Chapter 1
Mashups for the Emergency Management Domain

Axel Schulz and Heiko Paulheim

Abstract Emergency management applications support a command staff in disruptive disaster situations, such as earthquakes, large-scale floodings or fires. One crucial requirement to emergency management systems is to provide decision makers with the relevant information to support their decisions. Mashups can help here by providing flexible and easily understandable views on up-to-date information. In this chapter, we introduce a number of mashups from the domain of emergency management. An in-depth study of the mashup *MICI* shows how mashups can combine valuable information for ranking and filtering emergency calls to cope with information shortage and overload. We further discuss the use of Linked Open Data both as a source of additional information and a means for more intelligent filtering.

1.1 Introduction

Emergency management deals with coordinating operations in situations such as earthquakes, large-scale fires and floodings, or epidemics. There is a large variety of IT solutions supporting emergency management, ranging from intelligent messaging solutions to planning support systems and geospatial infrastructures. Mashups in this domain are especially helpful to provide information to the command staff.

Axel Schulz
SAP Research
Technische Universität Darmstadt, Telecooperation Group
e-mail: `axel.schulz@sap.com`

Heiko Paulheim
Technische Universität Darmstadt, Knowledge Engineering Group
e-mail: `paulheim@ke.tu-darmstadt.de`

Information is the most important resource in emergency management as it is necessary for decision making. Decision makers must know what is happening where and at which time, and they need to know the severity and relevance of all incidents. Hence, obtaining and communicating important information in real time is a crucial task.

Milis and van de Walle show that, despite the existence of numerous research projects on the topic, IT is rarely used in emergency management nowadays for information management. The main problem is that IT solutions are applicable only if there is "a member in the crisis management with an IT background" [43], but the members of emergency management organizations are often non-IT experts [14]. Thus, the barriers for using IT solutions need to be significantly lowered.

With the advent of the Web 2.0, information across the web has been combined in various light-weight mashups, which often have a steeper learning curve than the heavy-weight IT solutions existing in the field. This way of quickly creating representations of multiple information sources that can be viewed in a web browser became very popular over the last years. Especially in emergency management, where decision makers have to identify and react to risks of disasters, this easy understandable way of summarizing information may be valuable to improve the situational picture and minimize the risk of wrong decisions.

While the Web and especially the Web 2.0 is a rich and comprehensive source of information, the potentially valuable information often remains unused because the sheer amount of information cannot be handled efficiently, as it exceeds human cognitive limits. Especially in this case, mashups can help by processing and visualizing that information in a useful, structured way.

In this chapter, we present several mashup approaches for filtering and presenting information relevant for decision makers in emergency management. In an in-depth study, we present our prototype MICI which helps classifying and structuring an unordered stream of data from the fire brigade from Seattle.

The rest of this chapter is structured as follows: In section 1.2, we describe the domain of emergency management and discuss particular requirements to mashups in that domain. In the subsequent sections, we discuss examples for mashup solutions in the field. Section 1.12 discusses an in-depth view on the mashup application *MICI*. We conclude the chapter with a short summary and an outlook on future work.

1.2 The Emergency Management Domain

Emergency management is "the discipline and profession of applying science, technology, planning and management to deal with extreme events that can injure or kill large numbers of people, do extensive damage to property, and disrupt community life" [23]. Improving situational awareness, i.e., "the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future" [24], is one of the main goals for efficient decision making as it is critical for the success of the operation in the first 72 hours of a disaster [25].

At the beginning of a crisis, the emergency management staff begins to gather and analyze information from multiple sources. These might be web sites, governmental information, and information provided by units in the field via radio or mobile phone. In this case, human reconnaissance is required for two reasons [34]: first, to know the current state of an object (e.g., a bridge: is it bending?) and second, to compare it the "normal" conditions (does the bridge usually bend that way?). In this case, the decision maker has to cope with a concurrent information overload and shortage. While information provided in social media can be manifold, this kind of information is unstructured and has further to be verified. On the other hand, the amount of information from on-site rescue squads and bystanders is low. Using mashups, spatial, temporal, and thematic information about events can be used for increasing situational awareness.

Recent research shows the crucial importance of information for emergency management. With the increasing adoption of smartphones with multiple sensors and a constant internet connection [19], humans as soft sensors became valuable sources of crisis information in participatory sensing environments. Social media has been widely used for sharing messages about incidents, e.g. Twitter was used to report on incidents like the Oklahoma grass fires and the Red River floods in April 2009 [53] or the terrorist attacks on Mumbai [28]. Plenty of useful situational information was provided by citizens in these cases. Furthermore, the analysis of social media is valuable for detecting events in realtime or tracking diseases [47, 51]. Social media reports are often more timely and detailed than reports on official news sites, but on the other hand more noisy and more difficult to handle in terms of quantity.

In addition to generating new information, crowdsourcing mechanisms have been applied several times in emergency management to filter and aggregate information, e.g. during the Haitian earthquake. In this case, crowds can be used to significantly decrease the time for generating a map of the affected area [36].

Beside social media, the Semantic Web and Linked Open Data (LOD) have grown as another branch of information. While the social web consists of text, images, and videos, all of which cannot be processed by intelligent agents easily, the Linked Open Data cloud contains semantically annotated, formally captured information. Especially the increasing amount of Linked Open Government Data (LOGD) from the public sector [16] contributes about 42 % to the LOD cloud [17]. This information source could also provide valuable background information for decision-making, however, methods and tools for using that data in emergency management are still being developed.

There are various characteristics that make mashups attractive for the emergency management domain. First, emergency management relies on *timely information*. Mashups involving Web 2.0 and social media sources such as Twitter can provide more timely information than official or commercial news web sites [15]. Second, emergencies are *disruptive and unforeseeable* by nature, which makes it hard to combine all the required information sources upfront. The flexible nature of mashups helps combining the required sources of information in an on-the-fly manner to serve the right information at the right time.

On the other hand, introducing mashups also introduces a set of challenges. One particular requirement of emergency management is the need for *reliable information*, another particular challenge is dealing with *information overload*, which can be easily introduced when presenting the command staff an unfiltered stream of Twitter messages. Thus, mashups in the emergency management domain require *intelligent information filtering and ranking* mechanisms. In the following chapters, we will show how mashups are used to address information needs in the emergency management domain, and discuss approaches to address the above challenges.

1.3 Emergency Management Mashups

Various approaches exist for using mashups in the emergency management domain. Those approaches differ on several dimensions. For categorizing the approaches, we use the following three dimensions: information sources, processing methods, and user interfaces.

A mashup usually comprises many *information sources*. Those may range from static web content (such as text snippets taken from web sites, as well as web mining approaches extracting structured information from the web) to Web 2.0 sources (such as Twitter messages or pictures on Flickr) and government-provided data, and more recently, Linked Open Data as a source of structured information is used. Apart from these mere web sources, sensors (such as water level or earthquake sensors) may be used just as human observations (such as traffic conditions), provided as soft sensors. While human sensors are close to Web 2.0 sources, the former employ specialized interfaces directed at the emergency management use case, while the latter are casual interfaces that are used by humans not primarily with the purpose of supporting emergency management.

