

# SensePlace2: GeoTwitter Analytics Support for Situational Awareness

Alan M. MacEachren<sup>1,2,3</sup>, Anuj Jaiswal<sup>1,3</sup>, Anthony C. Robinson<sup>1,2</sup>, Scott Pezanowski<sup>1,2</sup>, Alexander Savelyev<sup>1,2</sup>, Prasenjit Mitra<sup>1,3</sup>, Xiao Zhang<sup>1,4</sup>, and Justine Blanford<sup>1</sup>

GeoVISTA Center<sup>1</sup>, Department of Geography<sup>2</sup>, College of Information Sciences and Technology<sup>3</sup>, and Computer Science & Engineering<sup>4</sup>, The Pennsylvania State University

## ABSTRACT

Geographically-grounded situational awareness (SA) is critical to crisis management and is essential in many other decision making domains that range from infectious disease monitoring, through regional planning, to political campaigning. Social media are becoming an important information input to support situational assessment (to produce awareness) in all domains. Here, we present a geovisual analytics approach to supporting SA for crisis events using one source of social media, Twitter. Specifically, we focus on leveraging explicit and implicit geographic information for tweets, on developing place-time-theme indexing schemes that support overview+detail methods and that scale analytical capabilities to relatively large tweet volumes, and on providing visual interface methods to enable understanding of place, time, and theme components of evolving situations. Our approach is user-centered, using scenario-based design methods that include formal scenarios to guide design and validate implementation as well as a systematic claims analysis to justify design choices and provide a framework for future testing. The work is informed by a structured survey of practitioners and the end product of Phase-I development is demonstrated / validated through implementation in *SensePlace2*, a map-based, web application initially focused on tweets but extensible to other media.

**KEYWORDS:** social media analytics, scenario-based design, geovisualization, situational awareness, text analytics, crisis management, spatio-temporal analysis.

**INDEX TERMS:** H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval—Information filtering, relevance feedback; H.5.2 [Information Interfaces and Representation]: User Interfaces—Graphical user interfaces, Evaluation/methodology.

## 1 INTRODUCTION

Establishing and maintaining Situational Awareness (SA) is a fundamental requirement for effective crisis management. The information contained in social media has been recognized by the crisis management community (e.g., the August, 2010 Red Cross Emergency Social Data Summit [1]) and by researchers in crisis informatics [2] as a potentially rich source of time-critical information that can enhance situational awareness and alert practitioners to events and situations needing their attention. Leveraging social media for crisis situational awareness remains a challenge. Reasons include: large volumes of information, limited means to sort relevant from irrelevant information (with relevance including thematic and geographic components), challenges to distinguish reliable from unreliable information, and limited tools to extract meaning from the generally ill-structured and cryptic formats used in social media applications.

IEEE Symposium on Visual Analytics Science and Technology  
October 23 - 28, Providence, RI, USA  
978-1-4673-0014-8/11/\$26.00 ©2011 IEEE

This paper reports on development of a web-based geovisual analytics application, *SensePlace2*, through which information contained in social media can be gathered and analyzed to support situational awareness in crisis management and related application domains<sup>i</sup>. The initial *SensePlace2* version is focused on foraging for and sensemaking with crisis-relevant information extracted from Twitter. Twitter is a starting point for our ongoing work because it is the most widely used microblogging application, with the number of ‘tweets’ reaching 110 million per day in January, 2011 and escalating (<http://tiny.cc/s8pbc>). Twitter is also special among social media because most tweets are public, thus tweets are more useful for situational assessment than other social media. Plus, the 140 character limit of tweets, the increased use of hash tags (particularly for crisis situations), and the potential for precise tweet locations all add to potential usefulness. Twitter, of course, is only one data source for SA and Twitter has a user age bias in the 25-55 age range. Thus, while this paper focuses on Twitter as the information source, the software architecture is extensible to ingest any relevant text sources.

A focus of the research reported is web-based place-time-theme foraging and filtering of tweet content within a visual interface that provides geographic, temporal, and thematic overview and detail as well as support for analysis histories. In this paper, we focus particularly on leveraging explicit geographic information in tweets (when users add place or coordinate information to tweets), explicit (but imprecise) geographic information contained in the user’s profile (their timezone), and implicit geographic information that can be inferred through identification of place-related hashtags or automated entity extraction methods. An example view from the interface is shown in Fig. 1.

The key innovation presented in this paper is the integration of linked visual-computational methods and a place-time-entity conceptual framework in a working prototype grounded in both theory and practice informed by a structured survey of professionals and guided by scenario-based design methods [3, 4]. In subsequent sections we provide background, present results of the practitioner survey, provide an overview of the application and its architecture, outline our use of scenario-based design and demonstrate that *SensePlace 2* can support it, present a second brief use case, and sketch our plans for future work.

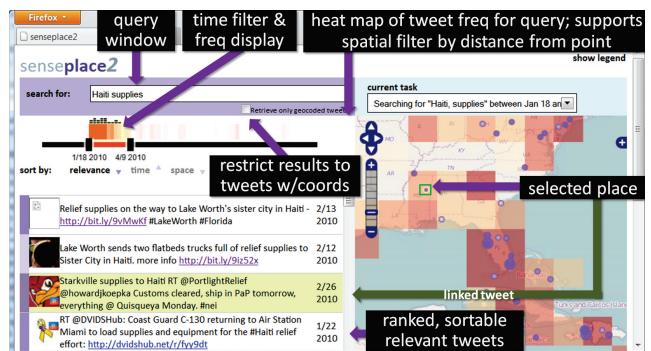


Fig. 1. *SensePlace2* interface: time constraint on query for ‘Haiti supplies’ plus spatial selection of relevant results.

## 2 BACKGROUND

Here we provide a brief overview of selected research streams that are directly relevant to our visual analytics approach for utilizing social media to support crisis management SA.

