

# *National Security and Social Media Monitoring*

## *A Presentation of the EMOTIVE and Related Systems*

Martin D. Sykora\*, Thomas W. Jackson, Ann O'Brien, Suzanne Elayan

Information Science Department  
Loughborough University  
Loughborough, United Kingdom  
\*M.D.Sykora@lboro.ac.uk

**Abstract**—Today social media streams, such as Twitter, represent vast amounts of ‘real-time’ daily streaming data. Topics on these streams cover every range of human communication, ranging from banal banter, to serious reactions to events and information sharing regarding any imaginable product, item or entity. It has now become the norm for publicly visible events to break news over social media streams first, and only then followed by main stream media picking up on the news. It has been suggested in literature that social-media are a valid, valuable and effective real-time tool for gauging public subjective reactions to events and entities. Due to the vast big-data that is generated on a daily basis on social media streams, monitoring and gauging public reactions has to be automated and most of all scalable – i.e. human, expert monitoring is generally unfeasible. In this paper the EMOTIVE system, a project funded jointly by the DSTL (Defence Science and Technology Laboratory) and EPSRC, which focuses on monitoring fine-grained emotional responses relating to events of national security importance, will be presented. Similar systems for monitoring national security events are also presented and the primary traits of such national security social media monitoring systems are introduced and discussed.

**Keywords**—social media monitoring; national security; information retrieval; natural language processing; Twitter

### I. INTRODUCTION

User generated content comes in many forms; however, vast quantities of informal text based messages shared by users of social media applications, such as on micro-blogging applications, present an exciting and potentially highly valuable live stream of opinionated, emotional and informative datasets. For instance, the crisis response community has used Twitter and similar social media with some success to help deal with crisis management during conflicts, natural or manmade disasters [1, 2]. Numerous businesses are now monitoring social media in order to gauge their brand’s popularity and that of their products [3]. Governments and law enforcement organisations are now also actively seeking ways to monitor and anticipate violence, and to analyse public response to various events of national security importance [4]. Reference [5] and [6] point out the importance of gauging the public response to terrorism events from social media, and specifically highlight the importance of automatic sentiment detection in Tweets. There is also a need to geo-locate tweets and organise the vast social streams in order to provide the most relevant

output in a user interface for monitoring tasks [7]. In [8] social media monitoring was presented as a useful new source for situational awareness during terrorism events; however, the authors were careful to point out the ethical issues with sharing on Twitter, as oversharing during times of terror, as the authors have shown, can aid the attackers agenda. In this paper we review several recent social-media monitoring systems from the literature, in the security domain. The primary elements of these systems are briefly presented, a gap in sentiment analysis capabilities is highlighted and our own EMOTIVE system, a project funded by DSTL (United Kingdom based *Defence Science and Technology Laboratory*) and EPSRC, which predominantly focused on fine-grained emotion extraction, is briefly introduced within this context.

The remainder of the paper is structured as follows. Section 2 introduces some background considerations for social stream monitoring systems. Section 3 describes a compilation of the essential features expected from most national security monitoring systems, and subsequently five specific monitoring systems from literature are reviewed. The EMOTIVE system, developed by the authors is introduced in section 4 and the paper is concluded in section 5.

### II. BACKGROUND

Social media is often the first source to break the news, only then followed by main stream news outlets. This has been the case, as early as the 2008 Mumbai attacks, where individuals on location broke the news via Twitter [9], or in the July 2009 Jakarta bombings, where Twitter broke the news [10]. Interestingly, it was found that even earthquakes, ranging from seismic intensity scale 3 or more, were reported quicker by Twitter users as opposed to the relevant Japanese agencies [11]. However, social streams such as Twitter were also found to be a great source for gauging or polling public opinion during relatively event less time-periods [12]. Unfortunately there is a very wide range of different content and styles of messages communicated on Twitter [13], which requires selective interpretation of the messages. Nevertheless the potential value of social media streams is quite evident from the highlighted literature above. However, there is a need for automated tools and techniques to monitor and help analyse the vast social-media content [14]. Automated analysis and monitoring is necessary in order to deal with the enormous amounts of social-media content being generated every day, as

purely human based monitoring is not feasible. As opposed to product, services and business related social media monitoring applications [15], throughout this paper we focus on the needs of monitoring and analysis applications relevant to national security, which are understood to encompass unplanned natural / manmade disasters, terrorist attacks, violent incidents, and scheduled expected events with some national significance.