The information from various sources needs to be further *processed* in order to get meaningful insights from the raw data. In this chapter, we concentrate on intelligent mashups, thus, we review approaches where the information sources are automatically or semi-automatically enriched with further knowl-

Table 1.1 Overview of Mashup Approaches

Mashup	Information Sources						Processing Methods					User Interface		
	Static Web Content	Web 2.0	Government Data	Linked Open Data	Sensors	Citizen Sensors	NLP, Information Extraction	Semantic Annotation	Machine Learning, Clustering	Information Filtering	Crowd Sourcing	Map	Multimedia	Other
Twitris		X		X		X	X	X	X			X	X	
SemSor		X		X				X	X	X		X	X	X
Disaster 2.0						X					X	X	X	
Linked Sensor Middleware				X	X			X		X		X	X	X
Twitcident		X		X			X	X		X				X
Repopulation Indicators for New Orleans			X									X		
IBISEYE					X							X		
EDIS			X		X					X	X	X		
Live Earthquake					X							X		
Healthmap	X		X							X		X		
MediISys	X	X					X			X		X		X
LA Fire Tweets, SwineFluTweets, Iran Protest Tweets		X										X		
WikiCrimes						X				X		X		
LA Fires, Bushfire Incidents, BushfireConnext	X		X							X		X		
PakReport						X						X		
Ushahidi		X				X				X	X	X		
FindShelter	X					X			X	X		X		
MICI			X	X				X		X		X		

edge. When dealing with text, natural language processing (NLP) methods are often used for extracting relevant topics and/or information snippets from text, or for classifying text messages. Semantic annotation is used for tagging and linking information items with further metadata and external knowledge, such as Linked Open Data sources. Machine learning and clustering can be employed for further processing the data, e.g., for forming relevant clusters, identifying duplicate messages or outliers. In cases where the problem is too hard to fully be automated, crowdsourcing may be employed, e.g., by asking humans to categorize pictures.

The processed information can then be presented to the emergency management staff in various different *user interfaces*. While map interfaces are the most wide spread, some user interfaces also include multimedia data. Furthermore, other interfaces exist, such as expert query interfaces.

Table 1.1 summarizes the approaches discussed in this chapter, and shows which information sources, processing methods, and user interfaces they comprise. In the subsequent sections, we discuss those approaches in more detail.

1.4 Twitris

Sheth et al. present an approach using Twitter for sensemaking, e.g. for the identification of events [33]. Twitris[1] extracts perceptions provided by citizens of real-world events and presents related information to these events. One goal of the approach was to cope with the problems of existing NLP techniques that do not work well on informal text. Therefore, it was necessary to keep the semantics of user defined conventions on social platforms, like hashtags, user names, "@"s. The idea of the approach is to extract context information from microblogs and SMS and to take background information into account, like news, pictures, video or Wikipedia.

The Twitris system uses spatial, temporal, thematic and sentiment analysis on tweets and SMS to analyze events. The whole pipe consists of six steps [50]:

1. The Twitter Search API is used for fetching tweets, while the Ushahidi API is used for collecting SMS. For crawling social media, a set of keywords is extracted using a semantic model from DBPedia [12] and a statistical analysis. Therefore, related concepts from DBPedia to the events at hand are used. The statistical analysis is based on trends, e.g. hashtags used by users and analysis from Google Insights.
2. The text is analyzed based on spatio-temporal-thematic (STT) bias to extract event Descriptors. To identify N-gram summaries for tweets, first, a Spatio-Temporal cluster for an event is created. TFIDF is used to generate the n-grams from these data. Second, spatial, temporal and thematic bias

[1] Twitris is available on `http://twitris.knoesis.org`

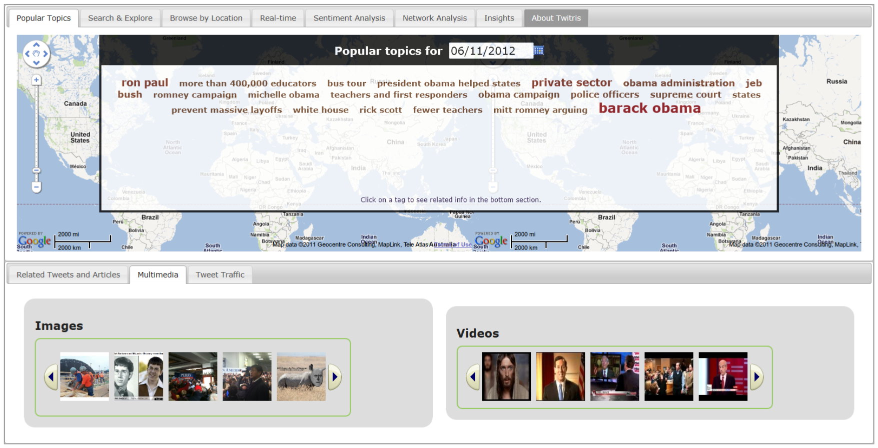

Fig. 1.1 Twitris User Interface showing topic analysis and related multi media content.

to the n-grams are used to weight the relevance for the events. Last, domain models are created using the event context.
3. Semantics are captured from internal (annotations, related posts), external (external sources) and mined internal context (sentiment analysis).
4. Domain models are created automatically to understand the meaning of event descriptors.
5. Entities are semantically annotated for knowledge discovery. Twarql [42] is used to represent social media as LOD.
6. External information sources are integrated using the semantic similarity between the contexts.

The UI has components for theme, time and space. All descriptors for events are represented on this map using the spatial attribute and the date. Additional information are represented using widgets, e.g. for media items, graphs, images, videos, and sentiment analysis. An example screenshot of the UI is shown in Fig. 1.1.

1.5 SemSor

The SemSor project supports the situational assessment for emergency management based on social media analysis [31]. The system constantly crawls social media sources like Twitter, Flickr or YouTube. The textual entries of the social media items are annotated with links to entities in the semantic web (the LOD cloud).

Spreading activation [22] is used on Linked Open Data to identify related entries that might also be useful, starting from the identified entities.

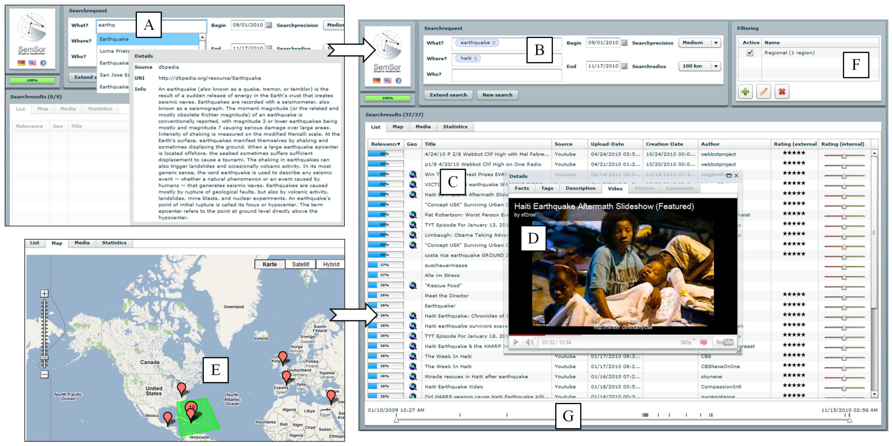

Fig. 1.2 Screenshots of different components of the SemSor UI [31]: forming a search query (A,B). Relevant information are displayed in a list (C) or in a separate window (D). Spatial (E, F) and temporal filters (G) can be applied.

The access on Linked Open Data resources using SPARQL can trigger the spreading activation. Furthermore, contextual information like location, time, or weather data can be used to change the mechanisms. The user's actions can be used to refine the link weights and to adapt the activation energies. Since spreading activation is not limited to a specific domain or data set, the SemSor approach is feasible for all LOD data sets.

