initially created center contains concepts like: attack, minist, blast, polic, and build. When assigning other non-geo-referenced data items to the cluster, which are closely related, additional important concepts like govern (government) were added. In addition, also the order of the labels changes. It shows the most frequently used concepts for describing the situation. This is additional information that helps the crisis management command to gain more insight into the situation. The effect can be observed in other clusters and in other data sets as well. The location of the shooting is identified since there is a data item annotated with the coordinates of the Utøya island. Due to the nature of the attack, there are only very few data items related to the shooting on the island.

The results can be used to understand the situation and find out the most relevant crisis-related sub-events. The more geo-referenced information is available, the more detailed are the results. Nevertheless, the approach shows quite good performance for fewer geo-annotated data items, as long as these data items contain the most important sub-events.

VI. Conclusion

In this paper, we utilize a clustering approach for crisis-related sub-event detection based on geo-referenced data. Due to the importance of visual data of the situation in crisis management, we applied the approach to data sets obtained from Flickr and YouTube.

We suggested a two-phase clustering approach that prioritizes geo-referenced data items. Compared to existing approaches, the proposed technique does not purely filter out terms; rather, it contains an intelligent mechanism for detecting important sub-events. These sub-events can be used to support crisis management or after-the-fact analysis to accurately perceive the situation. Experiments performed on our data sets show promising results in applying this approach. They also show good performance for data sets with fewer geo-referenced annotations.

In the future, we aim to extend this work by including time-based aspects into a real-time streaming approach for sub-event detection. Another possible extension is to identify concepts (e.g., fire, people, explosions etc. as in [27]) from visual content for the inclusion in the sub-event detection process. Additionally, we also plan a pre-processing stage where picture and video content analysis is used to filter out duplicate or similar data items.

Acknowledgment

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement n°261817 and was partly performed in the Lakeside Labs research cluster at Alpen-Adria-Universität Klagenfurt.

References

[1] S. R. Hiltz, B. van de Walle, and M. Turoff, "The Domain of Emergency Management Information," in *Information Systems for Emergency Management*. Armonk, New York: B. van de Walle, M. Truoff and S. R. Hiltz, 2010, vol. 16, pp. 3–19.

[2] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, "Microblogging During Two Natural Hazards Events: What Twitter May Contribute to Situational Awareness," in *Proc. of the 28th Inter. Conf. on Human Factors in Computing Systems*, ser. CHI '10. New York, NY, USA: ACM, 2010, pp. 1079–1088.

[3] S. Liu, L. Palen, J. Sutton, A. Hughes, and S. Vieweg, "In Search of the Bigger Picture: The Emergent Role of On-Line Photo-Sharing in Times of Disaster," in *Proc. of the Information Systems for Crisis Response and Management Conf.*, 2008.

[4] F. Bergstrand and J. Landgren, "Information Sharing Using Live Video in Emergency Response Work," in *Proc. of the 6th Inter. Conf. on Information Systems for Crisis Response and Management*, 2009.

[5] D. Pohl, A. Bouchachia, and H. Hellwagner, "Supporting Crisis Management via Sub-Event Detection in Social Networks," in *Inter. Conf. on Collaboration Technologies and Infrastructures*, Toulouse, France, 2012.

[6] ——, "Automatic Sub-Event Detection in Emergency Management Using Social Media," in *First Inter. Workshop on Social Web for Disaster Management (SWDM), In conjunction with WWW'12*, Lyon, France, 2012.

[7] A. Marcus, M. S. Bernstein, O. Badar, D. R. Karger, S. Madden, and R. C. Miller, "Twitinfo: Aggregating and Visualizing Microblogs for Event Exploration," in *Proc. of Annual Conf. on Human Factors in Computing Systems*. New York, USA: ACM, 2011, pp. 227–236.

[8] M. Mathioudakis and N. Koudas, "TwitterMonitor: Trend Detection over the Twitter Stream," in *Proc. of the Inter. Conf. on Management of Data*, ser. SIGMOD '10. New York, USA: ACM, 2010, pp. 1155–1158.

[9] H. Becker, M. Naaman, and L. Gravano, "Beyond Trending Topics: Real-World Event Identification on Twitter," in *Proc. of the Fifth Inter. AAAI Conf. on Weblogs and Social Media (ICWSM'11)*. AAAI, 2011, pp. 438–441.

