SemSor: Combining Social and Semantic Web to Support the Analysis of Emergency Situations

Philipp Heim
Institute for Visualization and Interactive Systems (VIS)
Universitätsstraße 38
Stuttgart, Germany
philipp.heim@vis.uni-stuttgart.de

Dennis Thom
Institute for Visualization and Interactive Systems (VIS)
Universitätsstraße 38
Stuttgart, Germany
dennis.thom@vis.uni-stuttgart.de

Thomas Ertl
Institute for Visualization and Interactive Systems (VIS)
Universitätsstraße 38
Stuttgart, Germany
thomas.ertl@vis.uni-stuttgart.de

ABSTRACT

In this paper we introduce SemSor, a system developed especially for the analysis of emergency situations. It constantly collects information from sources of the Social Web, maps it to unique resources in the Semantic Web and uses the annotated information as basis for the situation analysis. If an emergency situation needs to get analyzed, four steps are required: First, all information that is already known about this situation must be entered in the SemSor-GUI. Second, the entered information needs to be mapped to resources in the Semantic Web. Third, using these resources as starting nodes, a spreading activation is applied along the relationships within the Semantic Web to find relevant Social Web information. And fourth, the newly identified information is visualized according to different dimensions and can be filtered and explored by the user. In an iterative process, new insights can be used to refine the query and thus improve the activated information until a comprehensive analysis of even complex situations is possible.

Categories and Subject Descriptors

H.3.1 [**Content Analysis and Indexing**]: Indexing methods; H.3.3 [**Information Search and Retrieval**]: Information filtering, query formulation, relevance feedback; H.5.2 [**User Interfaces**]: Graphical user interfaces (GUI)

Keywords

Social Web, Web 2.0, Semantic Web, Social Semantic Web, situation analysis, spreading activation, interactive information retrieval

1. INTRODUCTION

Analyzing emergency situations is difficult in cases where either the agent who does the analysis is not close by and thus is not able to directly hear or see what is going on, or the situation itself is distributed in space or time so that one single person alone has difficulties in getting an overview. These cases hold true, e.g., in *emergency operation centers (EOCs)*, where neither the agent who receives an emergency call is onside nor the person who has made the call has usually a comprehensive picture of the situation. However, having a comprehensive picture is especially important for the analysis of emergency situations in order to take the right actions and thus prevent all kinds of damage.

To overcome the difficulties in analyzing emergency situations, we propose an approach that combines the advantages of the Social Web with those of the Semantic Web. The idea is to scan Social Web entries, semantically annotate their content and use spreading activation to find exactly those entries that are useful for the analysis of a specific emergency situation. The idea of combining Social and Semantic Web to a *Social Semantic Web* has already been described e.g. in [10] and implemented in many applications, e.g. within the *WeKnowIt*-Project [6]. Also tools have been developed that use this idea to support the analysis of emergency situations [8, 13]. In these tools the found Social Web entries are often arranged on a map to provide an overview of the geographical extent, e.g. on the *Interactive Fire Map* [3], or to extract relevant information via geographical filters [14]. Also popular are timelines that order Social Web entries according to the date of their creation and thus support an understanding of the chronology of events [15]. However, none of these approaches use spreading activation to find semantically related Social Web entries automatically.

The general idea of spreading activation in semantic graphs has first been introduced in [9]. Initially a set of starting nodes is labeled with activation energy, which then is iteratively propagated to other nodes that are linked to the starting nodes. Links can be weighted in order to control the spreading of energy. This can be used, for instance, in recommender systems to adapt the content of a web site to the current context of its visitors. Next to the users' concrete information needs, contextual information like location, time, role, or weather conditions can be used to spread the activation differently and thus to find information that is relevant with respect to a specific context [12]. In addition, every user action can lead to refined link weights and activation energies to account for individual preferences and interests. However, this requires the semantic graph to be stored locally or on a server with write permissions in order to be able to adapt the weights and activation energies accordingly.

.