The search query and the results are presented in different views as shown in Fig. 1.2. In those views, the search query can also be manipulated, which causes a changing of the weights used by the spreading activation mechanism, which might identify other relevant information. A result browser is showing all results, ordered by their semantic relevance. Furthermore, all relevant entries can be viewed on a map or on a timeline. Both views can be used to define additional temporal and location constraints to filter the information.

1.6 Disaster 2.0

The Disaster 2.0 system follows the idea of using citizen sensors and crowdsourcing to manage information about natural disasters [46]. To that end, the main entities like events (fire, flood), allocated resources (policemen, firemen) and damages (victims) are explicitly modelled in the system with unique identifiers. Information about those entities can be obtained by asking citizens.

The Disaster 2.0 interface visualizes the information using a Google Map view as shown in Fig. 1.3. On this, information sources are provided as markers with additional descriptions and the corresponding location. Additional

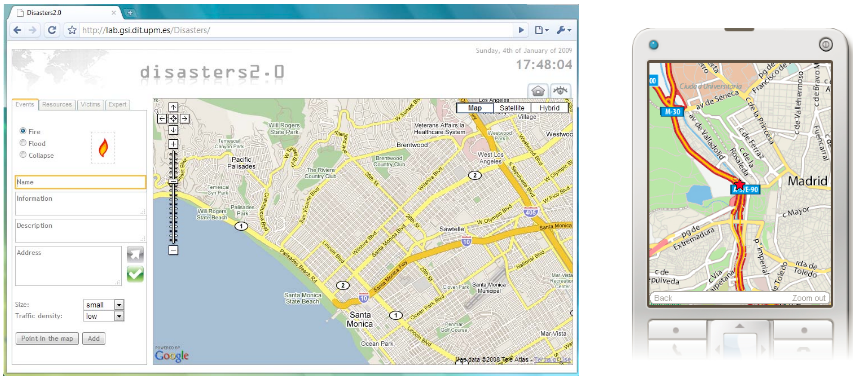

Fig. 1.3 Screenshot of the web client of Disaster 2.0 and the mobile UI for displaying disasters on a map [46].

information might be the magnitude of the disaster or the traffic around the affected area. Furthermore, the map enables the manual association of different information sources by dragging items on each other. For example, victims can be related to disasters this way. Important buildings for the rescue operations can also be displayed. In this case, a simple selection is possible based on the information provided by Google Maps.

A mobile application has been developed in the project as another way of displaying the contents of the map. This application can be used to refine this information, like the correct position of resources. With this mobile application, crowd sourced information can be obtained on a larger-scale, using the workforce of many citizens in the disaster area.

For the support of rescue forces, a rule engine is used to automatically assign incidents to the closest available resources. The rules use information such as the type and size of the incident, the number of victims, and traffic information. These rules can manually be edited by emergency personnel. This expert system works in real time, i.e., updates in the availability of resources are directly triggered to disasters and vice versa. Furthermore, a multi agent system for supporting the decision making has been developed. This system enables rescue squads to query the active disasters and to assign on these. Furthermore, this system enables the coordination of all rescue squads as these send their status and location information to the systems

1.7 Linked Sensor Middleware

The Linked Sensor Middleware (LSM) platform is designed to combine sensor data and the Semantic Web based on Linked Data principles [40]. The purpose of the middleware is to make sensor streams usable by integrating

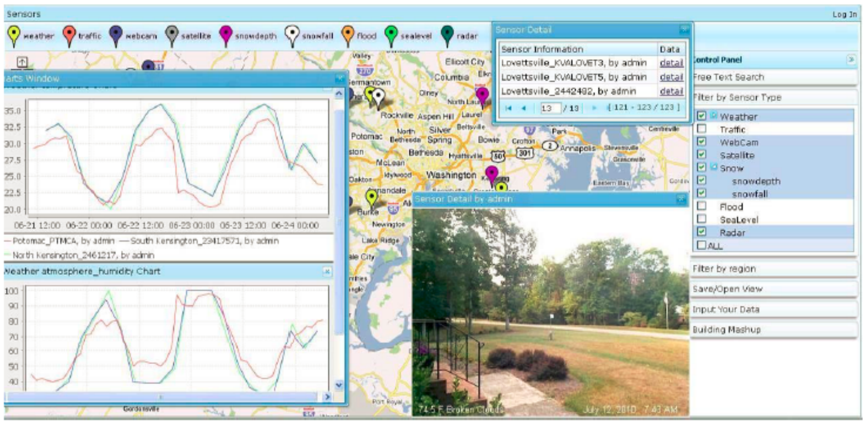

Fig. 1.4 Screenshot of the UI of the Linked Stream Middleware [40].

them with existing information. The whole concept is based on the idea of Linked Stream Data [49] to provide a possibility to publish heterogeneous sensor data as Linked Data to enable the connection to other Linked Data collections. The middleware is divided into four layers:

1. The data acquisition layer provides a wrapper to collect sensor data. These data are transformed to RDF, annotated with links and stored in a database. Based on the Semantic Sensor Ontology[2], two layers are defined for annotating sensor data: the static layer describes the sensor metadata, while the dynamic layer contains a graph-based stream from the various sensor reading.
2. The Linked Data layer, provides access to the sensor data and links to additional information in the LOD cloud. For example links are provided to external sources like DBpedia [12], Geonames [3], or LinkedGeoData [13] to identify points of interest near a sensor location.
3. The Data Access Layer enables the querying of resources based on a remote SPARQL endpoint. This can be used to filter or integrate existing data into new streams.
4. The application layer provides access to the information offered by the system. For example, the sensor information can be displayed on a map where information like weather, departure times of trains, street cameras are shown. An example visualization is shown in Fig. 1.4.

The Live Linked Open Sensor Database project [39] is based on the Linked Sensor Middleware. Two application scenarios for emergency management are mentioned: Possible threats can be taken into account, if additional information from sensor sources is available like the wind force or temperate in

[2] `http://purl.oclc.org/NET/ssnx/ssn`

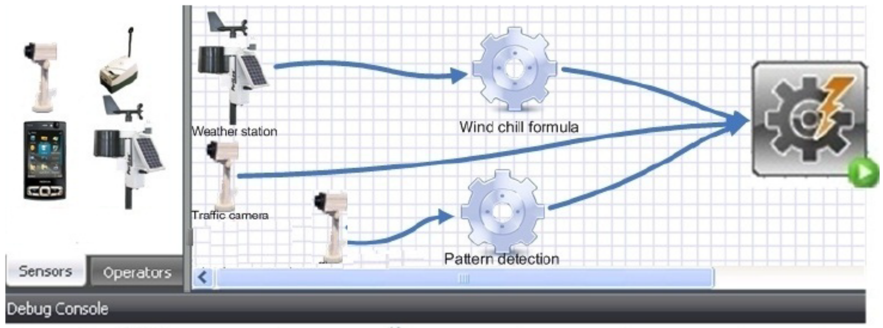

Fig. 1.5 Screenshot of the Sensor Composer based on DERI pipes [39].

case of a fire. Furthermore, during diseases the combination of user-generated content and sensor data might help to prevent the spread of a disease.

Currently 200.000 sensor readings have been integrated into the platform. The different sensors are displayed a map [27] that enables the filtering of data sources with browsing facets. Therefore, text input, a taxonomy or spatial information can be used for filtering. Additionally, the SWEET ontology[3] is used to provide vocabularies for natural phenomena's like blizzards or snow fall which can be used for filtering too. Using the view, current and historical sensor data can be compared. The DERI Pipes engine[4] has been integrated, to enable the user to visually combine sensor data sources for publishing them as new sensors. The view is shown in Fig. 1.5.