### III. PRIMARY ELEMENTS OF A MONITORING SYSTEM

In order for a system to facilitate monitoring of a certain event or entity of interest, efficient extraction of messages, geo-location, emotion evaluation, clustering and organization of the tweet messages, and an intuitive user-interface are necessary [16, 14]. Specifically in the field of terrorism informatics [5] suggested a framework for tweet filtering and extraction, followed by sentiment detection, exploration of demographics and finally, reporting and presentation of the enriched tweet datasets. We argue, based on prior work [16, 17, 7, 18, 19, 20], that a system design which integrates and unifies different techniques to facilitate effective monitoring in various situations should hence be composed of at least the following system steps:

1. Keyword / Keyphrase monitoring or first event detection, filtering and extraction
2. Accurate geo-location detection
3. Emotion detection and evaluation
4. Tone of tweet message detection, further semantic enrichment and organisation
5. User-interface visualisation

The idea behind these system steps is that in item 1, sparse text messages are monitored and retrieved, and some spam is filtered out at this early stage. Some approaches require manual input of keywords to monitor, other infer them automatically. For instance in [5] keywords were based on automated named entity recognition of regularly re-checked trending twitter topics. On the other hand in [21] monitoring was performed on an unfiltered stream of tweets (or tweet firehose), using first story detection. Essentially tweets that are similar are grouped together, using a kind of nearest neighbor or clustering techniques; however, to allow scalability for “Big Data”, Locality-Sensitive Hashing was used, and seems to be a very feasible solution to the problem. In steps 2, 3 and 4 further analysis substantially enriches the retrieved text messages by providing context through extracting location details, communicated emotions, and various features from the tone of the messages, which can be used to help organize the tweets. These three steps can also be understood to simply be steps in automated semantic enrichment [22]. The amount of semantic enrichment may vary from system to system, but for instance in the EMOTIVE project (section IV) the dominant focus was step 3, i.e. fine-grained emotion detection, although some geo-location and simple tone detection was also performed in EMOTIVE. The work in [23] provides further support for why this kind of information and the resulting situational awareness is very useful in security incidents monitoring. In item 5 the visualisation and possibly a faceted user-interface allows to explore the enriched sparse text message data (items; 2, 3, 4).

Thanks to item 2, map based visualisations are largely possible, with the possibility of overlaying various dimensions of information over a map interface. All system steps discussed must be adapted to (near) real-time monitoring scenarios, and require the use of efficient algorithms and data-structures to handle big-data, irrespective of parallelization architectures and scalable frameworks that could speed up the techniques in their own right.

In the rest of this section a brief overview of five systems for monitoring crisis and national security or emergency situations, using Twitter are presented. The systems are relatively recent efforts and some of them are still being developed. **Crisis Tracker** [16] organises Twitter messages into stories and categorises them by type of incidents and displays them on a geographic map (geo-map), although much of its functionality is supported by crowdsourcing, i.e. human users annotating content, rather than an automated algorithm performing the tasks. It is currently being developed for uses in the crisis management and monitoring community, as part of a PhD thesis, and an experimental, live deployment run of the system has completed in September 2012, in which crisis mapping volunteers used the system to track incidents in the Syrian conflict. **Crisees** [17] is a prototype system for aggregating social media content for crises monitoring, and its basic capabilities are extraction and monitoring of multiple events from social media with a geo-map visualisation, with a basic (lexicon based) emotion (sentiment polarity) classification. **SensePlace 2** [7] is being designed for the crisis management community, and the system essentially allows to search by place (geo-map), theme and time. Detection of themes was aided by named entity recognition (NER) from tweet content. Location mentions from messages were also extracted to place a non geo-tagged tweet geographically. **Swiftriver** [18, 19] is Ushahidi’s semi-automatic linguistic text analysis sub-system (i.e. a large portion of the processing is conducted by human operators), but it is still in development (currently in beta) without any relevant published evaluation of the system, other than some initial material and open-sourced code<sup>1</sup>. SiLCC is the main module for performing some of the automated linguistic text analysis; however, the interface itself is mostly concerned with establishing veracity of messages and sources, with the possibility for Swiftriver to be integrated with an Ushahidi instance. **Twitcident** [20] is developed in co-operation with Netherland’s emergency services, to provide contextual social media information for significant reported emergency incidents, which may have a public dimension and hence may be of interest for monitoring purposes, as people on Twitter may provide information of causalities, risks or damages. The system’s main focus is not necessarily a geo-map display (although available in a secondary view) but rather an event-specific display of Tweets that is searchable by a faceted user-interface, with further statistics on tweets matching the search criteria. Table I highlights the main features of the above mentioned systems, where the features of each system are grouped under the following headings; 1-extraction of messages, 2-geo-location, 3-emotion evaluation,