### 2.1 Situational Awareness

Endsley [5] described SA as a three-part process of perception, comprehension, and projection (into the future to make predictions) that leads to decision making and then to actions. Klein [6] argues that sensemaking and SA are distinct; he contends that SA is about the knowledge *state* of individuals tasked with monitoring and interpreting a situation and making decisions about how to act. Sensemaking, in contrast, is characterized as a *process*. But, SA depends on a process of *situational assessment*. The process of sensemaking as outlined by Pirolli and Card [7], although originally developed to model intelligence analysis, also provides a framework for how the initial two stages of the SA process are (or should be) carried out. Endsley's 'perception' is supported through the foraging loop and 'comprehension' is supported through the sensemaking loop. The outcome of the sensemaking loop in crisis management is often a 'projection' about the future situation given the comprehended current situation plus estimates of changes in the forces driving the crisis and/or responses to it by professionals and the public.

In the domain of power grid operations, Greitzer, et al [8] make a strong case for application of a sensemaking perspective to SA. One of their key points is that traditional information technology support for SA has focused on increasing information throughput, thus scaling human capabilities to process information. They argue that a sensemaking approach addresses SA at a deeper level with a focus on understanding the process through which analysts construct a "plausible 'story' of what is going on." They then propose that it is important to explicitly enable strategies for organizing information into coherent "chunks" and for defining "explicit connections across levels of abstraction to improve awareness of patterns and relationships".

For SA in a domain like crisis management in which "place" is critical, the now classic approach to organizing information into place, time, and concept/theme components is a logical basis for organizing a visual interface that supports foraging and sensemaking activities generally along with more specific SA tasks such as information chunking and the identification of patterns and relationships (for background on place-time-object models for information representation and use, see [9], [10, 11], and [12, 13]). A place-time-object organization of information enables users to pose the fundamental where, what, and when questions that are critical to SA.

In spite of the potential for visual analytics to enable SA, there has been relatively little attention in visual analytics to SA as a goal. One exception is work by Livnat, et al [14] who review the core concepts of SA and focus specifically on applying what they term the *w<sup>3</sup>* premise; that events have where, what, when attributes. They propose a visualization paradigm that addresses these attributes explicitly in the visual representation and interaction mechanisms of tools to support SA; the paradigm combines all components into a ring-based depiction. In one of several examples, they describe application of their visual representation device to visual analysis of event correlation to support emergency awareness. The explicit recognition of *where*, *what*, and *when* is also central to work on application of geovisual analytics to SA reported by Tomaszewski and colleagues [15-17]. An important component of that research is a conceptual and practical framework for developing information foraging and sensemaking tools that help analysts find and process documents ingested into an analytical process of geo-historical (space-time)

context creation to underpin information interpretation in crisis and humanitarian situations.

In other research with a crisis management focus, Kim et al [18] introduce mobile devices with map-based visual analytics methods providing SA support for emergency response; but they do not adopt a particular conceptual approach to SA nor do they focus explicitly on any systematic analysis of SA strategies or tools. In contrast, Kohlhammer, et al [19] introduce a visual analytics approach to decision-making grounded in an understanding of both SA and naturalistic decision-making. In relation to SA, they cite Endsley's [20] three levels and describe the comprehension step as based on synthesis of disjoint elements to "form a holistic picture" that includes understanding of object and event significance. They argue that visual analytics methods for SA must go well beyond enabling information perception to support comprehension in the context of existing goals.

Several key insights derived from work on SA outlined above guide work reported here. First, we have derived a three-part working definition of crisis management SA: *the process of determining the state of the world relevant to crisis management goals, the result of that determination (the knowledge state), and the use of the knowledge achieved to make decisions about action or predictions about future states*. Second, since place and time are critical to crisis management, we adopt a place-time-theme approach to foraging for and sensemaking with fragmented data. Third, we consider understanding of object and event significance and identifying connections across levels of abstraction to be critical to SA and focus (in initial work reported here) on using integrated overview-detail methods applied to both place and time to facilitate that understanding.

### 2.2 Social Media in Crisis Management

Social media have been acknowledged by many in the crisis management research and practice communities as a tool for both communicating and collecting information and a growing body of research reflects this interest and development [21-23]. Twitter specifically has drawn considerable research attention as a tool to gather information, disseminate information, coordinate actions, and support SA [24-27].

In work with a specific geographic component, Vieweg, et al [2] provide evidence that microblogs, using Twitter or other means, have a potential role in enabling common SA among practitioners in crisis situations. Specifically, Vieweg and colleagues collected, hand coded, and analyzed tweets during a pair of concurrent events (a grass fire in Oklahoma and a flooding event on the Red River having impacts from Winnipeg, MB to Fargo, ND). They were able to identify both geo-location and situational update information in Tweets that are directly relevant to building SA. The authors concluded that most local users will include geolocation in at least one event-relevant tweet and that geolocation information will differ in kind and amount depending on the nature of the crisis event and the stage within the crisis lifecycle (e.g., in the grassland fire event, geolocation information was included in 40% of all on-topic tweets, but only 18% of on-topic flood event tweets contained location information). They also found that place-relevant information occurred without specific locations mentioned in 6-8% of tweets (e.g., west central edge of town). In relation to situational update information, approximately half of on-topic Tweets in both events included situational update information. One additional finding is that tweets with either geolocation or situational update information are both more likely to be retweeted than tweets in general.

The concept of volunteered geographic information (VGI) is a complementary research area focused on utilizing the potential of crowd-sourcing to generate useful geographic information to support many activities. These range from basic mapping (e.g.,

Open Street Map) to direct support for crisis management risk communication (e.g., CrisisMappers.org, Ushahidi). Zook, et al [28] provide a recent overview of VGI/social media applied to crisis management with a focus on Haiti.

Drawing from sources cited and many others, we identify five distinct uses for social media in crisis management:

- a) to disseminate information to the public (e.g., for alerts)
- b) to gather information from the public (e.g., crowdsourcing)
- c) to coordinate with crisis management professionals
- d) to monitor activities of crisis management professionals
- e) as input to situational assessment for crisis management

### 2.3 Social Media Mapping & Analytics

There is a growing volume of research on analytical methods to leverage Twitter and other social media-derived information for many applications, including crisis management, investigative journalism, and political analysis. Here, we highlight a few key developments that informed our work.

Dörk, et al [29] present a strategy for using Twitter (and related microblogging) as a “backchannel” for ongoing live events. Through experience developing and using their *Visual Backchannel interface*, they identify four design goals: *summarize the conversation, integrate the ‘now’ with the ‘recent’, extend presence of the present* (make now persist long enough to be useful), *provide flexible time windows, make visual activity adjustable* (through query and filtering), and *encode organic feel*.

