Crisis Information Management in the Web 3.0 Age

Axel Schulz
Technische Universität Darmstadt
aschulz@tk.informatik.tu-darmstadt.de

Heiko Paulheim
Technische Universität Darmstadt
paulheim@ke.tu-darmstadt.de

Florian Probst
SAP Research
f.probst@sap.com

ABSTRACT

The effectiveness of emergency response largely depends on having a precise, up-to-date situational picture. With the World Wide Web having evolved from a small read-only text collection to a large-scale collection of socially created data accessible both to machines and humans alike, with the advent of social media and ubiquitous mobile applications, new sources of information are available. Currently, that potentially valuable information remains mostly unused by the command staff, mainly because the sheer amount of information cannot be handled efficiently.

In this paper, we show an approach for turning massive amounts of unstructured citizen-generated content into relevant information supporting the command staff in making better informed decisions. We leverage Linked Open Data and crowdsourcing for processing data from social media, and we show how the combination of human intelligence in the crowd and automatic approaches for enhancing the situational picture with Linked Open Data will lead to a Web 3.0 approach for more efficient information handling in crisis management.

Keywords

Emergency Management, Linked Open Data, Social Media, Participatory Sensing, Crowdsourcing

INTRODUCTION

During the last years, the adoption of smartphones equipped with multiple sensors and a constant internet connection increased rapidly. They are widely used in participatory sensing environments (Burke, 2006), e.g. the creation of incident reports in crisis situations. In this case, social media platforms, the essential ingredient of the Web 2.0, are widely adopted for spreading messages about incidents. For example, Twitter was used to report on incidents like the Oklahoma grass fires and the Red River floods in April 2009 (Vieweg, Hughes, Starbird, Palen, 2010) or the terrorist attacks on Mumbai (Goolsby, 2009).

Besides social media, the Semantic Web (sometimes coined *Web 3.0*), and the Linked Open Data (LOD) cloud as another branch of information is rapidly growing on the Web. While the Social Web consists of text and media, all of which cannot be processed by intelligent agents easily, the LOD cloud contains semantically annotated, formally captured information. The Social and the Semantic Web have grown in isolation from each other for a while, but approaches have been proposed for combining both branches of the web, leveraging both the collective intelligence of the crowd, as well as the formal reasoning capabilities on structured data provided by intelligent agents (Gruber, 2008).

Both, the Web 2.0 and the Web 3.0 contribute to a rapid change of the information landscape. However, only few of the available information sources are used actively by decision makers in emergency management. During our evaluations with experts from the emergency management domain, we figured out that the potential information overload is the main reason why these new information sources are not taken into account. Furthermore, it is not clear how high quality data can be retrieved. Nevertheless, it is unquestioned that the massive stream of user generated content contains pieces of highly relevant information that is not known to the decision maker. This is especially the case in the beginning of a large catastrophic incident. Harvesting these information will contribute to a better situational picture finally leading to an improved situational awareness compared to a situation where this information is not available at all. To overcome the problem of the information overflow, the combination of Social and Semantic Web can be helpful: Linked Open Data can enhance Web 2.0 contents by pre-classifying and semantically enriching it. Our vision foresees future crisis management systems, which use structured and relevant information taken from social media, mobile applications and the Semantic Web to increase the situational awareness of decision makers.

The rest of this paper is structured as follows: In section 2, we review related approaches. Section 3 provides a detailed description of our web 3.0 process for crisis information management, followed by a preliminary evaluation in section 4. We conclude with a short summary and an outlook on future work.

RELATED WORK

There are only few approaches for structuring and augmenting information provided by citizens in a crisis management context, or that utilize data sources such as Linked Open Data.

For structuring information about incidents reported via Tweets, the *Tweak the Tweet* project[1] proposes a syntax which supports a more efficient extraction of relevant data, which has to be used before tweets are sent in. For filtering incoming information, Ushahidi, especially known from the Haitian earthquake, is a platform that facilitates crowdsourcing during disasters (Okolloh, 2008).

The application of Linked Open Data in emergency management applications has been discussed in several approaches. (De Faria Cordeiro, Marino, Campos, Borges, 2011) describe an architecture which makes use of the Linked Data paradigm to represent and process information in emergencies. Ortmann et al. have discussed a similar approach (Ortmann, 2011). In their scenario, texts from social networks such as Twitter, are turned into structured information, such as Linked Open Data and enriched using volunteers. One of the rare works combining Web 2.0 and LOD is *SemSor* (Heim and Thom, 2011). In this approach, social media is constantly crawled and information is tagged with links to entities in LOD. This allows the retrieval of information that is not syntactically related to the situation at hand.