[10] T. Terpstra, A. de Vries, R. Stronkman, and G. L. Paradies, "Towards a Realtime Twitter Analysis During Crises for Operational Crisis Management," in *Proc. of the 9th Inter. ISCRAM Conf.*, Vancouver, Canada, April 2012.

[11] J. Rogstadius, V. Kostakos, J. Laredo, and M. Vukovic, "Towards Real-Time Emergency Response using Crowd Supported Analysis of Social Media," in *Proc. CHI Workshop on Crowdsourcing and Human Computation: Systems, Studies and Platforms*, 2011.

[12] N. Ireson, "Local community situational awareness during an emergency," in *3rd IEEE Inter. Conf. on Digital Ecosystems and Technologies (DEST '09)*, june 2009, pp. 49 –54.

[13] T. Rattenbury, N. Good, and M. Naaman, "Towards Automatic Extraction of Event and Place Semantics from Flickr Tags," in *Proc. of the 30th Annual Inter. ACM SIGIR Conf. on Research and Development in Information Retrieval*, ser. SIGIR '07. New York, USA: ACM, 2007, pp. 103–110.

[14] J. Yin, A. Lampert, M. Cameron, B. Robinson, and R. Power, "Using Social Media to Enhance Emergency Situation Awareness," *IEEE Intelligent Systems*, vol. PP, no. 99, p. 1, 2012.

[15] A. Jaffe, M. Naaman, T. Tassa, and M. Davis, "Generating Summaries and Visualization for Large Collections of Geo-Referenced Photographs," in *Proc. of the 8th ACM Inter. Workshop on Multimedia Information Retrieval*. New York, NY, USA: ACM, 2006, pp. 89–98.

[16] C. Zhou, D. Frankowski, P. Ludford, S. Shekhar, and L. Terveen, "Discovering personally meaningful places: An interactive clustering approach," *ACM Trans. Inf. Syst.*, vol. 25, no. 3, Jul. 2007.

[17] D. Yates and S. Paquette, "Emergency Knowledge Management and Social Media Technologies: A Case Study of the 2010 Haitian Earthquake," *Inter. Journal of Information Management*, vol. 31, no. 1, pp. 6–13, 2011.

[18] Y. Yang, J. G. Carbonell, R. D. Brown, T. Pierce, B. T. Archibald, and X. Liu, "Learning Approaches for Detecting and Tracking News Events," *IEEE Intelligent Systems*, vol. 14, no. 4, pp. 32–43, Jul. 1999.

[19] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to Information Retrieval*. Cambridge University Press, 2008.

[20] T. Kohonen, "The Self-Organizing Map," *Proc. of the IEEE*, vol. 78, no. 9, pp. 1464 –1480, Sept. 1990.

[21] C. Fellbaum, Ed., *WordNet: An Electronic Lexical Database*. MIT Press, 1998.

[22] L. A. Avila and J. Cangialosi. (2012, Aug.) National Hurricane Center. [Online]. Available: http://www.nhc.noaa.gov/data/tcr/AL092011_Irene.pdf

[23] Public Health Emergency. (2012, Aug.) 2011 Mississippi River Flooding. [Online]. Available: http://www.phe.gov/emergency/news/sitreps/Pages/2011mississippi-flooding.aspx

[24] BBC News Europe. (2012, Aug.) Timeline: How Norway's Terror Attacks Unfolded. [Online]. Available: http://www.bbc.co.uk/news/world-europe-14260297

[25] ——. (2012, Aug.) England Riots: Maps and Timeline. [Online]. Available: http://www.bbc.co.uk/news/uk-14436499

[26] Public Health Emergency. (2012, Aug.) Hurricane Irene 2011. [Online]. Available: http://www.phe.gov/emergency/news/sitreps/Pages/irene-2011.aspx

[27] A. Natsev, J. Tešić, L. Xie, R. Yan, and J. R. Smith, "IBM Multimedia Search and Retrieval System," in *Proc. of the 6th ACM Inter. Conf. on Image and Video Retrieval*, ser. CIVR '07. New York, NY, USA: ACM, 2007, pp. 645–646.