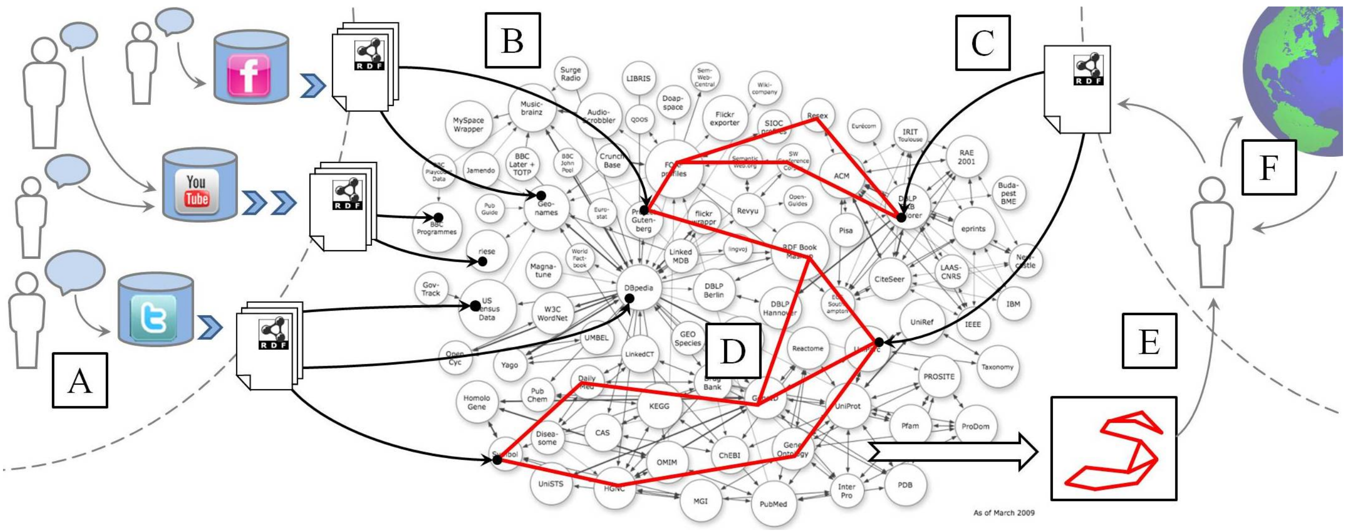

Figure 1: SemSor architecture: Social Web entries are constantly crawled (A) and mapped to semantic resources (B). If a situation has to be analyzed, all known information also needs to get mapped to semantic resources (C), which function as starting nodes for the spreading activation (D). The found Social Web entries are visualized and can be explored and filtered by the user (E). Gained insights, from the visualization or from external sources (F), can then iteratively be used to improve the activated information.

In this paper we introduce an approach that applies spreading activation in external semantic datasets and thus saves storage space and calculating capacities. Datasets in the *LOD cloud* [2] are accessed via *SPARQL* [5] queries to trigger the spreading activation and thus to find semantically related resources. Besides the low system requirements, the two main advantages of an outsourced spreading activation approach are: 1) The datasets are always up-to-date; no complicated methods for updating local copies are required. 2) Semantic relationships of all kinds and domains are used to activate relevant information; spreading activation is not restricted to a predefined set of resources, e.g. resources of a certain domain, but can include all domains contained in the LOD cloud. With the SemSor system, we present an prototypical implementation of our approach that facilitates the extraction of community information relevant to analyze a certain emergency situation. Even though the spreading activation takes place externally, the user can rate the relevance of the found Social Web entries to refine the search query and thus change the activation values until a thorough analysis can be achieved.

In the following we first describe the general SemSor architecture with all the components and steps that are required for the analysis of emergency situations and provide further details to each of the steps afterwards. This includes the crawling and annotating of Social Web entries, the initial query formulation, the spreading activation, the visualization and filtering, and the interactive query refinement. At the end of the paper a conclusion and an outlook on future work is given.

2. SYSTEM ARCHITECTURE

The SemSor System constantly scans Social Web sources, e.g. *Twitter*, *Flickr* and *YouTube*, for new entries (Fig. 1, A) and semantically annotates their textual content. Terms with a distinct meaning, e.g. geographical or temporal references, are therefore mapped to unique semantic resources of datasets in the LOD cloud, e.g. *DBpedia* [7] or *GeoNames* [1] (Fig. 1, B). Thus the system automatically creates a machine readable representation of the semantic that is contained in the found Social Web entries, which can later on be used to support the analysis.

Once a certain situation needs to be analyzed (Fig. 1, F), e.g. because of an incoming emergency call, everything that is known about this situation must also be mapped to unique semantic resources (Fig. 1, C). These resources are then used as starting nodes for the spreading activation [9] that is applied to find all semantic resources that might be of relevance (Fig. 1, D). In a next step, all the Social Web entries that have been annotated with at least one of the activated resources are collected and form the result set, which is presented to the user via multiple views (Fig. 1, E). The result set can interactively be explored and filtered by the user in order to gain new insights about the situation. Gained insights or news from external sources, e.g. from the first responders, (Fig. 1, F) can then iteratively be used to refine the search query, thereby activate new resources in the Semantic Web and thus improve the situation analysis. Due to the possibility to iteratively refine the search query, humans and computers can co-operate in this task.