<sup>1</sup> Code available on GitHub – <https://github.com/ushahidi/silcc>; Swiftriver sites – <http://swiftly.org/>;

#### 4-clustering and organisation of tweet messages and the 5-user-interface.

TABLE I. FIVE TWITTER MONITORING SYSTEMS (PUBLISHED IN ACADEMIC LITERATURE)

System	Features
CrisisTracker [16]	-EXTRACTION: <i>Manual search-terms</i> and their subsequent monitoring, and automated <i>breaking-news detection</i> -LOCATION: Geo-tagged Tweets only; consideration is given to users on the ground – i.e. on location users are highlighted by users (Crowdsourced). -EMOTION: n/a -ORGANISATION: Automated clustering of messages into Stories. Crowdsourced message / entity annotation by humans. -UI: Geo-Map interface, faceted search, colour coded message /event types, and main-detail view of stories-messages.
Crisees [17]	-EXTRACTION: Manual search-terms entry and subsequent monitoring -LOCATION: Geo-tagged – only lat/long tagged Tweets -EMOTION: Lexicon based sentiment word matching (+)ve/(-)ve -ORGANISATION: Simple list -UI: Geo-Map interface, individual tweet message display – colour coded by emotion, and display of images.
SensePlace2 [7]	-EXTRACTION: Manual search-terms entry and subsequent monitoring -LOCATION: Geo-tagged (lat/long) tagged Tweets, and inferring location from user profiles or tweet messages themselves (using Geonames), and the TtT standard is mentioned, although not clear whether implemented. -EMOTION: n/a -ORGANISATION: Based on the principle of Time, Theme and Location. -UI: A structured survey of professionals in crisis response was conducted to guide the UI design.
Swiftriver (Ushahidi) [18, 19]	-EXTRACTION: - -LOCATION: Gazetteer lists of place-name locations -EMOTION: n/a -ORGANISATION: Trustworthy sources are determined by Trustworthiness ratings, and context is added by tagging. -UI: Ushahidi’s UI consists of a map as main search interface of events, possible to select type of events.
Twitcident [20]	-EXTRACTION: Search-terms based on an incident profile (e.g. locations, persons) from emergency broadcasting services (incidents requiring public emergency services to take action). Incident profiles (search-terms) are continuously updated to adapt to topic changes. -LOCATION: Geo-tagged (lat/long) and location gazetteer lists with NER (Named Entity Recognition) and stated Twitter profile location -EMOTION: n/a -ORGANISATION: Messages are classified into reports about casualties, damages or risks; Classification of tone, i.e. Question or Answers; Explicit message filtering is employed -UI: faceted search interface

#### IV. THE EMOTIVE SYSTEM

EMOTIVE (Extracting the Meaning of Terse Information in a Geo-Visualisation of Emotion) is a project within which social media monitoring techniques targeted for the national security domain were developed, as part of a prototype system. The dominant focus of the work was to develop a novel ontology engineering and natural language processing (NLP) approach for fine-grained emotion extraction. A wide spectrum of explicit emotions are expressed on Twitter; however, existing approaches do not capture this richness of emotional expression [24, 25], hence our contribution in this area. Table I also highlights a gap in the emotion analysis capabilities of existing systems. Geo-location techniques were also explored in EMOTIVE, as only 1%-1.6% of tweets tend to contain explicit geo-coordinate meta-data, rendering map-based visualisations impractical. The techniques within the system were specifically targeted at the United Kingdom, since numerous instances of localised slang are contained in the

ontology. With respect to the system features discussed so far EMOTIVE has tackled the primary features as follows. **Extraction:** Manual search-terms entry and subsequent monitoring, although the system provides keyword suggestions to the analyst. **Location:** Geo-tagged, Tweak the Tweet (TtT) standard recognition, location check-in services and location gazetteer lookup. **Emotion:** Fine-grained emotion recognition from sparse text, specifically; anger, confusion, disgust, fear, happiness, sadness, shame and surprise; and emotional expression strength score rating for each identified emotion. **Organisation:** Tweets are organised based on time, location and emotion, where semantically related emotions and synonyms are dealt with. **UI:** An interactive and dynamic geo-map based visualisation, overlaid with topic and emotion information, and employing colour-coded tag-clouds.

Due to space constraints only some EMOTIVE details are briefly presented below, and a full description, with example ontology terms and a detailed evaluation and benchmarking results are available in [26].