Social networks as well as Linked Open Data contain a vast amount of information, which is way too large to handle for an emergency management staff. All the approaches above have shortcomings when it comes to deal with that large amount of information. Information filtering is either neglected completely or supposed to be performed manually by some information management officer, who may quickly become a bottleneck. In contrast, the work discussed in this paper tries to minimize the manual efforts for filtering information by introducing machine learning methods such as clustering and trained classifiers.

THE WEB 3.0 PROCESS FOR CRISIS INFORMATION MANAGEMENT

In Figure 1 three central steps for future crisis information management building on social media, Semantic Web technologies and crowdsourcing are shown. The steps are *information collection, information classification* and *information enrichment*. For all three steps discussed in the Web 3.0 crisis information management process, prototypes have been developed in the InfoStrom[2] project.

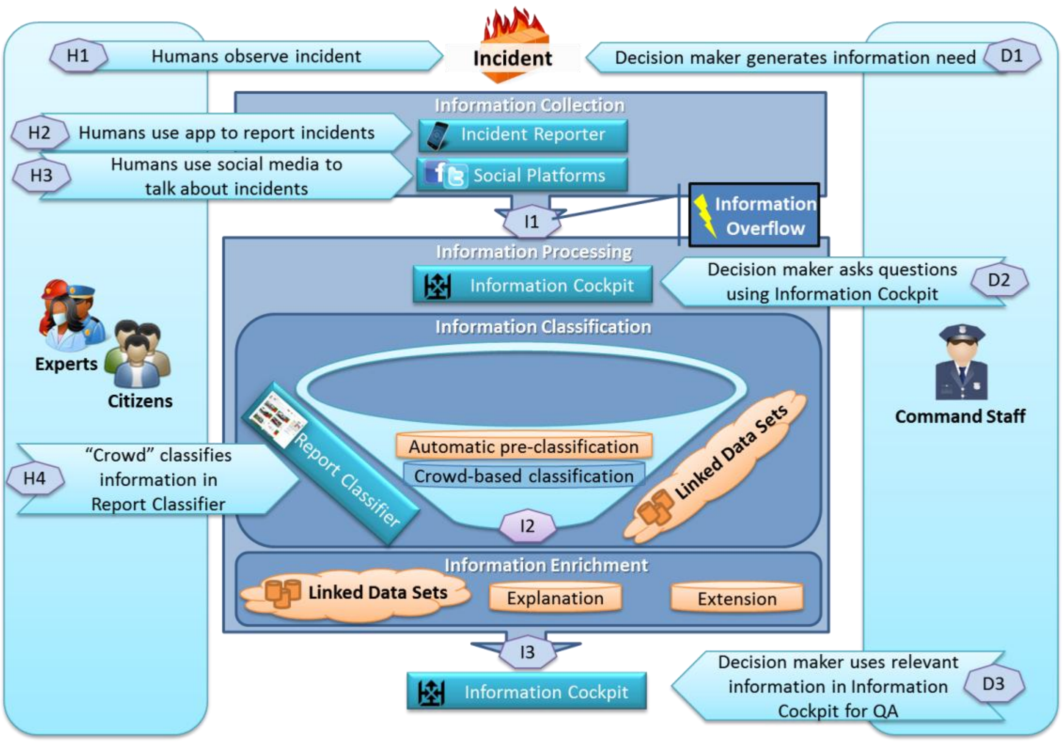

Figure 1. The Web 3.0 process for crisis information management

[1] http://wiki.crisiscommons.org/wiki/Tweak_the_Tweet [Accessed: 16-Feb-2011]

[2] http://www.infostrom.org/ [Accessed: 16-Feb-2011]

Information Collection

Today as well as in the future, citizens witnessing an incident (H1) will call for help, but they will also share what they just encountered using social media. With the massive spread of smartphones, this kind of communication will become a valuable information source for crisis management. Channeling and collecting this stream of information will be done via specialized mobile applications (H2) and social platforms like Twitter or Facebook (H3). To cover that use case, we developed the mobile *Incident Reporter* application that can be used by citizens for reporting. The *Incident Reporter* application allows the submission of images, audio, and textual description, which are submitted as clearly related information objects in contrast to existing social media platforms.

Information collected in that way results in an information base I1, that is unstructured, unsorted, and may contain redundancies. Thus, it cannot directly be used for satisfying the information need of the decision makers, but requires further processing.

Information Classification

In the *information classification* step the goal is to reduce the incoming flood of information (I1) to a set of reports that is relevant to the information need of the command staff using automatic and crowd-based classification. We chose the approach of having the command staff asking questions in order to articulate a particular information need (D2). Some example questions would be: "Are there still people in the burning building?" or “Can we cross the bridge with a 12 ton fire truck?”. We call this approach *question guided relevance rating*. Reports that are helpful for answering a question are considered to be relevant for the command staff. These steps can be done using the *Information Cockpit*, which is the central access point for the decision maker. From this module, the decision maker can call for participation in social networks.