3. CRAWLING AND ANNOTATING SOCIAL WEB ENTRIES

In order to use information that is contained in Social Web entries to support the analysis of emergency situations, the entries first have to be extracted and annotated by the crawler component of SemSor. Even though the Social Web contains a huge amount of data, only a minimum of this information needs to get stored in the SemSor database. In a first step, a broad multitude of Social Web data gets collected and evaluated according to a preconfigured metric that determines the *a priori relevance* of each individual entry. Within this metric, different properties of an entry like its source, the date and time of its creation as well as location-based data get extracted and serve as basis to calculate a weighted importance rating. The weights of the metric can be configured according to the individual needs of its users (e.g. a specific emergency response team) and provide a basic means to decide, which entries should be kept and which can be deleted if computational- or storage-resources become short. Following the collection and a priori evaluation, the Social Web entries are analyzed and certain terms in their textual contents are automatically assigned to unique resources in datasets in the LOD cloud by using services like e.g. *OpenCalais* [4]. Once new entries have been registered and evaluated, only their URLs and the URIs of the assigned semantic resources have to be kept for subsequent steps.

4. INTERACTIVE SITUATION ANALYSIS

The SemSor system supports the whole situation analysis process, including the initial query formulation, the search for relevant information via spreading activation, the visualization and filtering of the results as well as mechanisms to iteratively refine the query.

4.1 Initial Query Formulation

The method of query is based on common question schemes of emergency calls according to relevant aspects of a situation: "What has happened?", "Where did it happen?" and "Who is involved?" (Fig. 2, B). The agent is supposed to provide approximate answers to at least some of these questions and can further substantiate his query by providing boundaries for the temporal and spatial extent of the situation. In this process, the agent is assisted by an adaptive autocomplete feature, which will try to interactively map given search terms to resources in the Semantic Web. Throughout this process the definitions of proposed resources are provided in pop-up windows, e.g. corresponding *Wikipedia* articles (Fig. 2, A), to help users especially in the disambiguation of ambiguous input terms. Based on this procedure the system is able to get a reliable handle onto the relevant nodes in the Semantic Web. During an emergency situation, like e.g. the 2010 Haiti earthquake, disaster agents can obtain a general overview of the situation by performing a broad search on keywords like "Earthquake" and "Haiti". Based on the interactive mapping of search terms to semantic resources the query is annotated by SemSor and connected with resources in the Semantic Web. The nodes of this framework serve as the initially activated nodes in the spreading activation procedure.

4.2 Spreading Activation

Based on the initial activation of the user defined staring nodes (Fig. 3, A), a homogeneous spreading activation is applied along the relationships between the resources in the datasets. In order to automatically activate semantically related resources that might also be relevant for the situation analysis, the activation happens along the instance-relationship layer (Fig. 3, B) as well as the class-relationship layer (Fig. 3, C). Thus the resulting set of Social Web entries is not limited to those containing references to one of the user defined starting nodes only, but also includes entries referring to resources that are within a distinct semantic radius around the user defined starting nodes (Fig. 3, D);

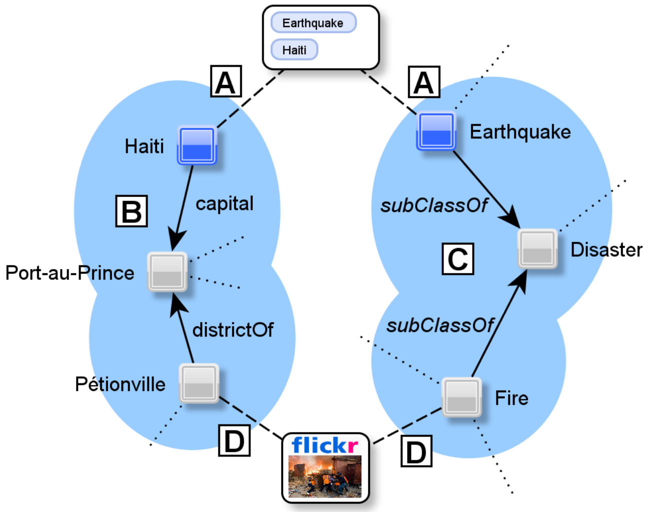

Figure 3: Search terms are interactively mapped to resources in semantic datasets (A) that function as starting nodes for the spreading activation that is applied along the links within these datasets (B and C). As a result, semantically related resources get activated and thus Social Web entries annotated with at least one of them get found (D).

this facilitates finding information that is relevant for the analysis of a certain emergency situation.