Keyword suggestions for monitoring were based on most frequently encountered hashtags in a set of tweets, and a uni-gram language model (trained on a subset of the 1-trillion Google n-gram corpus) segmented multi word hashtags into its constituent words for ease of viewing, e.g. #brutalviolentrioting → brutal, violent, rioting. Tweets are also indirectly filtered to only the ones that contain explicit expressions of emotion, which seems to be relatively effective in filtering out obvious spam and hijacked hashtag tweets.

Several approaches were explored for geo-location of non geo-tagged tweets. The GeoNames database of cities and towns in the UK was used as a gazetteer list to look up any place-name mentions. TtT standard’s geo-location convention was found to be less helpful as it was practically not used in the Twitter dataset that was analysed (see next paragraph for information on the dataset used). However geo-inference from related geo-services such as Foursquare, similar to work in [27], was more frequent, yet the number of users in our dataset was not considerably large (i.e. less than 1%), and hence the approach did not yield a significant improvement.

The main aspect of EMOTIVE was the development of an ontology which semantically represents eight basic, fine-grained emotions. We have based our emotions on the widely accepted Ekman’s basic emotions. Others, e.g. [28, 29] suggested extending the set of emotions used by Ekman, for various reasons. In our ontology we introduced confusion and shame in addition to Ekman’s six emotions, as we found shame to be a highly prevalent emotion on Twitter, and confusion seemed to be a relevant emotional indicator of situational awareness, which as argued in [8] is important during terrorist incidents. The EMOTIVE ontology contains well over 300 emotional terms. In addition to emotions, it covers negations, intensifiers, conjunctions, interjections, and contains information on the perceived strength (also known as activation level) of individual emotions, whether individual terms or phrases are slang or used in standard English and their associated POS (Parts-of-Speech) tags, where this aids to resolve ambiguity. More details on our custom POS tagger and the emotions extraction evaluation and benchmarking is

available in [26]. The study of language containing emotional expressions, required for the construction of the ontology, was performed by an English language and literature PhD level research associate, with training in linguistics and discourse analysis, during a three month period. In order to develop the ontology the RA sifted through around 600MB of cleaned Tweets on 63 different UK-specific search-terms / events based Twitter datasets (collected using Twitter Search API). This manual analysis focused on identifying commonly used explicit expressions of emotion.

Finally the organisation of tweets was performed mostly at the UI stage, as is the case with other systems [20] where tweets are viewed in a faceted interface. Similar and extreme emotions, and tweets from specific locations are grouped together in a faceted UI. These semantically enriched tweets are also displayed as points on an interactive geo-map.

## V. CONCLUSION

Several systems that show substantial promise in the security and crisis monitoring area were discussed and an attempt at identifying significant features shared by some of these tools was made. It was shown that existing systems and prototypes have not focused on emotion analysis. Given the gap in sentiment analysis capabilities, this paper presented the EMOTIVE system as a national security and emergency monitoring tool for emotional public response analysis to help automate tedious and inefficient processes in current monitoring tasks. Especially, fine-grained emotion monitoring.

