Roadmapping Discussion Summary
–
Social Media and Linked Data for Emergency Response

V. Lanfranchi[1], S. Mazumdar[1], E. Blomqvist[2], C. Brewster[3]

[1]OAK Group, Department of Computer Science
University of Sheffield, Regent Court – 211 Portobello Street,
S1 4DP Sheffield, UK
{v.lanfranchi, s.mazumdar}@dcs.shef.ac.uk

[2]Linköping University, Department of Computer and Information Science,
SE-581 83 Linköping, Sweden
eva.blomqvist@liu.se

[3]Operations and Information Management Group
Aston Business School, Aston University, UK
c.a.brewster@aston.ac.uk

Abstract. This paper provides a summary of the Social Media and Linked Data for Emergency Response (SMILE) workshop, co-located with the Extended Semantic Web Conference, at Montpellier, France, 2013. Following paper presentations and question answering sessions, an extensive discussion and roadmapping session was organised which involved the workshop chairs and attendees. Three main topics guided the discussion - challenges, opportunities and showstoppers. In this paper, we present our roadmap towards effectively exploiting social media and semantic web techniques for emergency response and crisis management.

Keywords: Social Media, Linked Data, Emergency Response, Crisis Management.

1 Introduction

Emergencies require significant effort in order for emergency workers and the general public to respond effectively. Emergency Responders must rapidly gather information, determine where to deploy resources and make prioritization decisions regarding how best to deal with the emergency. Good situation awareness [1] is therefore paramount to ensure a timely and effective response. Thus, for an incident to be dealt with effectively, citizens and responders must be able to share reliable information and help build an understanding of the current local and global situation and how this may evolve over time [2]. Information available on Social Media is increasingly becoming a fundamental source for Situation Awareness. During a crisis,

citizens share their own experiences, feelings and often, critical local knowledge. Integrating this information with Linked Open Data, (such as geographic or demographic data) could greatly enrich its value to better prevent and respond to disasters and crisis.

These characteristics make the automation of the intelligence gathering task hard, especially when considering that (i) documents must be processed in (near) real-time and (ii) the relevant information may be in the long-tail of the distribution, i.e. mentioned very infrequently. Common techniques for extracting information from text have been applied to Social Media content with alternate success. For instance, Named Entity Recognition (NER) techniques that extract semantic concepts have been shown to perform poorly on short and noisy social media content [3]. While annotation services and APIs are a highly stimulating research direction for understanding the content and context of social media streams, the aggregation and integration of multi-dimensional datasets, from different domains and large volumes of data still pose a significant technical challenge to development in this area.

Understanding and acting upon large–scale data of different nature, provenance and reliability is a significant knowledge management challenge. Decision-support and visualization techniques must be developed to enable data exploration and discovery for crisis management purposes. Social challenges involved in exploiting social media and Linked Open Data for crisis situations include: credibility, accountability, trustworthiness, privacy, authenticity and provenance of information.

2 Workshop Goals

The SMILE Workshop was aimed at gathering innovative approaches for the exploitation of social media and linked data, using semantic web technologies, for emergency response and crisis management. The workshop attempted to bring together expertise from several research areas such as Semantic Web and Linked Data, Social Sciences, Emergency Response and Information Scientists. The goal of the roadmapping discussion, perhaps the most important aspect of the workshop, was to understand knowledge practices within different communities and organisations and share ideas and expertise to pave the way for better exploitation of existing techniques and frameworks. The papers presented at the workshop are themselves included in these proceedings, hence, the rest of this summary paper is focused on summarising the roadmapping discussion.

3 Organisation

The attendees at the workshop who were willing to participate in the discussion were split into three smaller groups in a break-out session following the presentations. Each group consisted of 3-4 participants, including at least one workshop organiser. The groups were given 20 minutes to discuss and share ideas among the members, with at least one workshop organiser guiding the discussion and taking notes. The discussion was targetted toward three specific topics (implicitly assumed to be set in

the context of the scope of this workshop, i.e. focusing on social media, linked data and other semantic web technologies) :

- *What are the main challenges for Emergency Response?*
- *What are the unexploited opportunities of using Social Media and Linked Data in Emergency Response?*
- *What are the showstoppers that restrict exploitation for Emergency Response?*

At the end of the 20-minute session, the groups convened and shared their notes, thereby stimulating further discussion on these topics. The overall discussion lasted over an hour and resulted in several proposals and highly stimulating conversations. Below, we summarize the conclusions for the three questions being discussed.