1.8 Twitcident

Twitcident[5] is a mashup for filtering, searching and analyzing social media information about incidents [7]. The purpose is to automatically filter relevant information from social media and making these information accessible and findable dependent on the given context.

For incident detection, information about incidents published in the P2000 network[6] are collected, which is used by the public emergency services in the Netherlands. Therefore, the type of incident, its location and its temporal attributes can be retrieved. Furthermore, the incident is scaled for severity in this system. The Twitcident framework transforms these information into

[3] `http://sweet.jpl.nasa.gov/`

[4] `http://pipes.deri.org/`

[5] `http://wis.ewi.tudelft.nl/twitcident/`

[6] `http://en.wikipedia.org/wiki/P2000_(network)`

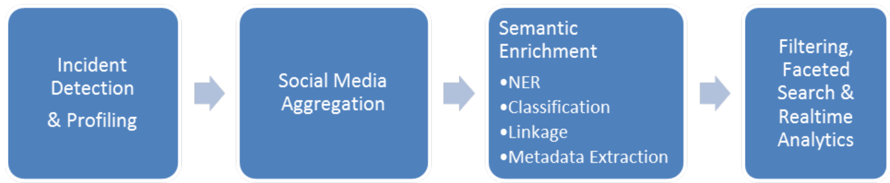

Fig. 1.6 The Twitcident processing pipeline based on [7].

an initial query for collecting potentially relevant Twitter messages (tweets) using the Twitter Streaming API[7].

Four steps as shown in Fig. 1.6 are applied for improving the filtering of tweets [8].

First, based on the incident information, an incident profile is created which describes the attributes of the incidents. This might be related concepts like locations or persons and weights of their importance for the incident. These profiles are constantly updated with changing information.

As a second step, a social media Aggregation module retrieves tweets based on the incident profile, related pictures and videos to a post.

As a third step, the Semantic Enrichment module consists of four different components for processing the information. Before applying the enrichment, non-English tweets are translated to English using the Google Translate API[8], because the components only work in this language. The NER component uses OpenCalais[9], DBPedia [12], AlchemiAPI[10] and Zemanta[11] to detect entities like persons and locations in tweets. The classification component classifies each tweet into different types of reports, using categories such as different damages and risks. Furthermore, the different senses of the Tweeter are differentiated based on manually defined rules containing the basic keywords of the senses. It can be identified if the tweeter sees, feels, hears or smells something. A linkage component extracts links from tweets. These links are used to provide additional context information. To that end, the content of linked web pages is analysed. In a last component, metadata about tweets is collected. Thus, the metadata like the user profile information provided by the Streaming API is used.

As a fourth step, keyword-based and semantic filtering can be applied to automatically identify relevant tweets for an incident. The keyword-based filtering can be used to define a query. The query itself is evaluated using a relevance model based on RM2 [38]. The semantic filtering is based on the enriched tweets.

[7] https://dev.twitter.com/docs/streaming-apis
[8] https://developers.google.com/translate/
[9] http://www.opencalais.com/
[10] http://www.alchemyapi.com/
[11] http://www.zemanta.com

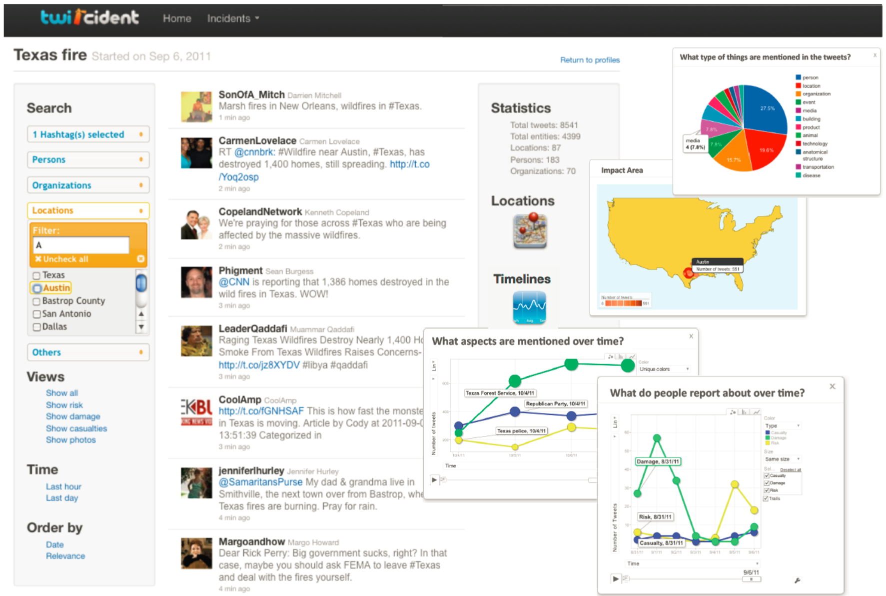

Fig. 1.7 Screenshot of the Twitcident interface, showing the search and filtering (left), the retrieved social media (middle) and the analytic components (right) [7].

The user is enabled to search the information based on faceted search [6] and analytics capabilities. A faceted search can be used to define pairs of attributes that have to be matched by all returned tweets. Different strategies for querying can be applied in this case [7]. The information found can be analysed using different widgets, for example the geographical impact can be identified. The filters applied in the faceted search will also be applied on the analytics widgets. The search and analytics functionality is available to client users in a web front end as shown in Fig. 1.7.

The framework has been evaluated with respect to the filtering capabilities. As a result, the faceted search seems to be the most appropriate approach for filtering tweets relevant for an incident.

1.9 Crisis Map Mashups

Map mashups are the combination of different sources of information that are displayed in some geographic visualization. The use of recently updated information sources like publicly available information, like governmental data, or crowdsourced information enable the view on spatial and temporally changing events. Therefore, crisis map mashups can be used to report, assist and manage emergencies. Zang et al. [54] describe map mashups as a common form of data mashups, because they are “the most visual and adaptable of the mashup options”. Several analyses on crisis map mashups have been made e.g. by [41].

Generally, crisis map mashups can be roughly categorized into mashups that show information from only one source, e.g., Twitter messages on a given topic, and that aggregate information from different sources. The latter often employ some intelligent processing for aggregation, e.g., topic and entity detection for relating pieces of information from different sources.

1.9.1 Non-aggregating Map Mashups

The *Repopulation Indicators for New Orleans* [30] is intended to visualize the repopulation in New Orleans, USA after the Huricane Katrina in 2005 [41]. To that purpose, information from the U.S. Postal Service Delivery Statistics and a Housing and Urban Development Census[12] is used. The *IBISEYE* [32] mashup provides an overview about current storms, their strengths and current weather alerts. In this case, information from `http://www.weatherflow.com/` is used.

The combination of user-generated content with geographic maps has become popular during the last years. *Los Angeles Fire Tweets* [9] displays all tweets related to the string “fire” in a 100miles radius around Los Angeles, USA. *Swine Flu Tweets* [11] shows the distribution of swine flu in the world using the Twitter API. *Iran Protest Tweets* [10] displays tweets related to the iran protests.

Vikalpa [20] is an initiative launched in 2008 to gather and display user-generated reports on election related violence. Reports are created by local journalists and shown on a map to locate related incidents.