In research that is mostly analytical, Shamma, et al [30] present results on identifying structure and semantics in microblogging (via Twitter) about media events and providing that information back to the microbloggers to enhance their experience, e.g., by uncovering emerging topics and trends in real time. Their approach includes attention to message frequency, message length (which are shorter during key parts of the event), number of messages with an @ designation (also fewer during the most interesting parts), and how conversational tweets are.

Diakopoulos, et al [31] integrate analytical methods with a visual interface to support journalists’ use of Twitter to supplement other information sources. They ground the work in the Pirolli and Card [7] sensemaking model. Their research focus is on extracting and representing relevant information from large streams of Twitter (or other social media). The approach and tools utilize computational content analysis methods that address four tasks: judging uniqueness, analyzing and classifying sentiment, assessing relevance, and extracting keywords. The latter two are also a focus in the SensePlace2 application we present below.

In work with photo sharing sites rather than tweets, Andrienko, et al [32] introduce a set of analytical techniques focused on reconstructing past events from activity traces. The tools integrate geovisualization, geocomputation, and statistical analysis methods that support integrated analysis of the spatial, temporal, and thematic components of data that includes both numeric attributes and text. The approach is able to scale to fairly large data sets due to a previously developed strategy for territory tessellation [33]. Ideas about both tessellation (to create overviews) and strategies (for space-time-theme parsing of information) informed decisions on SensePlace2.

Field and O’Brien [34], also with a geographic focus, present ideas on web-mapping of tweets. Citing the lack of attention to cartographic design and map usability of most existing efforts to map tweets, they propose strategies for representing multiple coincident messages in intuitive ways. They also describe a successful case study in which students in a field course used Twitter to support real-time collaboration on an assignment that required coordinated data collection and integration of the information collected (thus an activity not unlike those undertaken during crisis management).

## 3 SURVEY OF PRACTITIONERS

Here we present results of a structured survey of crisis/emergency management professionals who responded to questions about potential uses of social media in crisis management and provided feedback on an early design mockup for SensePlace2.

### 3.1 Understanding User Needs

As an input to our tool design and development process, we completed a needs assessment study with two primary goals: (1) to develop a clear understanding of how emergency management professionals currently use social media tools and (2) to elicit what emergency management professionals envision doing with social media geovisual analytics tools in the future.

To address these goals, we developed a 21 question survey delivered on-line using SurveyMonkey. To recruit participants, we posted a request to the International Association of Emergency Managers (IAEM) listserv and to FirstResponer.gov. Our request yielded 46 respondents. A few questions were skipped by one or a few respondents, thus some received fewer than 46 responses. The first survey question asked about work experience in emergency management. Respondents had a high level of expertise overall; the majority (53.7%) had over 9 years of experience, while 22% had between 6-9 years, 2.4% between 3-6 years, 12.2% between 1-3 years, and 9.8% claimed no experience working in emergency management.

The following sections reveal what we learned from this survey on our two key research goals: the current use of social media, and the envisioned use of geovisual tools that support social media analysis. Our results are reported with percentages for questions that allow multiple options to be selected, and by average subjective rating for questions that use a 5-point Likert agree/disagree scale.

### 3.2 Current Use of Social Media

To begin understanding how social media is used by emergency management professionals, we asked participants to characterize their use of social media tools in personal, professional, and crisis management settings. Facebook (97.4%), LinkedIn (68.4%), and Twitter (63.2%) are most common for personal use. For professional purposes, LinkedIn (67.9%), Twitter (53.6%), and Facebook (53.6%) are most common. With respect to crisis management activities (Fig. 2a), participants primarily use text messages to communicate with other professionals (60.9%), Facebook or other “friend” networks (52.2%), Twitter to gather information from the public (43.5%), and text messaging to communicate with the public (43.5%).

We also asked participants to assign a rating to the usefulness of microblogging tools like Twitter for each of the four key phases of crisis management (mitigation, preparedness, response, and recovery). Microblogging tools are currently considered to be more useful for response and recovery activities than for mitigation and preparedness (Fig. 2b).

Next, we asked participants to rate the usefulness of tools like Twitter for specific emergency management tasks. Our results suggest that such tools are most useful for dissemination, information gathering, and as inputs to situational assessment. There is weaker agreement that such tools are useful for monitoring other emergency management personnel or coordinating their activities (Fig. 2e). We did not provide or collect definitions of SA from respondents. However, one of us has been a member of the IAEM listserv since 2004 and identified 137 posts since then that mention SA; the emphasis in these posts is on both the state of being aware of all factors relevant to successful planning, response, and recovery as well as on the process of achieving that awareness.