Before providing the information objects to a crowd for filtering, they are pre-classified in the a*utomatic pre-classification* step to simplify and speed up the classification process. In this case, Linked Open Data is used. Text information obtained from social networks and human observers is usually very short and may contain noise such as typing errors, abbreviations, and colloquial language. This makes it hard to classify the information automatically. In this case, additional annotations of the information objects are helpful to perform an automatic pre-classification. Common tagging and annotation engines, such as DBPedia Spotlight[3] or Open Calais[4], are capable of augmenting textual information with semantic annotations and references to entities in Linked Open Data.

For example, the decision maker may ask a question such as "Is there a fire at university?" In this scenario, this question implicitly refers to the university of Darmstadt in Germany, although "Darmstadt" is not explicitly mentioned. A message that will be identified as being relevant may be "It is burning at TUD", where "TUD" is the commonly used abbreviation of "Technische Universität Darmstadt", which is the official name of the Darmstadt university. A tagging engine may now annotate the string "university" with the DBPedia category *dbpedia-owl:University*, as well as the string "TUD" with the DBPedia entity *dbpedia:Technische_Universität_Darmstadt*. Since there is a link (i.e., *rdf:type*) between the latter two, a similarity score can be computed between the question and the report in question, e.g., by counting the number of traversed links in Linked Open Data. Sorting the information objects by those similarity scores leads to a pre-classified collection of information, which has a higher relevance according to a question than other information items have. This augmentation process is shown in Figure 2.

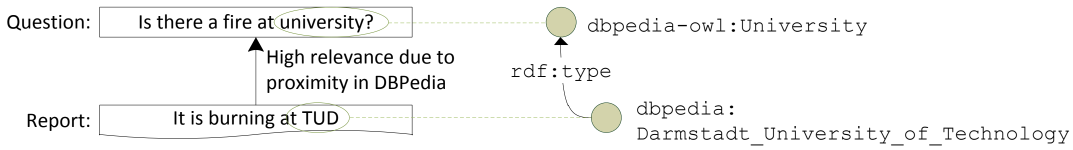

Figure 2. Example process for augmenting textual information with semantic annotations

By using additional machine learning methods, such as clustering, the dataset may be further prepared for being processed. To that end, additional information gathered from Linked Open Data may be used to generate features as input to a machine learning tool (Paulheim and Fürnkranz, 2011). As a result, clusters containing duplicates or very similar information items may be further reduced.

[3] http://dbpedia.org/spotlight [Accessed: 16-Feb-2011]

[4] http://www.opencalais.com/ [Accessed: 16-Feb-2011]

After the pre-classification, crowdsourcing mechanisms are applied in the c*rowd-based classification* step to analyze the information and filter it according to the relevancy for a specific question (H4). For that purpose, we implemented the *Report Classifier* application, which is based on the *question guided relevance rating*.

The crowd must not necessarily be trained rescue personnel for producing useful ratings, as many questions posed by the decision makers can be answered with common sense, as for example, the color or the smell of the smoke. All citizens can access the *Report Classifier* from everywhere, using a web browser. They can select a specific incident and one of the questions posed by the command staff. The application provides all user-generated reports from within the area of the selected incident so that each report can be rated with respect to its usefulness for answering the question. All ratings are then employed in a relevance model to determine a relevance score for each rated information object. Because this concept of the *Report Classifier* only works if the crowd reaches a critical mass, we added Social Network integration. This allows the distribution of a call for ratings to several Social Platforms, where interested people like local citizens use the Classifier to give their ratings.

The resulting dataset contains relevant information specific to questions, which can be further processed.

Information Enrichment

The resulting information is further augmented with additional information. First, this augmentation of the information base I2 is done in the *information explanation* step, as shown in Figure 1, by adding additional information using references to entities in Linked Open Datasets. E.g. the dataset with classified information might contain strings for which background knowledge might not be easy to find, which could be the chemical properties of sulfur to make a proper risk assessment near a sulfur reservoir. Using Linked Datasets like DBPedia[5], Freebase[6] or OpenCyc[7], we can offer additional information as annotations to provide essential knowledge.

Furthermore, the information base can be extended in the *information extension* step shown in Figure 1 by providing additional context information for the objects mentioned in the information stream, e.g., indicating other objects next to the incident that require protection. Important buildings might be hospitals in a reachable area or other protection-worthy buildings in danger like chemistry plants. This information could also be visualized on a map showing the area of danger using geo-referenced linked datasets, such as Linked Geo Data. Another example for information extension is the integration of information about persons relevant for fighting the incident. For example, the president of the university might not be known to the decision maker, but their office might serve as a relevant information hub for coordinating the rescue activities.