1.9.2 Aggregating Map Mashups

The Hungarian National Association of Radio Distress-Signalling and Infocommunications (RSOE) operates the Emergency and Disaster Information Service (EDIS) which provides a so called Disaster and Emergency AlertMap [44]). This map (see Fig. 1.8) displays emergencies obtained from governmental sources around the world, like wild fires, explosions as well as tsunami and earthquake warnings. Furthermore, detailed information about the events can be displayed in a separate view, e.g. the damage level and the number of affected people. The Live Earthquake [21] mashup is more specific for displaying earthquakes of the last seven days on a map and a timeline. Therefore, data

[12] `http://www.hudhdx.info/`

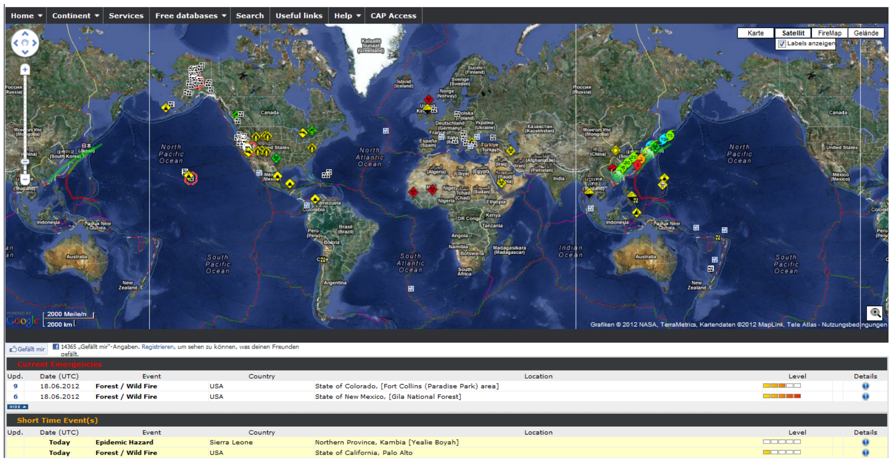

Fig. 1.8 Screenshot of the RSOE AlertMap.

from the U.S. Geological Survey[13], the European-Mediterranean Seismological Centre[14] and the GFZ Potsdam[15] is integrated.

Other governmental sources are used to present data from the public health sector. Healthmap [18] is provided by the Children's Hospital Boston, US, as a mashup for disease outbreak monitoring. In this case, official sources like ProMED Mail[16], the WHO[17] and the FAO[18] are used. Furthermore, news feeds from Google and chinese search engines are aggregated. All available information like the time, the type of a disease and the affected species are displayed on a map, where every pushpin displays the related news. All information can furthermore be accessed using the Outbreaks Near Me mobile application, where the same information as on the website are displayed. This app can also be used for reporting about new known outbreaks. MediISys [35] follows a similar approach to display information about diseases on a map. Furthermore, statistical analysis like recent and past events are provided. News from the WHO, CDC[19] and EMM[20] are aggregated and displayed in a list. All provided information can about alerts and important topics can further be filtered manually.

WikiCrimes [52] was created by the University of Fortaleza, Brazil. It allows the posting of user-generated reports about criminal activities. These

[13] `http://www.usgs.gov/`

[14] `http://www.emsc-csem.org/Earthquake/`

[15] `http://geofon.gfz-potsdam.de/eqinfo/eqinfo.php`

[16] `http://www.promedmail.org/`

[17] `http://apps.who.int/ghodata/`

[18] `http://www.fao.org/corp/statistics/en/`

[19] `http://www.cdc.gov/DataStatistics/`

[20] `http://www.hse.gov.uk/events/index.htm`

reports are aggregated and displayed on a map showing the hotspots of different types of crimes (e.g. theft, robbery). The view can be filtered based on the crime types, temporal attributes and the credibility of the reports. Furthermore, one can receive alerts about crimes in a predefined location.

For displaying warnings about wildfires, the Los Angeles Fires[4] mashup developed by the LA Times integrates several information sources. Reports from LA Times reporters, satellite images and viewer's comments are combined. Bushfire Incidents[2] follows the same approach for Australia. Additionally, weather information and the current position of the fire brigades can be displayed. BushfireConnect[1] is another approach for aggregating official information about fires in Australia with user-generated reports. Information about incidents and news can be displayed on a map and in a list. These can be filtered by the type of media and different categories like the types of the emergencies, official reports or community updates.

Another form of incident reporting has been used in PakReport[5], where users send information about floods via SMS. These information are displayed according to helpful background information like places with water or hospitals.

1.10 Ushahidi

The Ushahidi platform is an open-source crisis management application that was developed in 2008 [45]. Ushahidi uses crowdsourcing for incident reporting and aggregation and became world-wide known from the Haitian earthquake, where the platform was used for translating Creole text messages by crowdsourcing mechanisms. The platform has now been used for several campaigns over the last years, e.g. for tracking events around the Gaza Strip, as shown in Fig. 1.9.

In a first scenario during the elections 2008 in Kenia, citizens could post reports of incidents to the platform [37]. Reports can be send via SMS, mail or Twitter. In a simple mashup based on the Kohana framework[21] and Google maps, those user-generated incident reports are displayed. Using three different timelines, the information can be filtered [41]. The first timeline filters the temporal and spatial information. The second timeline shows the number of incidents in a time period and can be used to filter temporally. The third timeline shows the changes in the number of reports over time while showing the corresponding reports on the map.

For keeping the reliability of information high, reports from incidents which are potentially not trustable have to go through an approval process conducted by the team from Ushahidi [45]. In this case, reports from anonymous people are checked against other sources like the news. The code of the

[21] http://kohanaframework.org/

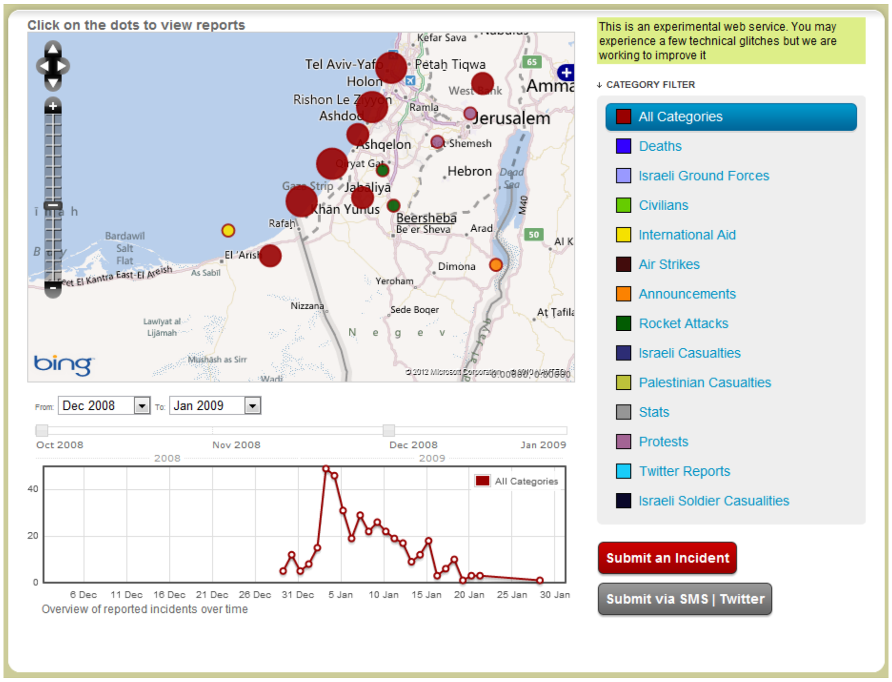

Fig. 1.9 Adoption of the Ushahidi platform from the Al Jazeera Labs for tracking events around the Gaza strip. `http://labs.aljazeera.net/warongaza/`

platform is freely available to adapt it to any situation. E.g. the framework behind Ushahdidi has been published as a cloud service as "Crowdmap" and can be used by anyone for mashup creation.

