Integrating Social Media with Existing Knowledge and Information for Crisis Response

Gareth J. F. Jones
Centre for Next Generation Localisation
School of Computing, Dublin City University
Glasnevin, Dublin 9, Ireland
gjones@computing.dcu.ie

ABSTRACT

Existing studies on social media in the context of crisis have studied the content of items and their patterns of transmission. However, social media content generated during a crisis will generally be unstructured and only reflect the immediate experiences of the authors, while the volumes of data created can make rapid interpretation very challenging. Crisis situations can be characterized with various expected attributes. In many situations there will be large amounts of information relevant to the situation already available. We argue that existing natural language engineering technologies can be integrated with emerging social media content utilization techniques for more powerful exploitation of social media content in crisis response.

Categories and Subject Descriptors

H.3.3 [**Information Search and Retrieval**]: Search and Retrieval

General Terms

Human Factors

Keywords

Social Media, User Contributed Content, Multimedia, Natural Language Engineering, Crisis Information Mining and Support

1. INTRODUCTION

Developments in the reporting of natural and man-made crisis situations in recent years have greatly increased their visibility. Reports from news agencies and social media sources often now show the developments in a crisis situation in real-time as they happen, often beginning within minutes of the start of an emerging crisis.

Real-time reporting enables interested parties inside and outside of the crisis to be informed of developments, but can also provide information to those coordinating efforts to respond to the crisis. In order to best exploit the available information it is important to understand the attributes of disaster situations, and the current responses of affected communities and how these may change when they are better informed of the situation via real-time media sources [7].

The 3rd Workshop on Social Web Search and Mining (SWSM2011): Analysis of User Generated Content Under Crisis July 28 2011, Beijing, China

Traditional broadcast media have long played a significant role in reporting crisis situations. The most significant change in reporting of events in recent years has been the emergence of Web 2.0 and social media sources which enable real-time publication of updates of the situation as it develops. An important feature of this content is that it is increasing being created on location-aware mobile devices providing GPS location of the point of transmission (on occasion location can potentially be extracted from the content itself). Thus temporal and location context provides a potentially very powerful means of tracking the development of a crisis situation. However, as is made clear in [2], best exploitation of social content requires careful understanding of its nature. Important studies are currently being made of how social media sources such as *Twitter* can be used in tasks such as earthquake response [3]. However, while there is significant research to be done in best utilizing individual social media streams, these will often form a component of a larger system [1].

We suggest that research in addition to focusing on the nature of social media content and its sociological impact on behaviour in crisis and the development of technologies to better exploit its use, we should also look at how its utility might be further increased by examining how it can be integrated both across multiple social media streams and with information and knowledge resources already available before the crisis arose. Established and emerging natural language and multimedia technologies provide a range of tools and techniques which can potentially form components of more complex integrated information management systems for crisis response.

The remainder of the paper considers how formal knowledge description and natural language technologies might be used within systems of this type.

2. CRISIS CLASSES AND ATTRIBUTES

The exploitation of social media sources in crisis situations has so far been rather informal. However, crises can be largely be categorised with expected attributes [5]. A basic classification of crisis features would include: natural and man-made situations, expected or unexpected, (e.g. random shooting), long term or short term and geographically widespread or local. Clearly many combinations are possible and some crises are not easily classified into one class or the other (e.g. an earthquake expected sometime in the coming decades). Taking this analysis further can lead to a topology of crises as proposed in [5]. This can include details of important attribute features of each crisis which

provides input to the requirements of search and mining in terms of vocabulary, entity types and data attributes to look for, leading to crisis type specific information extraction or question answering tools [4]. Thus when a crisis occurs it can be classified and available information mined to fill the relevant attributes slots, with empty slots providing prompts for workers to find potentially significant missing information.

3. INFORMATION SOURCES FOR CRISIS RESPONSE

The timestamp and location data often provided with social content can be used in various ways. Most obviously information associated with can be looked up in increasingly rich gazeetters and lists of significant toponym interest points. While social media sources can provide rapid sources of information about a developing crisis, their informal nature and their often short format mean that they are unlikely, on their own, to provide sufficient information to maximise their potential value. Taking account of this observation, existing information sources should also be incorporated. Such sources could include more static social media sources such as *Wikipedia*, and more formal content describing the location of the incident including details such as geographic or toponym features. Beyond this further useful relevant information may be located on the web at large.