The spreading activation in SemSor is implemented mostly as a remote process. On the client side, the process is only triggered and controlled but is run completely within external datasets on server side. Therefore SPARQL queries are sent to the datasets to find resources related to the starting nodes, which are then scored according to the semantical and topological properties of their relationships.

Related resources are found based on an approach described in [11]. Taking the starting nodes as roots, a *breadth-first search (BFS)* is applied to find all resources that are related to one of the starting nodes up to a predefined depth threshold. The depth threshold defines the number of nodes that are allowed between a starting node and a resource that can be activated. Thus having e.g. a depth threshold equal null restricts the activation radius to resources that are directly connected to one of the starting nodes. For each related resource that is found, the algorithm checks whether Social Web entries have been crawled that are assigned to it (e.g. the Flickr entry is assigned to Fire in Fig. 3).

All those Social Web entries are then activated according to topological and semantical aspects and thereby scored. In our implementation the extent of activation depends on three aspects: 1) The length of the relationship, e.g. the length between the Flickr entry and the starting nodes is two, 2) the connection types within the relationships, e.g. "district of" or "capital", and 3) the classes of the intermediate resources, e.g. "Port-au-Prince" is a city. It is also possible to define certain connection types or classes that should not be used to spread the activation, which is useful if someone is not interested in relationships that contain certain instances or connections.

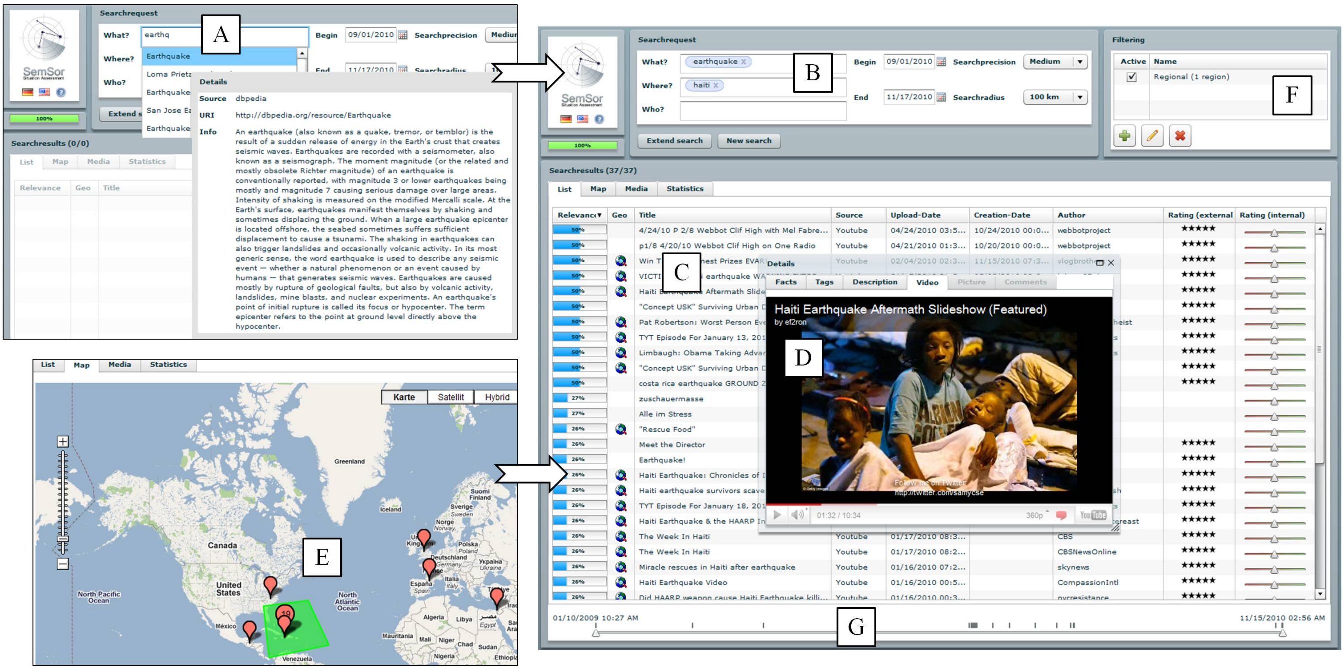

Figure 2: SemSor GUI: Search terms are interactively mapped to semantic resources (A) that together form the search query (B). Relevant information from Social Web sources is found automatically and shown in a list (C). Single entries can be examined in detail (D) and filters can be formulated according to various dimensions, e.g. spatial filters (E) or temporal filters (G).