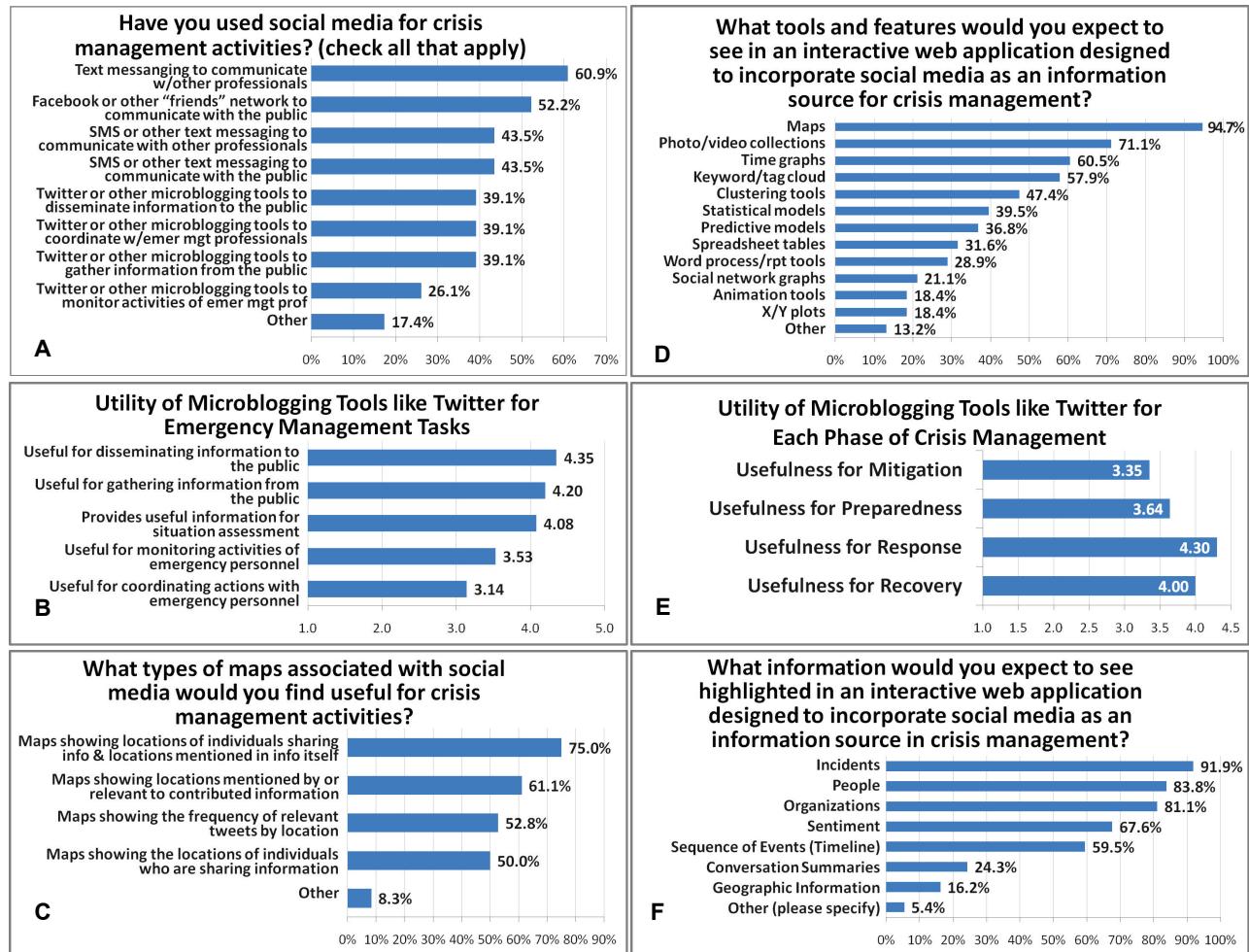


Fig. 2. Results from survey of emergency management personnel.

### **3.3 Envisioned Use of Geovisual Tools**

In the second major portion of our survey, we shifted focus toward how participants envision using geovisual tools that can take advantage of social media data sources. To begin, we asked participants to identify which types of geographic information they would find useful, and to tell us which types of maps they would expect to see in new social media enabled tools. Participants indicated that both the locations mentioned in or relevant to the contributed information (86.1%) and location of the person generating the information (80.6%) are useful for crisis management activities. Additionally, participants responded that they would find maps that show both types of geographic information at once to be maximally useful (Fig. 2c).

We then asked participants to identify the tools and features expected in an interactive, web-based application designed to incorporate social media as an information resource for crisis management (Fig. 2d). Maps (94.7%), photos/video collections (71.1%), time graphs (60.5%), keyword clouds (57.9%) and clustering tools (47.4%) were among the most popular choices.

Further, we asked participants to identify types of information they would expect to be able to analyze in the same type of application (Fig. 2f). Incidents (91.9%), sequence of events (83.8%), geographic information (81.1%), people (67.6%), and organizations (59.5%) were the most expected information types.

In a final series of questions, we asked participants to review a graphical mockup of the SensePlace2 interface (Fig. 3). First, we asked participants to describe how they might use such an application. 14 of the 25 total responses collected for this question highlighted its potential use to support situational awareness activities. The next most popular response category (4 responses)

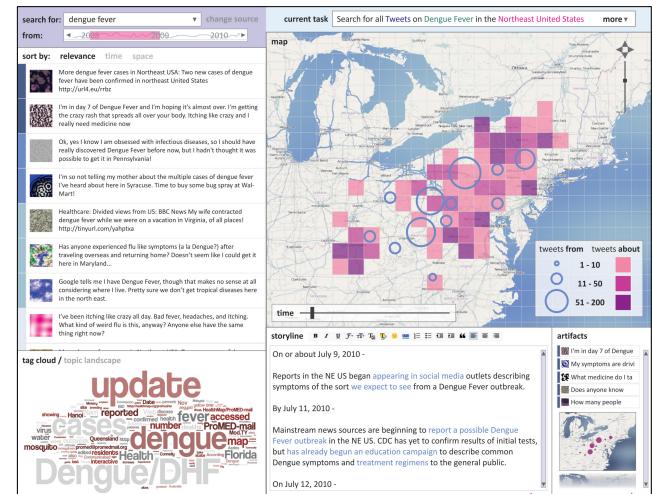


Fig. 3. SensePlace 2 mockup used in survey.

focused on its potential to support decision making during a crisis. The remaining five responses called for narrower use cases (logistical support, for example) or did not highlight a use case because they were unclear about the application's purpose.

Next, we asked participants what types of information or results they would expect to be able to share with colleagues if they were to use an application like SensePlace2. This poll revealed that printable maps (97%), a link that would launch the application with preloaded data (63.6%), pre-formatted text reports (51.5%) and static screen captures (51.5%) were preferred (Fig. 4).

Finally, we asked participants to tell us which tools/features they would add to our design, and which tools/features they suggest could be removed. We collected 18 responses to the former question, and the most popular suggestions (6 responses) were tools that would enhance data interoperability with additional media sources (photos and streaming video), and import/export capability to support flexible information sharing. The next most common category of new feature suggestions focused on enhanced reporting capabilities to create written reports and disseminate those quickly/easily to other analysts (4 responses). We also noted suggestions for more sophisticated mapping tools (3 responses) and basic interface controls to resize windows and exchange visualization elements (2 responses).

We collected 13 responses to our question asking participants to highlight features they would remove from our design, and the most popular suggestion was to remove the tag cloud (6 responses). In second place (5 responses) our participants said they would not change anything.

The results of our survey suggest that those in the emergency management domain are actively using social media tools, that they expect to be able to explore multiple kinds of geography associated with social media, and that they anticipate making use of this information primarily for SA activities. Our results provide a roadmap for current and future design iterations of SensePlace2 that include functionality for mapping tools, data interoperability and ingest features, and analytical reporting capabilities.