4 Challenges

Most of the challenges identified by the group are highly significant to the field, mostly owing to the need for aiding analysis over massive amounts of high velocity information in near real-time. While the social media and semantic web communities already struggle with the challenges of coping with large scale data, the urgency and impact of delays, and other implications of Emergency Response and Crisis Management makes the field even more challenging. Solution developers need to be fully appreciative of the immense potential of social media, but also aware of the repercussions of an imprecise decision and analysis outcome - lives may even be at stake. The technical challenges of discovering information from large volumes of constantly generated and changing social media, quickly and efficiently making sense of information to make actionable decisions and addressing the various technical issues arising out of processing social media content are already established as highly significant to progress in the area. A significant challenge identified by the participants was how information from individual users can be analysed across multiple social media platforms and be collated in a coherent manner. This is essential, as disambiguating identities in social media is a research challenge in itself, and lessons from the field need to be translated into emergency response to aid and assist in taking decisions in real-time. Additionally, information across different media needs to be correlated to present decision makers with a more comprehensive picture of incidents on the ground.

Lessons learnt from past events also need to be captured and fed back into solutions to provide improved analysis for future events. This will result in better and more informed decisions, thereby making a response more effective and efficient. Multi-linguality of social media information also presents a significant challenge. Critical information shared in other languages can be missed out if the solutions fail at handling multiple languages. Experience among the participants showed that this is a challenge even with smaller incidents on a national and local level, in today's mixed and globalised societies. In addition, information shared in different languages can also help decision makers understand a developing situation from different perspectives, as different users are affected in different ways. Solutions also need to

quickly identify the severity and the importance of a piece of information within the context of emergencies – identifying jokes, irony, humour, sarcasm are all essential in reducing noise.

In addition to the technical challenges, the group discussions focussed on addressing the "softer" issues with exploiting social media. One of the most significant challenges is assessing trust within social media – various metrics have been proposed so far that range from analysing followers and friends to weighting verified users. However, a methodological approach for assessing trust and trustworthiness is extremely important, especially in the context of critical domains such as Emergency Response. Assessing validity and reliability of social media data is also an important concern, as social media often spreads without minimal provenance information.

Among the challenges listed during the discussion, the legal one is perhaps one of the most difficult to be addressed by the various actors within the community. The legal community needs to play an essential part in the area, as presently the field lacks a legal framework to guide the exploitation of social media within the context of emergencies and crises. Legal, and privacy, aspects become even more challenging at a global level, where laws may vary greatly between countries involved in a crisis, as may the perception of and need for privacy among social media users of different countries.

5 Opportunities

Most of the opportunities that were identified related to the greater availability and exploitation of Linked Open Data and Semantic Web. Linked Data needs means to be quickly discovered and identified – in the event of an emergency, highly relevant information is harder to find, as information from knowledge sources across different platforms and frameworks need to be effectively aggregated. This, in addition to unpredictable down-times of essential data providers[1] are concerns that solution providers need to address before fully relying on such services and providers.

An opportunity that was recognised unanimously by the group was the need for a greater connectivity between Linked Data and physical sensors. The exploitation of sensors in the emergency response domain has been mostly limited to the sensor owners, providers and organisations. However, providing streaming access to sensors over Linked Open Data frameworks could stimulate an exciting research opportunity for scientists, solution providers and enthusiasts. Making available such information to the public can create highly useful applications and provide efficient solutions to problems that the community has been struggling with over the past years. Several challenges organised as parts of conferences such as the ISWC[2] and ESWC[3] challenges have seen highly stimulating entries over the years and continue to inspire

---

[1] http://labs.mondeca.com/sparqlEndpointsStatus/ lists Linked Data endpoints that are unavailable at the time as well as their uptime during the last 7 days. The large number of endpoints that are unavailable are a concern for the community.