## REFERENCES

- [1] S. Kumar, G. Barbier, M. A. Abbasi, H. Liu, "TweetTracker: An Analysis Tool for Humanitarian and Disaster Relief," Proceedings of the Fifth International Conference on Weblogs and Social, 2011.
- [2] A. Tapia, K. Bajpai, B. Jansen, J. Yen and L. Giles, "Seeking the trustworthy tweet: Can microblogged data fit the information needs of disaster response and humanitarian relief organizations," Proceedings of the 8th International ISCRAM Conference, 2011.
- [3] A. Tollinen, J. Jarvinen and H. Karjaluo, "Social Media Monitoring in the industrial Business to Business Sector," World Journal 2 (4), pp. 65-76, 2012.
- [4] D. Preotiuc-Pietro, S. Samangoei, T. Cohn, N. Gibbins and M. Niranjani, "Trendminer: An Architecture for Real Time Analysis of Social Media Text," Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media, 2012.
- [5] M. Cheong and V. C. S. Lee, "A microblogging-based approach to terrorism informatics: Exploration and chronicling civilian sentiment and response to terrorism events via Twitter," Journal of Information Systems Frontiers – Springer 13 (1), pp. 45-59, 2011.
- [6] K. Glass and R. Colbaugh, "Estimating the sentiment of social media content for security informatics applications," Security Informatics Vol. 1(1), pp. 1-16, 2012.
- [7] A. MacEachren, A. Jaiswal, A. Robinson, S. Pezanowski, A. Savelyev, P. Mitra, "SensePlace2: GeoTwitter analytics support for situational awareness," Proceedings of the Visual Analytics Science and Technology (VAST) IEEE Conference, 2011.
- [8] O. Oh, M. Agrawal and H. Rao, "Information control and terrorism: Tracking the Mumbai terrorist attack through twitter," Information Systems Frontiers Vol. 13, pp. 33-43, 2011.
- [9] C. Beaumont, "Mumbai attacks: Twitter and Flickr used to break news," The Daily Telegraph, 27th Nov. 2008.
- [10] P. Cashmore, "Mashable: Jakarta bombings - Twitter user first on the scene," Available from <http://mashable.com/2009/07/16/jakarta-bombings-twitter/>, 16th Jul. 2009.
- [11] T. Sakaki, M. Okazaki and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," Proceedings of the 19th international conference on World Wide Web, 2010.
- [12] B. O'Connor, R. Balasubramanyan, B. Routledge and N. Smith, "From Tweets to Polls: LinkingText Sentiment to Public Opinion Time Series," Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, 2010.
- [13] P. Andre, M. Bernstein and K. Luther, "Who gives a tweet?: evaluating microblog content value," Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work, 2012.
- [14] F. Johansson, J. Brynielsson and M. Quijano, "Estimating Citizen Alertness in Crises Using Social Media Monitoring and Analysis," Proceedings of the European Intelligence and Security Informatics Conference (EISIC), 2012.
- [15] A. Tollinen, J. Jarvinen and H. Karjaluo, "Social Media Monitoring in the industrial Business to Business Sector," World Journal, 2 (4), pp. 65-76, 2012.
- [16] J. Rogstadius, V. Kostakos, J. Laredo and M. Vukovic, "A real-time social media aggregation tool: Reflections from five large-scale events," Proceedings of the European Conference on Computer-Supported Cooperative Work (ECSCW), 2011.
- [17] D. Maxwell, S. Raue, L. Azzopardi, C. Johnson and S. Oates, "Crises: Real-time monitoring of social media streams to support crisis management," Advances in Information Retrieval, 2012.
- [18] Ushahidi, "Explaining Swift River," Available from <http://blog.ushahidi.com/index.php/2009/04/09/explaining-swift-river/>, Last Accessed 1st October 2012, 2009.
- [19] Ushahidi, "About: Introduction and History," Available from <http://www.ushahidi.com/about>, Last Accessed 1st August 2012, 2012.
- [20] F. Abel, C. Hauff, G. Houben, R. Stronkman and K. Tao, "Semantics + Filtering + Search = Twitcident Exploring Information in Social Web Streams," Proceedings of the 23rd ACM International Conference on Hypertext and Social Media, 2012.
- [21] S. Petrović, M. Osborne and V. Lavrenko, "Streaming first story detection with application to Twitter," Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, 2010.
- [22] S. Bird, E. Klein and E. Loper, "Natural Language Processing with Python," O'Reilly Publishers, 2009.
- [23] H. Chen, Y. Zhou, E. Reid and C. Larson, 2010. "Special issue on terrorism informatics," Information Systems Frontiers, 2010.
- [24] M. Thelwall, K. Buckley and G. Paltoglou, "Sentiment Strength Detection for the Social Web," Journal of the American Society for Information Science and Technology 63, pp. 163-173, 2012.
- [25] M. Choudhury and S. Counts, "The Nature of Emotional Expression in Social Media: Measurement, Inference and Utility," Technical Report: Microsoft, 2012.
- [26] M. Sykora, T. W. Jackson, A. O'Brien and S. Elayan, "EMOTIVE Ontology: Extracting fine-grained emotions from terse, informal messages" IADIS Intelligent Systems and Agents Conference, 2013.
- [27] Y. Ikawa, M. Enoki and M. Tatsubori, "Location inference using microblog messages," Proceedings of the 21st international conference companion on World Wide Web, pp. 687-690, 2012.
- [28] R. Plutchik, "Emotion: A Psychoevolutionary Synthesis," Longman Higher Education, 1980.
- [29] J. Sabini and M. Silver, "Ekman's basic emotions: Why not love and jealousy?," Cognition and Emotion 19 (5), pp. 693-712, 2005.
- [30] K. Gimpel, N. Schneider, B. O'Connor, D. Das, D. Mills, J. Eisenstein, "Part-of-speech tagging for twitter: Annotation, features, and experiments," Technical Report, 2010.
- [31] A. Ritter, S. Clark, Mausam and O. Etzioni, "Named Entity Recognition in Tweets: An Experimental Study," Proceedings of Conference on Empirical Methods in NLP, 2011.