#### 4 APPLICATION DESIGN AND IMPLEMENTATION

This section introduces our design process and initial system implementation and testing. First, we outline the scenario-based design (SBD) development process applied [4], initial SensePlace2 functionality it prompts, and present an SBD claims analysis focused on key design choices. Second, we detail the tweet processing system. Third, we report on initial validity assessment of that tweet processing.

##### 4.1 SBD & SensePlace2 Functionality

Our tool and interface design approach follows SBD principles [3]. SBD is a method and set of techniques through which future system use is concretely described early in development. Development is guided by narrative descriptions of envisioned use episodes to achieve a system enabling these use experiences. The

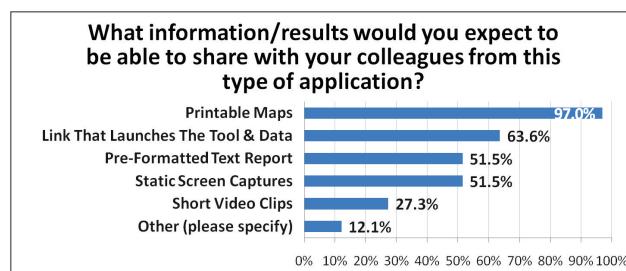


Fig. 4. Responses about sharing information.

process includes a sequence of four scenario types: problem, activity, information, and interaction. Space does not permit a detailed discussion of all SBD scenarios used in our design process; one led to the mockup assessed in the practitioner survey reported above, a sequence of others led to the system as detailed below. We present one composite, abbreviated scenario that we use here: (a) to illustrate design goals (derived from past work with practitioners, practitioner survey input, and review of scientific literature and current practice) and (b) as a mechanism (together with the accompanying video demonstration of capabilities) to validate the SensePlace2 implementation.

##### 4.1.1 SBD example scenario

In our scenario (demonstrated with SensePlace2 and real data), a hypothetical emergency management professional (Karen) working for a hypothetical international emergency response organization (the International Crisis Response Agency - ICRA) has been tasked with monitoring Twitter feeds from the U.S. and neighboring countries as input to SA during the aftermath of the Haiti 2010 earthquake. One primary ICRA objective that Karen is focusing on is to identify activities and attitudes related to the relief effort that might not be included in official situation reports from the government agencies and NGOs involved. She is particularly interested in efforts to gather and send relief supplies and to provide support for refugees who are displaced by events. In relation to relief supplies, many organizations as well as individuals will attempt to contribute supplies, provide logistical support, and even to travel to the crisis location to help. The donation and delivery of supplies by volunteers can be ineffective or even disruptive since response organizations may be unaware of volunteer efforts until those efforts have generated too many of some supplies, too few of others, and perhaps have even tried to ship supplies before facilities are available to accept them.

Karen starts her situational assessment with an initial overview of all tweets from the past 6 months that include "Haiti" as a location; this provides a general context within which to interpret interpret activities. The heatmap shows considerable activity across the globe but little clear structure; thus some narrowing of focus is needed to uncover useful information on the current state of affairs in Haiti and what supplies may be needed. Karen first looks at the most relevant 500 tweets, but this does not look useful since it starts with tweets about the Haitian Ambassador giving a press conference. She begins to drill down to more useful information by dragging the timeline handle to limit the time range to 10 weeks leading up to the present. Karen then decides to start by investigating tweets about supplies related to Haiti by adding "supplies" to the query. This creates some results that look more useful; she finds information about supply shipments and by mousing over the 3<sup>rd</sup> tweet finds that "Starkville" is in Mississippi and that a ship that started there cleared customs and will arrive tomorrow in Port au Prince (illustrated in Fig. 1).

Next, Karen reviews the most recent tweets by sorting the tweet list by time. By clicking on some of the tweets at the top, she finds several tweets from "Today's News Network" that are not news to her but also finds that Yele Haiti and the Carma Foundation have partnered for medical supplies and aid (Fig. 5). Karen goes back to relevance order by clicking on the relevance column. Then Karen notices a couple locations on the map having fairly large numbers of tweets, so she clicks on the large circles in Tampa, Sarasota and Port Everglades; this brings those tweets to the top of the list. From this, she sees that a cargo ship has left Port Everglades, some supplies arrived in Sarasota, and the President of Haiti will be accepting school supplies in Tampa.

While much of the activity in moving supplies is going through Florida, Karen is also interested in understanding the status of activities in other parts of the country. So, she changes the query

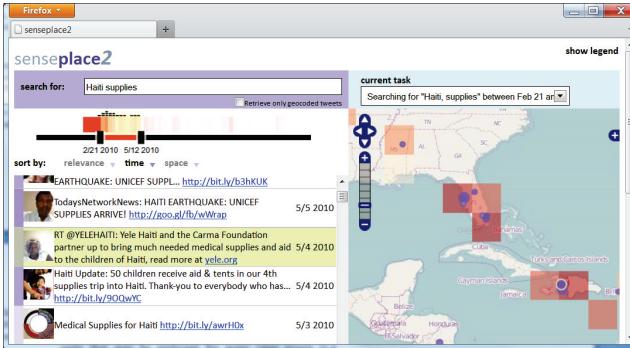


Fig. 5. Use of temporal sorting to find organizational activities.

to the somewhat broader topic of “Haiti Relief” and uses the Alt-click spatial search feature pointing at Louisville, KY to identify the 500 closest relevant tweets. She Ctrl-clicks two tweets, bringing them to the top of the list (Fig. 6). Karen is surprised at the range of fundraising activities that ICRA was not aware of.

#### 4.1.2 SensePlace2 functionality

The SensePlace2 interface includes 5 core components: query panel, timeline display/control, tweet list, tweet map, and history view. Each component is described in more detail below.

*Query Panel:* The query panel provides a minimalistic interface to specify terms for text-based queries.

*Timeline display/control:* The Timeline control supports query filtering with time range sliders. A ‘heatbar’ signifies tweet density over time using a sequential color scheme with dark red representing high frequency. An ordinal histogram (a variant of a sparkline) above the heatbar signifies normalized query frequency results, divided into low (non-zero), medium, and high.