[2] http://challenge.semanticweb.org/2013/

[3] http://2013.eswc-conferences.org/program/mashup-challenge

new research. Such initiatives, more targeted towards streaming data and potentially even specific to our domain, should be organised as part of larger events and conferences to help researchers and enthusiasts explore ideas and contribute toward addressing the challenges.

Using multi-lingual Linked Data corpora for addressing issues arising out of exploiting multi-lingual social media was suggested. However, solution developers need to be aware of the provenance, validity and authenticity when using such data sources as primary sources of information. The application of crowd sourcing techniques for validating and verifying information was suggested as one of the means to alleviate at least some of the challenges (assessing trust, data validity and reliability) that were identified. Such techniques are already being trialled and developed as part of several research projects, such as WeSenseIt[4]. Another solution that was proposed, to help address the same challenges, was to identify "verified" users during emergencies.

In addition to making Linked Data more accessible and available, the group also discussed the need to make social media data more accessible. While user privacy and security are of prime concern to data providers and researchers alike, it is important to make more research data available for solutions targeting Emergency and Crisis Management. While this is a subject that has been debated several times, it is extremely important to identify the opportunities arising out of accessing research datasets. A more collaborative approach together with data providers and services that release anonymised research datasets is essential, to ensure that data and user privacy is protected, however preserving essential content within the data.

6 Show Stoppers

The topics of minimal data availability, assessment of provenance, validity and trust, have been discussed unanimously as significant showstoppers that hinder progress in the area. Trustworthy sources of information in social media are difficult to identify, specifically with various restrictions applied by data providers on user profiles as well as possible absence of provenance information within content. In addition to social media sources, assessing the trustworthiness of Linked Data providers is also a significant challenge.

Identifying and accessing isolated data islands was also identified as a significant showstopper in the field. Several organisations have extremely useful resources that are often not accessible by solution developers. Converting such essential sources of information into Linked Data, and establishing connections within the Linked Data Cloud would provide an excellent opportunity for the community, but the lack of these resources are at the moment an important showstopper.

The lack of working public demonstrations of prototype systems was also identified as an important showstopper. The availability of such prototype systems, and dissemination of such systems within external communities, stimulate discussion

---

[4] http://wesenseit.eu/

and harbour collaborative efforts, which can be extremely valuable for the progress of the community.

7 Conclusions

The workshop identified a set of different challenges and current showstoppers, such as the scarcity of data (both sensor streams and Linked Data sets of the domain), the lack of provenance and trust information (or methods to assess trust), the unclear legal situation with respect to the use of social media data, as well as technical challenges such as the exploitation of noisy social media data streams of high volume in near real-time. The Emergency Response and Crisis Management domain certainly sets some added requirements on the semantic web technologies to be used, e.g. handling international settings with multi-lingual content and high change rate of both information, the underlying concepts, and the format of the information, the need for high trustworthiness of information an conclusions drawn from data, and the urgency aspect that requires near real-time analysis and presentation. However, it is also exactly these challenges that make semantic web technologies useful, and potentially even essential, for solutions in this domain – addressing information access, integration and analysis in a highly configurable and "open" manner, while interpreting the semantics behind the information to achieve truly relevant and accurate results.

Our intention is to make this workshop a successful yearly event, bringing together researchers and practitioners in the field, exploring solutions to the challenges that were identified. By including a yearly roadmapping discussion we will be able to follow the progress of the field, as well as identify new challenges that arise.

References

1. Wong, W. & Blandford, A. (2004) Describing Situation Awareness at an Emergency Medical Dispatch Centre. In Proc. Human Factors and Ergonomics Society's 48th Annual Meeting. Santa Monica, CA: HFES. 285-289.
2. Endsley, Mica R (1995). "Toward a theory of situation awareness in dynamic systems." Human Factors: The Journal of the Human Factors and Ergonomics Society 37.1 (1995): 32-64.
3. Alan Ritter, Sam Clark, Mausam, and Oren Etzioni. (2011) Named entity recognition in tweets: an experimental study. In Proceedings of EMNLP '11. Association for Computational Linguistics, Stroudsburg, PA, USA, 1524-1534.