With the increasing amount of data on the platform, the Ushahidi, Inc. developed the SwiftRiver toolset as a complimentary product [29]. The SwiftRiver toolset provides several APIs for filtering and structuring data from multiple information sources. Social media, blogs or mobile applications can be used as an information source. The fetched data can be processed by the different APIs. For example, the location of the information is identified. Content is analyzed lexically and duplicates are removed. As a result, contextualized messages are created, which can be processed by asking questions like "Where is the person at" or "What has happened" based on the predefined tags provided in a taxonomy.

1.11 FindShelter

FindShelter, the approach described by [26] is based on crawling web pages for important information like emergency accommodations as this source of

Fig. 1.10 FindShelter using a visualization based on Gaia. Black points are hospitals, white points are emergency accommodations.

information cannot easily be aggregated by decision makers. For this, relevant information about entities are retrieved and provided as a web feature service (WFS). The aggregation process is structured as follows.

First, information sources are defined manually. In this case, web pages of a specific domain can be used. As a second step, relevant information are crawled from the web site using breadth-first search. In a third step, information of the same type is extracted and aggregated. To that end, regular expressions are used to extract the relevant information, e.g., for phone numbers and addresses. For completion of the retrieved data it is proposed to use external sources, e.g. using the Google Geocoding API to complete geographic information, and yellow pages for providing additional contact information. In a fourth step, data from different web sites is merged. For every retrieved entity, missing data is added and the attributes are replaced with the newest information. For describing entities, more detailed information can be added using this approach. E.g. the different areas of expertise of a hospital can be added. In this case, a synonym table is used to map retrieved terms to standard denotations. As a last step, the retrieved data is published as RDF and as a OGC-standard web feature service and can be displayed on a map view, as shown in Fig. 1.10.

While it is possible to search directly for hospitals with a search engine like Google Maps, the FindShelter approach is also capable of retrieving and aggregating information on more complex entity types, such as emergency shelters, which can comprise, e.g., schools, convention centers, and sports halls.

1.12 Case Study: MICI

MICI (Mashup for Identifying Critical Infrastructure) [48] is a mashup application which showcases how Linked Open Data can be used as relevant background knowledge in emergency applications. It addresses several central information needs of emergency response staff:

- Many incoming messages need to be prioritized. In situations where the resources are limited and a large number of individual incidents exist, it is necessary to decide which ones need response most urgently.
- Background knowledge about incidents is required. For example, it is required which nearby schools possible require evacuation.
- Information overload has to be avoided. Despite the large number of incidents and the even larger amount of potentially relevant background knowledge, the command staff still has to be able to get a clear situational picture.

MICI reads Open Government Data, e.g., as RSS data, which contains information about incidents. Users may define rules which classify the severity of an incident, e.g.: *if there is a fire within 50m radius of a gas station or a gas pipeline, the severity is high*. For evaluating such a rule, the background information about nearby objects, such as gas stations, is taken from Linked Geo Data [13], a data set within Linked Open Data which contains information about several objects with geo coordinates. For our prototype scenario, we use the fire call dataset from the city of Seattle[22], which provides a list of fire calls with type and coordinates, among others.

1.12.1 User Interface

The main screen of the prototype shows two views on the current situational picture, as depicted in Fig. 1.11. On the left-hand side, a list of incidents is displayed, which can be filtered according to the incidents' severeness. The right hand side shows a map view. When the user selects an incident in the list, that incident and all affected infrastructure objects are displayed in the map.

Rules are used both to find relevant infrastructure objects, as well as to calculate the severeness of an incident. To define which objects are relevant

[22] `http://data.seattle.gov/Public-Safety/Seattle-Real-Time-Fire-911-Calls/kzjm-xkqj`

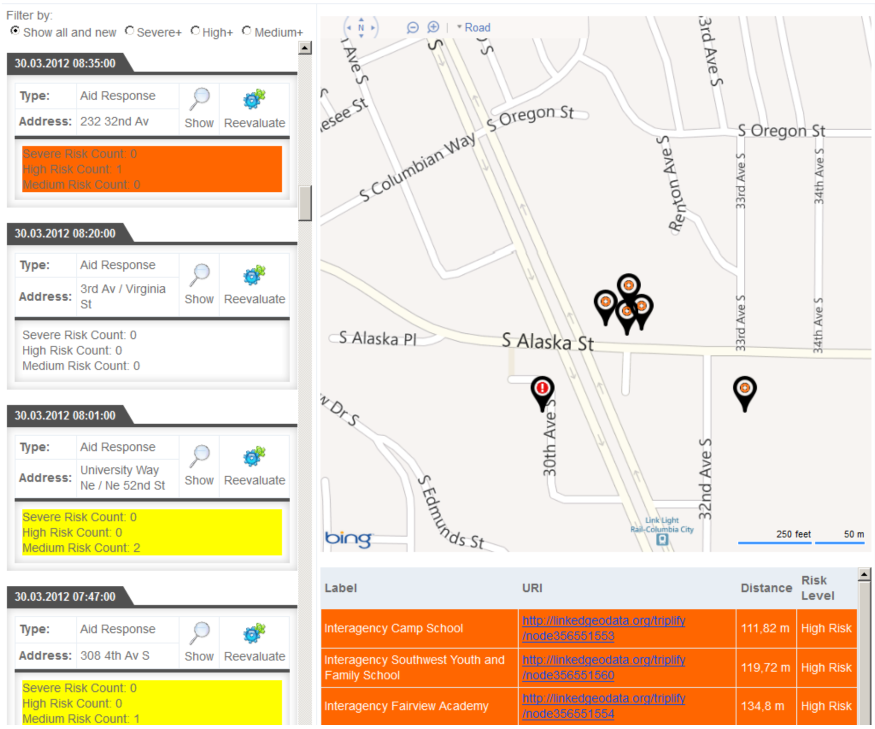

Fig. 1.11 MICI main screen, showing a list of color-coded incidents (left), the map view of an incident including potentially affected infrastructure (top right) and a list view of those infrastructure objects (bottom right)

for which type of incident, users may create their own rules on the rule panel. To that end, they assign a set of object types to an incident type and define a radius and a degree of severity. For defining a rule, the user needs to know about the potential types of objects on which background knowledge is available, i.e., the ontology of Linked Geo Data defining object types such as gas stations, schools, etc. This ontology is used as a vocabulary for the user to define rules, while the vocabulary for the incident types depends on the RSS source used.

As the rule sets are relatively constant, editing the set of rules is something which is typically done once, and not at the time of an incident. During an emergency, the command staff will rather work with the main screen, which provides a clear situational picture to the emergency staff.