4. CONTENT ANALYSIS

Techniques of varying complexity could be used to process the available content. For example, relevant vocabulary could be identified using a combination of location data, geographic and toponym features, and crisis attributes. This would then be available to adapt other language-based applications operating in the system. Other techniques that could be used include methods such as multidocument summarization to produce an integrated view of the available information, perhaps highlighting key facts in the information. In order to identify missing attributes of the crisis a question answering process could be applied over the available data. When working effectively this would could make available important information relating to the crisis situation and the environment in which it is taking place. However, since the data is from multiple very heterogeneous sources, locating this information reliably is likely to prove very challenging. In order to support search, one might look at applying methods such as document expansion to social content to increase its retrievability or alternatively to align it with more formal content describing the topic.

A further challenge of using this content may arise since, as observed previously, the content may be multilingual. For example, individuals may be issuing "tweets" in different languages. Indeed the same individual may issue tweets from their account in different languages for the attention of different readers. The availability of multilingual content which may contain unique or corroborating data means that crisis information processing may benefit from the introduction of cross-language search and question answering tools or machine translation of content. Domain adaptation of language technologies tools could potentially be supported using crowdsourcing methods. This has the attraction of enabling many of those who wish to contribute to relief efforts beyond making a financial payment, to make use of relevant knowledge remotely.

The availability of location metadata with the very large archives of socially contributed and manually tagged images which are becoming available, such as *Flickr*, potentially enables pictures taken in a crisis to be automatically matched with existing images from similar locations and annotated using their manual tags [6]. Information on many places and situations appears widely on the web. A further way of gaining useful information about places and developing events can be realised by searching the general web. Suitable queries and shallow extraction rules for the current type of crisis and the general location could be realised by extending the caption expansion methods described in [8].

5. CONCLUDING REMARKS

Crisis represents an opportunity to utilize information search and mining technologies in socially significant ways. Some of the required technologies have been under development for other tasks for many years, others are developing rapidly in parallel with the emergence of large scale social media sources. Some can be used directly, others will require considerable adaptation to the task of crisis support. Determining how to develop powerful information systems for crisis support will depend on careful examination of opportunities and requirements of different situations. However, all the pieces are in place in terms of data supply and technologies to undertake focussed research efforts into better using the available information to reduce the impact of crisis on affected communities.

6. REFERENCES

[1] C. Caragea et al. Classifying text messages for the Haiti earthquake. In *Proceedings of ISCRAM 2011*, Lisbon, Porugal, 2011.

[2] B. De Longueville, R. S. Smith, and G. Luraschi. "OMG, from here, I can see the flames!": a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proceedings of LBSN '09*, pages 73–80, 2009.

[3] P. Earle, M. Guy, R. Buckmaster, C. Ostrum, S. Horvath, and A. Vaughan. OMG Earthquake! can twitter improve earthquake response? *Seismological Research Letters*, 81(2):246–251, March/April 2010.

[4] R. Grishman. Information extraction: Techniques and challenges. pages 10–27, Frascati, Italy, 1997.

[5] S. Gundel. Towards a new typology of crises. *Journal of Contingencies and Crisis Management*, 13(3):106–115, 2005.

[6] G. J. F. Jones, D. Byrne, M. Hughes, N. E. O'Connor, and A. Salway. Automatic semantic annotation of landmark images with web mining. In *Proceedings of SAMT 2010*, Saarbrücken, Germany, 2010.

[7] L. Palen and S. B. Liu. Citizen communications in crisis: anticipating a future of ict-supported public participation. In *Proceedings of CHI 2007*, pages 727–736, 2007.

[8] A. Salway, L. Kelly, I. Skadina, and G. J. F. Jones. Portable extraction of partially structured facts from the web. In *Proceedings of IceTAL 2010*, Reykjavik, Iceland, 2010.