*Tweet list:* The tweet list depicts the 500 most relevant tweets for any query. Tweet text along with accompanying metadata (posting date, user profile image and relevance ranking) is displayed as a scrollable list. The default term frequency-based Lucene relevance rating is not very discriminating for tweets. So, we use Lucene’s “boosting” (weighting) mechanism; weighting is adjustable. Our default weights (in rank order) by incidence of: places, organizations, hashtags, and dates/currency/times. To enhance list readability, a color-coded strip on each tweet depicts its relevance ranking. The list can be sorted by relevance (the default), time, and location. To support the survey request for access to photos and video, URLs in tweets are live links.

*Tweet map:* Following logic outlined in the background section, the tweet map supports simultaneous tweet distribution overview and detail plus an ability to select places or apply spatial filtering. A *heatmap* provides the overview (gridded with transparent fill using the same sequential color scheme as on the timeline). It depicts tweet frequency filtered by concept, time, and place query specifications; frequency is grouped into 5 quantile categories. Ideally, tweet frequency would be normalized. But there is no clear method for normalization that will be consistently valid. Options include: total population per cell (this can be obtained from the global population grid data set but those data have zero values for ocean cells even though tweets can come from ships within sight of some event); twitter accounts/cell (a value that is probably not possible to obtain); or the ‘at risk’ population (e.g., if the focus is on flooding, normalization could calculate population at specific elevations). Thus, our initial implementation uses raw frequency per cell. A gridded heatmap is used rather than a choropleth map (in which political unit shading represents data magnitude) because: (a) choropleth maps are misleading if counts rather than normalized data are depicted; with no clear choice on how to normalize data, territories are normalized through a regular tessellation; (b) a grid map provides a better overview because it

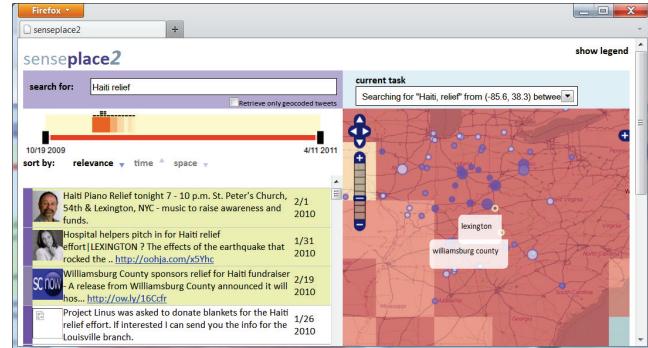


Fig. 6. Use of spatial query to find regional activities.

avoids the visual distraction of political units (which are irrelevant for many crisis events). Locations for the top 500 tweets are highlighted with graded circle sizes (small = 1-5, medium = 6-15, large = >15). To meet needs identified in the survey, from and about location is distinguished, with *from* in green and *about* in purple. Clicking on a location highlights linked tweets and moves them to the top of the tweet list. Ctrl-click allows multiple locations to be selected. The map has normal pan and zoom controls but also serves as an interface to spatial filtering; Alt-click finds the 500 relevant tweets closest to the clicked location.

*History view:* Whereas the query panel, timeline and tweet list define, display and refine query content, the history view stores queries for reuse (a step toward the “launch with preloaded data” request from the survey). Currently this view provides a dropdown menu of each query as a combination of thematic, temporal and spatial filters used and presented in natural language terms for easy review.

A new query can be initialized in two ways - by using the query panel to enter new key words or by using the timeline or spatial query tool to set the time frame or locality of study. In either case, as soon as the new query is launched it is recorded and displayed in the history view. After the query returns, the timeline control, tweet list and tweet map are updated. An existing query can be refined in three ways. First, key words can be adjusted, second, a new spatial query can be launched to refine the original location, and third, timeline controls can be adjusted. Similar to launching a new query, those changes will be recorded in the history view and reflected in timeline control, tweet list and tweet map.

#### 4.1.3 SensePlace2 SBD claims analysis

SBD claims analysis identifies system features expected to have important consequences for users. Claims about the features represent both a justification for initial design decisions and testable contentions to follow up on in subsequent user testing. Figure 7 details key claims for SensePlace2 as presented here.

#### 4.2 SensePlace2 Architecture

A primary task to support SA is collecting tweets about specific topics relevant to crisis management, i.e. earthquake, tsunami, etc, and then storing these tweets in a format conducive to interactive analysis. To achieve these goals we wrote a distributed twitter crawler integrated with a postGRES database for storage. Our crawler utilizes the Twitter API that returns relevant tweets to user-input keywords (“earthquake”, ‘cyclone’, etc.) within a seven-day period. The API calls return up to 1000 relevant tweets and auxiliary metadata (e.g., creation time, tweet id, user id etc.) in JSON format. Data are parsed, cleaned, and written to a postGRES database. When tweets corresponding to a keyword are stored, we also store the most recent tweet id retrieved for a keyword. We then also query the Twitter API using the keyword and the most recent tweet id for the 1000 most recent tweets. In

Organizing the interface to support query and display using separate map, timeline, and text views
+ fits with a standard conceptualization of the world into place, time, and theme components
+ enables what, where, when questions
+ provides context for events that are happening in place
- may under-represent important people and organizations
- may emphasize dimensions of events and processes over the events and processes themselves
The combination of heatmap depicting overall frequency and graduated circle depiction of the relevant from & about tweets
+ provides integrated overview + detail
+ allows an analyst to quickly explore places w/relevant info
- may hide relevant tweets that do not make the 500 cut-off
- may be confusing to analysts who are only familiar with typical web maps with limited use of abstract symbols
The heatmap and ordinal histogram combination
+ supports simultaneous overview + detail allowing analysts to understand the temporal context for relevant query results
+ allows analysts to quickly zero in on key events
- may prompt analysts to focus too much attention on active times / events versus the most important tweets
Analytical recall supported by natural language query list
+ supports recall of past search parameters using easy to understand language
+ helps the analyst understand the fundamental questions asked by combinations of interactions
- may not help analysts quickly browse to identify particularly important/unimportant results
Keyword search as an input to exploration and analysis
+ is a method that most users will expect and one that requires little or no training to understand
+ matches the fact that the underlying data is a text corpus
- relies on input stemming from analyst interests and assumptions, rather than on significant patterns that have come from an objective analysis of the entire dataset

Fig. 7. SBD Claims analysis for SensePlace2.

addition timestamps are used to prevent old tweets from being downloaded on subsequent queries to the API. Note that since Twitter assigns a unique ID to each retweet, any retweets are stored as unique database entries.