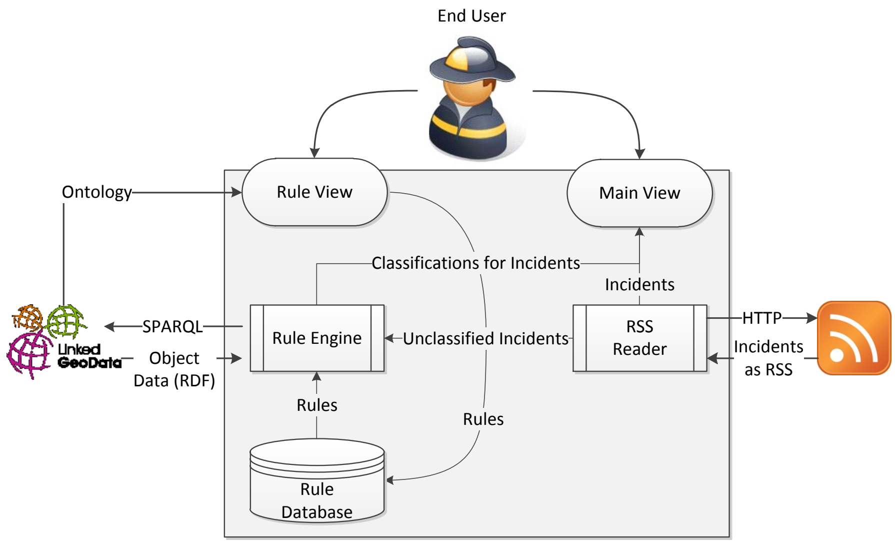

Fig. 1.12 Prototype architecture

1.12.2 Architecture

Figure 1.12 shows the architecture of our prototype. The RSS reader reads incidents from an RSS source. The rule engine is responsible for classifying incidents based on rules entered by the users, using background knowledge from Linked Geo Data. A rule is a tuple of the form:

$$\langle incident\_type, radius, object\_types, class \rangle \tag{1.1}$$

The incident type is provided by the RSS schema of the incident source, the object types are classes from Linked Geo Data, and the classifications are an ordered set of degrees of severity. The above example would be formalized as follows:

$$\langle Fire, 50, \{\texttt{lgdo:Fuel}, \texttt{lgdo:Pipeline}\}, severe \rangle \tag{1.2}$$

For classifying incidents, the rule engine uses the maximum of all radii defined for formulating a SPARQL[23] query to retrieve objects nearby the incident. Each rule is then evaluated against the set of retrieved objects, and the maximum classification of all firing rules is used (for example, if one rule states that an incident is severe, while two more state that it is medium, the overall assessment is severe). Furthermore, those objects that made a rule fire are included in the incident information.

It is noteworthy that each rule serves two purposes:

[23] http://www.w3.org/TR/rdf-sparql-query/

1. Classifying incidents. An unordered list of incident messages is classified into different categories of severity. This allows the emergency staff to identify the most urgent messages.
2. Relevant information objects are picked from Linked Open Data. Since only the objects that made a rule fire are attached to the incident, rules help separating important information objects (such as gas stations and schools) from non-important ones (such as trash bins or traffic lights).

The outcome of the rule processing step is a set of classified incidents, augmented with relevant information objects. These processed incidents can then be used to provide intelligent views. For example, incidents may be filtered by severity, marked with different colors and/or symbols on a map, etc. By adding the corresponding objects responsible for making a rule fire, information on potentially harmed infrastructure (such as gas stations and pipelines) can also be provided to the end user, e.g., in a map view. Thus, the user may not only see that an incident is severe, but also get direct access to the objects that caused that rating, i.e., the name and the position of a specific gas station.

1.13 Conclusions and Future Perspectives

In this paper, we have shown a survey of mashup approaches for emergency management. The need to have timely, accurate, and complete information is a paramount requirement in emergency management. At the same time, it is hard to define the actual information sources in detail upfront, since emergencies always are most often unforeseeable. Thus, the flexible and light-weight nature of mashups is a good fit for complementing IT landscapes for emergency management.

Existing mashup infrastructures make use of diverse information sources, including classical web sources as well as Web 2.0, sensor information and Linked Open Data. Citizen sensors also come into play when specific information is required which can only be delivered by a human on site. However, most mashups only use one or two types of information sources. Richer information infrastructures leveraging the information contained in the deep web together with timely Web 2.0 and sensor data as well as official government data and integrating them into one clear situational picture still need to be developed. How to build complex information infrastructures like these, using sophisticated information filtering and aggregation mechanisms, while at the same time maintaining the light-weight paradigm of web mashups, is still subject to research.

While finding, aggregating, and filtering information is a topic already well covered by the mashup research in the domain, another crucial requirement in emergency management is information reliability. Some first works for addressing these issues, e.g., by crowdsourcing, are currently emerging, but they are still rare. For establishing mashups in the emergency management domain, addressing information quality will be a crucial issue. Apart from crowdsourcing, other possible approaches are weighting information sources or searching for mutual evidence in diverse information sources.

The information used in mashups is most often very timely, especially when using Web 2.0 information sources. However, presenting an analysis in real time is often very difficult. Recognizing a possible trend in hindsight is easy when looking at past data, but recognizing a beginning disruption is a harder problem. For example, the authors of *Twitris* admit that their spatio-temporal-thematic analysis is done with a week of lag. This shows that timely information processing is a topic of active research.

In summary, this chapter has shown that mashups can provide a clear value in the emergency management domain. As this is an active research area, future solutions will be even more powerful than the examples shown in this chapter, and eventually, mashups will be used in the future by emergency management staffs to help them make the right decisions.

References

1. Bushfire connect (2012). URL `http://bushfireconnect.org/`
2. Bushfire incidents (2012). URL `http://www.aus-emaps.com/fires.php`
3. Geonames (2012). URL `http://www.geonames.org/`
4. Los angeles fire map (2012). URL `http://www.latimes.com/news/local/la-me-la-fire-map-html,0,7464337.Htmlstory/`
5. What you see about floods (2012). URL `http://pakreport.org/flood2010/`
6. Abel, F., Celik, I., Houben, G.: Leveraging the Semantics of Tweets for Adaptive Faceted Search on Twitter. In: L. Aroyo, C. Welty, H. Alani, J. Taylor, A. Bernstein, L. Kagal, N. Noy, E. Blomqvist (eds.) ISWC'11 Proceedings of the 10th international conference on The semantic web, *Lecture Notes in Computer Science*, vol. 7031, pp. 1–17. Springer Berlin Heidelberg (2011)
7. Abel, F., Hauff, C., Houben, G.: Twitcident: fighting fire with information from social web streams. In: Proceedings of International Conference on Hypertext and Social Media. Milwaukee, USA (2012)
8. Abel, F., Hauff, C., Stronkman, R.: Semantics + Filtering + Search = Twitcident Exploring Information in Social Web Streams Categories and Subject Descriptors. In: International Conference on Hypertext and Social Media. Milwaukee, USA (2012)
9. Ajmani, V.: Los Angeles Fire Tweets (2008). URL `http://www.mibazaar.com/lafires.php`
10. Ajmani, V.: Iran Election protests (2009). URL `http://www.mibazaar.com/irantweets.html`
11. Ajmani, V.: Swine Flu Tweets (2009). URL `http://www.mibazaar.com/swineflu.html`
12. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: Dbpedia: A nucleus for a web of open data. The Semantic Web **4825**(Springer), 722–735 (2007)
13. Auer, S., Lehmann, J., Hellmann, S.: LinkedGeoData – Adding a spatial Dimension to the Web of Data. In: International Semantic Web Conference (ISWC) (2009)
14. Babitski, G., Bergweiler, S., Grebner, O., Oberle, D., Paulheim, H., Probst, F.: SoKNOS - Using Semantic Technologies in Disaster Management Software. In: The Semantic Web: Research and Applications (ESWC 2011), Part II, pp. 183–197 (2011)