4.3 Visualization and Filtering

Based on their activation, the Social Web entries are visualized in the SemSor GUI. While the search still continues, all Social Web entries that were already discovered by the activation are presented through the result browser in different user-selectable views (e.g. tabular, map and statistic view). Every view provides the opportunity to obtain pictures, videos and other user generated content related to the situation (Fig. 2, D). The standard view is a listing of entries sorted by their individual semantic relevance (Fig. 2, C). In order to get an overview of the spatial distribution of possibly relevant entries, the agent can view them on a map that can also be used to formulate geographical constraints, e.g. to show only results that refer to a certain geographic region (Fig. 2, E). To further explore the set of results a time line offers an overview over the temporal distribution of events (Fig. 2, G) and allows to formulate temporal constraints, e.g. to show only results that refer to a specific period of time. To further explore the quality and diversity of results the agent has the opportunity to analyze diagrams that show the composition of chosen result-subsets by author, location or tag-categories. By acquiring these initial impressions, the agent can further asses the nature and extent of the situation and initiate subsequent steps.

4.4 Interactive Query Refinement

At this point the advantages of the interactive features in SemSor come into play. If the resulting set generated by the initial spreading activation (Fig. 4, A and B) and diminished by the user defined filters is yet not sufficient, the agent can further refine and expand his initial query by rating single result items on a continuous scale (Fig. 4, C) and by dragging additional tags from the items to the query fields. High ratings of some Social Web entries can then lead to the activation of semantic resources which are directly connected to those entries (Fig. 4, D). Thus after rating the items, the agent is given the possibility to restart his search, but this time, the spreading activation will execute starting also from the newly activated items (Fig. 4, E). Through this "pollination" and subsequent activation of remote nodes in the semantic graph the agent is given the chance to discover relevant regions and new Social Web entries that were not included or even near his initial query (Fig. 4, F). Based on the filters and the interaction procedure, the agent is able to cope with the enormous flood of Social Web data that can be found in connection with emergency situations and make beneficial use of them.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we described an approach that combines the Social and the Semantic Web in order to support the analysis of emergency situations. Exploiting information from the Social Web is especially useful when: (1) the situation is distributed in space or time, (2) first responders are not on site, or (3) the situation cannot or only partly be observed; e.g. this can be the case if a situation takes place within a building or is hidden behind some obstacles.

The automatic semantic annotation of entries in the Social Web as well as the interactive mapping of entered search terms to unique resources in the Semantic Web allows information to be found not only by string matching but also according to its meaning. Therefore spreading activation is applied in external semantic datasets that offer both up-to-date and comprehensive information on all kinds of topics.

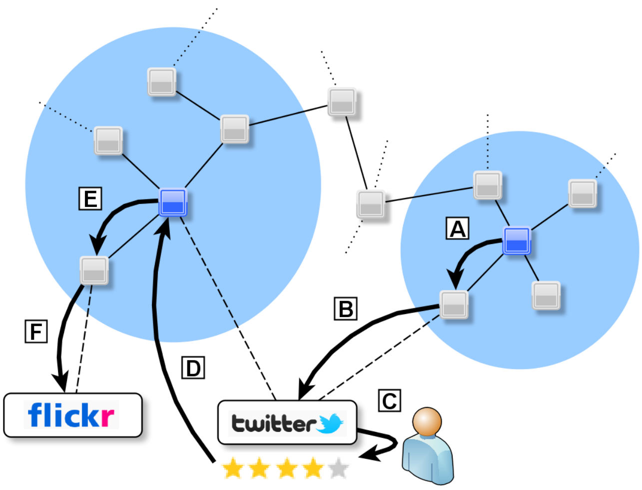

Figure 4: Once an entry has been found via spreading activation (A and B), the agent can rate it as relevant (C) and thus indirectly expand the search query (D). This leads to the activation of other nodes (E) and thus can produce Social Web entries as result set that are only distantly related to the user defined search query (F).