The result of the *information enrichment* step is a dataset (I3), containing the information objects resulting from *information classification step* (I2) that are now (partly) explained and augmented with relevant context information. These can now be used by the decision maker.

PRELIMINARY EVALUATION

For analyzing the validity of our approach, we have conducted a series of preliminary user studies with 12 experts with 2 to 20 years of experience in the emergency management domain. Our goal was to receive qualitative feedback on how our process of generating reports using mobile applications and filtering that information using crowd-sourced mechanisms are applicable. The general impression of the prototypes was, that the “presented information is well structured” and the presented prototype “applications are useful” in the work of an emergency response staff. Furthermore, our way of handling the information overflow was seen as promising and helpful for the command staff.

Correctness	Completeness	Timeliness	Objectivity	Credibility	Relevance
4.25 (0.95)	3.5 (0.57)	5.25 (1.26)	4.75 (0.95)	4.75 (0.5)	4.5 (0.57)

Table 1. Average subjective measures for the provided information on a seven point Likert scale with 1 = low and 7 = high (standard deviation in parenthesis)

[5] http://dbpedia.org/ [Accessed: 16-Feb-2011]

[6] http://www.freebase.com/ [Accessed: 16-Feb-2011]

[7] http://cyc.com/opencyc [Accessed: 16-Feb-2011]

The participants filled out subjective ratings for six categories for the provided information. The average score on a seven point Likert scale and the standard deviation for each subjective questionnaire is shown in Table 1.

Although the participants tended to agree that the timeliness of information is high, their perceptions of the completeness of the information were less optimistic (which corresponds with the open world assumption underlying most of the information the web). The objectivity, credibility and relevance of the information are moderately positive. Nevertheless, the overall impression of the presented prototypes was very positive with an average score of 2.8 (0.83) on a Likert scale (1 = positive and 7 = negative). Furthermore, the amount if information provided in the *Report Classifier* application had an average subjective measure of 4.5 (0.57) (1= too low, 4 = appropriate, 7 = too high), which shows that the filtering is already very good.

CONCLUSION AND FUTURE WORK

We presented an approach and prototypes for a new way of crisis information management, combining mobile applications, the Social and the Semantic Web. We have shown how the potential information overload of user-generated content can be efficiently reduced in such a way that a set of relevant information for the command staff results. In our approach, we use crowdsourcing and Linked Open Data to enhance, classify, and filter the information at hand. The result of our approach is a structured dataset, which will enhance the situational picture of the command staff, resulting in an increased situational awareness. With a set of preliminary evaluations, we have shown that our approach is applicable to handle the information overflow for the command staff.

Future work needs to be done to apply machine learning methods for pre-classification. To avoid information overload induced by adding potentially interesting information, it is also necessary to not only classify the information items produced by observers, but also the Linked Open Data entities and their relations between each other. Furthermore, Linked Datasets themselves can be annotated based on their usage. This makes it necessary to take the context of a posed question into account. For example, a dataset might be relevant for a question in a specific context, but not in another one.

ACKNOWLEDGMENTS

The work is partly funded by a grant of the German Federal Ministry for Education and Research (No. 13N10712).

REFERENCES

1. Burke, J. (2006) Participatory sensing. IEEE Internet Computing 14, 1, 12-42.
2. De Faria Cordeiro, K., Marino, T., Campos, M. L. M., and Borges, M. R. S. (2011) Linked Open Data and the Design of Information Infrastructure for Emergency Management Systems. ISCRAM 2011, 1-6.
3. Goolsby, R. (2009) Lifting Elephants: Twitter and Blogging in Global Perspective. Social Computing and Behavioral Modeling, 1-6.
4. Gruber, T. (2007) Collective knowledge systems: Where the Social Web meets the Semantic Web. Web Semantics Science Services and Agents on the World Wide Web 6, 1, 4-13.
5. Heim, P. and Thom, D. (2011) SemSor: Combining Social and Semantic Web to Support the Analysis of Emergency Situations. *Proceedings of the 2nd Workshop on Semantic Models for Adaptive Interactive Systems SEMAIS*.
6. Okolloh, O. (2008) Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. Participatory Learning and Action 59, 65-70.
7. Ortmann, J., Limbu, M., Wang, D., and Kauppinen, T. (2011) Crowdsourcing Linked Open Data for Disaster Management. Framework.
8. Paulheim and Fürnkranz (2012) Unsupervised Generation of Data Mining Features from Linked Open Data. To appear in: International Conference on Web Intelligence, Mining and Semantics (WIMS'12).
9. Vieweg, S., Hughes, A. L., Starbird, K., Palen, L. (2010) Microblogging during two natural hazards events. ACM Press.