15. Beaumont, C.: New York plane crash: Twitter breaks the news, again (2009). URL http://www.telegraph.co.uk/technology/twitter/4269765/New-York-plane-crash-Twitter-breaks-the-news-again.html
16. Bizer, C.: The Emerging Web of Linked Data. IEEE Intelligent Systems **24**(5), 87–92 (2009)
17. Bizer, C., Jentzsch, A., Cyganiak, R.: State of the lod cloud (2001). URL http://www4.wiwiss.fu-berlin.de/lodcloud/state/
18. Brownstein, J., Freifeld, C.: Healthmap: the development of automated real-time internet surveillance for epidemic intelligence. Eurosurveillance **12**(48), 3322–3324 (2007)
19. Burke, J., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M.B.: Participatory sensing. In: In: Workshop on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications, pp. 117–134 (2006)
20. Centre for Monitoring Election Violence: Election day violence (2008). URL http://cmev.wordpress.com/maps/
21. Clausen, J.: Live Earthquake Mashup (2012). URL http://www.oe-files.de/gmaps/eqmashup.html
22. Collins, A., Loftus, E.: A spreading-activation theory of semantic processing. Psychological review **82**(6), 407–428 (1975)
23. Drabek, T.E., Hoetmer, G.J.: Emergency Management: Principles and Practice for Local Government. International City Management Association, Washington, DC, USA (1991)
24. Endsley, M.R.: Design and evaluation for situation awareness enhancement. In: Proceedings of the Human Factors Society 32nd Annual Meeting, *Aerospace Systems: Situation Awareness in Aircraft Systems*, vol. 32, pp. 97–101. Human Factors and Ergonomics Society, Human Factors and Ergonomics Society (1988)
25. Federal Emergency Management Agency: FEMA 2011 – 2014 Strategic Plan Overview (2011). URL http://www.fema.gov/about/2011_14_strategic_plan_overview.shtm
26. Fritz, C., Kirschner, C., Reker, D., Wisplinghoff, A.: Geospatial Web Mining for Emergency Management. In: GIScience 2010 - Extended Abstracts, pp. 1–5 (2010)
27. Galway, D.: Sensor Masher examples (2009). URL http://sensormasher.deri.org/
28. Goolsby, R.: Lifting Elephants: Twitter and Blogging in Global Perspective. In: Social Computing and Behavioral Modeling, chap. 2, pp. 1–6. Springer US (2009)
29. Gosier, J.: Swift River 2011 (2011). URL http://www.slideshare.net/Ushahidi/swiftriver-2011-overview
30. Greater New Orleans Community Data Center (GNOCDC) : Population and Loss of Children across the New Orleans Metro Area (2009). URL http://www.gnocdc.org/repopulation
31. Heim, P., Thom, D.: SemSor: Combining Social and Semantic Web to Support the Analysis of Emergency Situations. In: Proceedings of the 2nd Workshop on Semantic Models for Adaptive Interactive Systems SEMAIS (2011)
32. Herald-Tribune: IbisEye: Your 2012 Hurricane source (2012). URL http://ibiseye.com/
33. Jadhav, A., Wang, W., Mutharaju, R., Anantharam, P.: Twitris: Socially Influenced Browsing. In: Semantic Web Challenge 2009, demo at 8th International Semantic Web Conference,. Washington, DC, USA (2009)
34. Johnson, D., Zagorecki, A., Gelman, J.M., Comfort, L.K.: Improved Situational Awareness in Emergency Management through Automated Data Analysis and Modeling. Journal of Homeland Security and Emergency Management **8**(1) (2011)
35. JRC European Commission: MedISys (2012). URL http://medusa.jrc.it/medisys
36. Kamenetz, A.: Open Source Maps Are Helping the World Bank Save Lives in Haiti (2010). URL http://www.fastcompany.com/1554072/open-source-developers-help-the-world-bank-save-lives-in-haiti
37. Keay, A.: Ushahidi and Crowdmap: micro streaming as time-binding media. 3 PM Journal of digital research and publishing **1**(2), 116–125 (2010)

38. Lavrenko, V., Croft, W.B.: Relevance based language models. Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval SIGIR 01 **pages**(1), 120–127 (2001)
39. Le-Phuoc, D., Parreira, J.X., Hausenblas, M., Han, Y., Hauswirth, M.: Live linked open sensor database. Proceedings of the 6th International Conference on Semantic Systems - I-SEMANTICS '10 p. 1 (2010)
40. Le-Phuoc, D., Quoc, H., Parreira, J.: The Linked Sensor Middleware – Connecting the real world and the Semantic Web. In: Semantic Web Challenge 2011, ISWC 2011, April 2005 (2011)
41. Lui, S., Palen, L.: The New Cartographers: Crisis Map Mashups and the Emergence of Neogeographic Practice. Cartography and Geographic Information Science **37**(1), 69–90 (2010)
42. Mendes, P., Passant, A., Kapanipathi, P.: Twarql: tapping into the wisdom of the crowd. In: Proceedings of the 6th International Conference on Semantic Systems (I-SEMANTICS 2010) (2010)
43. Milis, K., De Walle, B.: IT for Corporate Crisis Management: Findings from a Survey in 6 different Industries on Management Attention, Intention and Actual Use. 40th Annual Hawaii International Conference on System Sciences HICSS07 **24**, 24–24 (2007)
44. National Association of Radio Distress-Signalling and Infocommunications: RSOE EDIS (2012). URL http://hisz.rsoe.hu/
45. Okolloh, O.: Ushahidi, or 'testimony': Web 2.0 tools for crowdsourcing crisis information. Participatory Learning and Action **59**(January), 65–70 (2008)
46. Puras, J.C., Iglesias, C.A.: Disasters 2.0: Application of Web 2.0 technologies in emergency situations. In: J. Lundgren, S. Jul (eds.) Proceedings of the 6th International ISCRAM Conference, vol. 37. Citeseer (2009)
47. Sakaki, T., Okazaki, M.: Earthquake shakes Twitter users: real-time event detection by social sensors. In: WWW '10 Proceedings of the 19th international conference on World wide web, pp. 851–860 (2010)
48. Schulz, A., Paulheim, H.: Combining Government and Linked Open Data in Emergency Management. In: AI Mashup Challenge 2012 (2012)
49. Sequeda, J.F., Corcho, O.: Linked Stream Data: A Position Paper. In: K. Taylor, A. Ayyagari, D.D. Roure (eds.) International Workshop on Semantic Sensor Networks 2009, vol. 522, pp. 148–157. CEUR, Kerry Taylor and Arun Ayyagari and David De Roure (2009)
50. Sheth, A., Chen, L.: Understanding Events Through Analysis Of Social Media. Tech. rep., Kno.e.sis (2011)
51. Signorini, A., Segre, A.M., Polgreen, P.M.: The Use of Twitter to Track Levels of Disease Activity and Public Concern in the U.S. during the Influenza A H1N1 Pandemic. PLoS ONE **6**(5) (2011)
52. University of Fortaleza: WikiCrimes - Mapping crimes collaboratively (2012). URL http://www.wikicrimes.org
53. Vieweg, S., Hughes, A.L., Starbird, K., Palen, L.: Microblogging during two natural hazards events. In: Proceedings of the 28th international conference on Human factors in computing systems CHI 10, pp. 1079–1088. ACM Press (2010)
54. Zang, N., Rosson, M., Nasser, V.: Mashups: who? what? why? In: CHI '08 extended abstracts on Human factors in computing systems, pp. 3171–3176 (2008)