Because of the outsourced spreading activation and the minimum information that is needed to store the annotated Social Web entries – an entry is represented by its URL and the URIs of the annotated resources only – SemSor is able to handle the huge amounts of available data and find information relevant for the analysis of a certain emergency situation. The found information is visualized in multiple views and can be explored and filtered by the agent based on individual information needs. If the result set is not yet sufficient, the query can interactively be expanded or narrowed down. Single entries can be rated as relevant or as irrelevant which changes the starting nodes and thus can result in other entries to be found by the spreading activation. The query can be refined until a sufficient analysis of the emergency situation is possible.

In its current implementation, the SemSor system is most suitable to analyze current or past emergency situations (cp. Fig. 2). Since many people use the Social Web to comment on emergency situations, relevant information is available even while a situation is happening. Together with the fact that only seconds are required from the time a new comment is uploaded to when it can be found in SemSor, it is already possible to facilitate the analysis of current and past emergency situations.

However, analyzing current or past emergency situations often cannot prevent them from happening. In order to prevent emergency situations a preventive analysis is required. First signs of a forthcoming emergency situation need to be detected and interpreted in the right way so that the right actions can be initiated. Besides the agent triggered search, this would require SemSor to automatically scan the Social Web entries for new topics and trends and iteratively produce an overview of the current situation. If certain topics get popular or unusual changes can be detected an alarm could be raced automatically that could force an agent to check the situation and decide on the right actions to prevent a possible emergency situation from happening.

6. ACKNOWLEDGMENTS

The prototypical implementation of our approach, the SemSor system, was developed within a student project by Clint Banzhaf, David Schmid, Dominik Jäckle, Edwin Püttmann, Jochen Seitz, Johannes Dilli, Marijo Macet, Nico Ploner, Steffen Bold, Stephan Engelhardt and Thomas Michelbach. We would like to thank them for their excellent work. Furthermore we thank Bernhard Schmitz, Michael Wörner and Thomas Schlegel for co-supervising the project.

7. REFERENCES

[1] Geonames. `http://www.geonames.org/`, 2006.
[2] Linking open data community project (lod). `http://esw.w3.org/SweoIG/TaskForces/CommunityProjects/LinkingOpenData`, 2007.
[3] San diego wildfires 2007 interactive fire map. `http://www.signonsandiego.com/firemap/`, 2007.
[4] Opencalais. `http://www.opencalais.com/`, 2008.
[5] Sparql protocol and rdf query language. `http://www.w3.org/TR/rdf-sparql-query/`, 2008.
[6] Weknowit project. `http://www.weknowit.eu`, 2010.
[7] S. Auer, C. Bizer, G. Kobilarov, J. Lehmann, R. Cyganiak, and Z. Ives. DBpedia: a nucleus for a web of open data. In *Proc. of the 6th Int. Semantic Web Conf. (ISWC'08)*, volume 4825 of *LNCS*, pages 722–735. Springer, 2008.
[8] J. Borsje, L. Levering, and F. Frasincar. Hermes: a semantic web-based news decision support system. In *Proc. of the 2008 ACM symposium on Applied computing*, SAC '08, pages 2415–2420, New York, NY, USA, 2008. ACM.
[9] A. M. Collins and E. F. Loftus. A spreading-activation theory of semantic processing. *Psychological Review*, 82(6):407 – 428, 1975.
[10] T. Gruber. Collective knowledge systems: Where the social web meets the semantic web. *Web Semantics: Science, Services and Agents on the World Wide Web*, 6:4–13, 2008.
[11] P. Heim, S. Hellmann, J. Lehmann, S. Lohmann, and T. Stegemann. Relfinder: Revealing relationships in rdf knowledge bases. In *Proc. of the 4th Int. Conf. on Semantic and Digital Media Technologies (SAMT 2009)*, pages 182–187. Springer, 2009.
[12] T. Hussein and J. Ziegler. Adapting web sites by spreading activation in ontologies. In *Proc. of the Int. Workshop on Recommendation and Collaboration (ReColl '08)*, 2008.
[13] N. Ireson. Local community situational awareness during an emergency. In *Proc. of the IEEE Int. Conf. on Digital Ecosystems and Technologies (DEST 2009)*, pages 49–54. IEEE, 2009.
[14] C. P. Julio and C. A. Iglesias Fernández. Disasters 2.0: Application of web 2.0 technologies in emergency situations. In *Proc. of the 6th Int. ISCRAM Conf.*, Gothenburg, Sweden, 2009.
[15] S. B. Liu and L. Palen. Spatiotemporal mashups: A survey of current tools to inform next generation crisis support. In *Proc. of the 6th Int. ISCRAM Conf.*, Gothenburg, Sweden, 2009.