The SensePlace2 *Text Information Channel Backend* is depicted in Figure 8. Due to Twitter call limitations, the crawler typically sends approximately 150 keyword requests per hour. Once tweets are loaded into the database, separate applications analyze tweets for coordinates (depicting where tweets are from) and named-entities such as locations, organizations, persons, hashtags, URLs etc (depicting what and where tweets are about). Next extracted entities are written to separate location and entity tables. Lastly, locations extracted are georeferenced using Geonames.

Approximately 1% of tweets include coordinates. While this is a small proportion, it has gradually increased from less than 0.5% in Nov. 2009. It is also likely that Twitter users will be willing to turn location on during a crisis event and there is evidence that efforts to develop crisis-specific hashtags can have an impact in real events [35], and these hashtags include location. Thus, ‘from’ locations are likely to be more frequent in future crises.

The overall system architecture of the *Twitter Information Channel storage and processing backend* (Fig. 9) consists of a central database that contains data tables, associated named entity tables, location tables, and text search indices for each information channel that support both the text processing tools and the UI. The architecture is designed such that a wide variety of information channels can be easily integrated and allow the

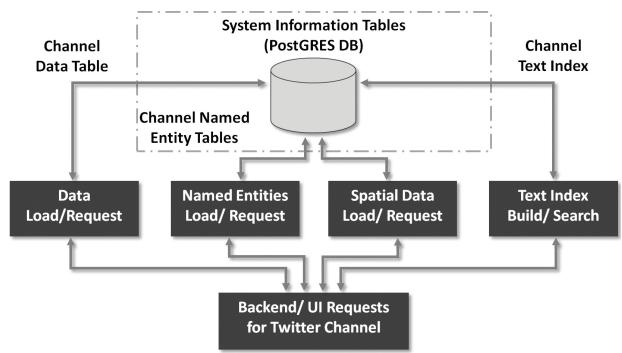


Fig. 8. SensePlace2 Text Information Channel Backend.

user interface/ processing software to quickly access and process this information. Each channel of information that is currently stored consists of four backend processing daemons:

1. *Channel Search/ Data Load Daemon*: Each channel has an associated daemon that collects information of interest. For example, the Twitter channel search daemon currently uses a table with queries passed on to the Twitter Search API. Results are processed as described above. Each channel data table typically contains a single text column of interest (in this case the text within tweets returned from a search). This column information is stored in the system information table as shown in Figure 8 as it will be used by the text processing software.

2. *Named Entity Extraction Daemon*: The named entity extraction daemon extracts the named-entities contained within the text present for each channel and writes this information to the associated named-entity tables in the database. Four broad categories of named entities are extracted: people, locations, organizations and others. The current daemon uses a custom developed ANNIE named-entity extractor based on GATE [36]. The named-entity extraction daemon is a distributed application that runs on multiple machines allowing a relatively large volume of text to be processed in real time.

3. *Geocoding Daemon*: Named-entity extraction generates a location table containing text information for locations within a channel (e.g. “Haiti”). While this is suitable for text querying, it is insufficient to perform spatial querying on this channel, e.g., to find the closest tweets to a location. The Geocoding daemon processes the location table (for all channels) and geocodes locations with coordinates of the best matching location returned from the Geonames.org web service. These point coordinates are stored and indexed within the location table allowing for quick generation of heat maps and retrieval of relevant tweets in a desired location neighborhood.

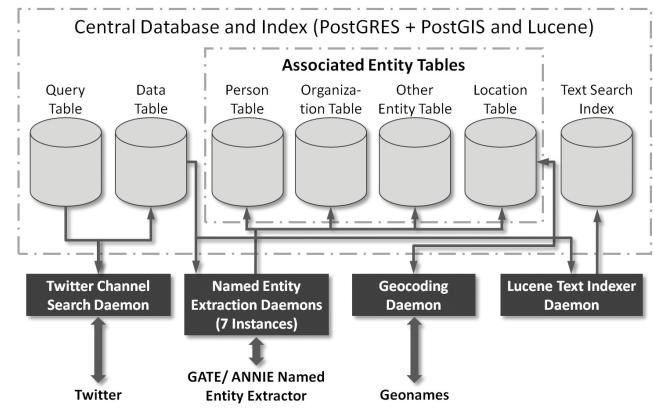


Fig. 9. Twitter info channel storage & processing backend.

**4. Lucene Text Indexer Daemon:** As discussed in #1, each channel typically contains a single text column in the channel data table. For example, for the twitter channel the data table contains one column that stores all tweets that were extracted by the channel search daemon. The Lucene text indexer daemon creates a fielded text index for the channel text for full text querying as well as for more advanced retrieval of relevant tweets within a date range. The main fields are text, location, organization, people and other entities; boosting of selected fields (e.g., location) is used to improve relevance ranks.

Fig. 10 and 11 show tweets loaded by the Twitter Channel Search daemon and the corresponding entities extracted by the Named Entity Extraction daemon (Locations and Organizations/ Types, tools also extract persons). Fig. 12 shows the location table for the Twitter channel and the spatial coordinates that were assigned to the locations by the Geocoding daemon.

### 4.3 Tweet Entity Extraction Accuracy

To measure entity extraction, we first found a suitable query “Haiti AND earthquake” and searched the database for tweets that *matched* and had been *processed* by the entity extractor and ordered them by date. The choice of this keyword was specifically due to the fact that all tweets matching “Haiti AND earthquake” will contain at least one location namely “Haiti” and will provide a suitable experimental dataset for measuring entity extraction accuracy. Output from the dates 2010-01-12 to 2010-02-18 matching this query is shown in Fig. 13.

A suitable date was then chosen such that the total tweet count is manageable for hand verification. We chose the date 2010-02-05, which corresponded to a total of 1074 tweets that had been processed by the entity extractor; we extracted these for further processing. Each tweet was then checked to confirm whether the location ‘haiti’ was extracted. In addition, if other locations were present they were marked as either correctly or incorrectly extracted. Table 1 illustrates location extraction results.

The column ‘multiple locations only’ represents location extraction results from tweets having multiple locations. The ‘all locations’ column represents location extraction results from all tweets having at least one location, i.e., all 1074 tweets. In each, the sub-column column ‘all’ represents tweets where all locations were correctly extracted. The column ‘partial’ represents the tweets where at least one of the many locations was extracted. The column ‘incorrect’ represents the tweets where at least one location was incorrectly extracted.

Since retweets were included above, results may have been biased as a result of correct or incorrect location extraction of a

A	B	C	D
sysid	tweettext	name	type
3	21 www.mrlfest8.com The H1N1 vaccine is still in short supply, as despite U.S. government assure	united states	country
4	22 RT @Doc_D: Reading about h1n1 in pregnancy. Any cases at UW?	reading	city
5	26 Sunnow Tv Show --- H1N1 Spread Prevention [WBNGTV]. News Sports. Weather Binghamton, new york	new york	province/estate
6	32 Mother of Kansas girl who died in Indu during FFA convention says she later tested positive for H1N1 kansas	kansas	province/estate
7	48 H1N1 Flu Virus: Government Of Canada Provides Travel Vaccines	canada	country
8	52 China: AH1N1 accounts for 80 pct of China's total flu cases, official http://bit.ly/mlymCa#t	china	country
9	60 H1N1 flu vaccine shortage impacting public flu clinics in Baltimore - http://tinyurl.com/h5bpsi#t baltimore	baltimore	city

Fig. 9. Example tweets, extracted locations, and types.

A	B	C	D	E	F	G	
sysid	tweettext	name	type	sysid	geocoded	geonameid	coordinates
1058	HHS waives certain Medicare, Medicaid, CHIP requirements for H1N1 flu. <a href="http://ow.ly/xkDAfb">http://ow.ly/xkDAfb</a>	medicare	organization				
1063	Further rise in swine flu cases (BBC News). <a href="http://bit.ly/kjkSfr">#h1n1</a>	bbc	company				
1066	Emergency approval from FDA for new swine flu test (AP via Yahoo! News). <a href="http://bit.ly/lK7Im7">#h1n1</a>	fda	organization				
1066	Emergency approval from FDA for new swine flu test (AP via Yahoo! News). <a href="http://bit.ly/lK7Im7">#h1n1</a>	yahoo	company				
1070	Health and Human Services Announces EPA Registered Hard Surface Disinfectant Formulations	environmental protection agency	organization				
1077	North Canton Company Fights Against H1N1 with New EPA Registered Hard Surface Disinfectant, Formula#t environmental protection agency	environmental protection agency	organization				
1096	The line-up at NY Civic Centre at 11:48 for H1N1 shot. Cops said 4 - 5 wait. <a href="http://pic.gd/972f1">http://pic.gd/972f1</a>	my civic centre	organization				

Fig. 10. Example Tweets, extracted organizations and types.

A	B	C	D	E	F	G
locid	name	type	sysid	geocoded	geonameid	coordinates
3	19401 pakistan	country	97002 t	1168579 (30.70)		
4	19402 florida	province/estate	97004 t	1455751 (28.750408-82.5000976)		
5	19407 united states	country	97008 t	6252001 (39.76,-98.5)		
6	19408 france	country	97008 t	3017382 (46.2)		
7	19416 turkey	country	97018 t	298795 (39.35)		
8	19417 delhi	city	97019 t	1261481 (28.612818962100977,23.11401367188)		

Fig. 11. Example Locations extracted from the Tweets and their Geonames IDs and latitude, longitude coordinates.

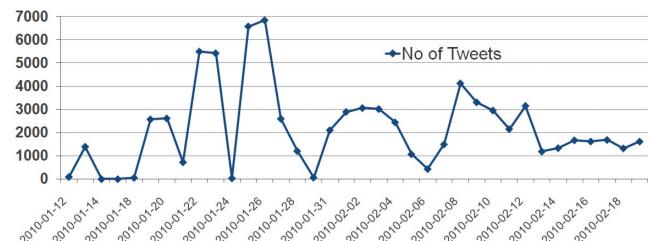


Fig. 8. Time plot of tweets with Haiti AND earthquake.

tweet that has high retweets in the dataset. To overcome this problem, the initial dataset was cleaned to remove tweets having greater than 75% overlap. To achieve this, the Overlap Coefficient was used [37]. The initial dataset containing 1074 tweets was reduced to 627 tweets using string similarity. Table 2 illustrates location extraction accuracy results on this smaller dataset.

Table 1. Location extraction results

	'Haiti'	Multiple Locations Only			All Locations		
		Extraction	All	Partial	Incorrect	All	Partial
% correct	92.2	74.4	21.1	0.04	89.2	4.4	1.3
Extracted Correctly	992	166	47	1	958	47	14
Total	1074	223	223	223	1074	1074	1074

Table 2. Location accuracy on cleaned dataset

	'Haiti'	Multiple Locations Only			All Locations		
		Extraction	All	Partial	Incorrect	All	Partial
% correct	91.4	69.2	24.2	0.08	88.2	4.6	1.4
Extracted Correctly	573	83	29	1	553	47	9
Total	627	120	120	120	627	627	67

## 5 SENSEPLACE2 ILLUSTRATIVE USE-CASE

In this section, we present a brief second use-case. Together with the demonstrated ability of SensePlace2 to support the design scenario outlined above, this use-case provides evidence toward validation of the place-time-theme approach to use of Twitter for SA support. This use-case reports on our application of SensePlace2 to understand the situation throughout the BP oil spill in spring of 2010. The use-case illustrates additional tool functionality added in response to practitioner survey input.

First, a query for “oil, slick” was used to begin investigating concerns about the impact of oil being generated by the uncapped BP well. Not surprisingly, initial concerns were focused on the distance of the slick to various locations and the potential impact on local economies. To obtain a more localized picture, the query was restricted to tweets with coordinate location (enabling a distinction between *from* and *about* locations matches the top geospatial request in our survey). Several locations along the Gulf Coast are highlighted using the Ctrl-click feature (Fig. 14). In addition to identifying some localized concerns about impacts on beaches, this result also highlights activities related to response personnel; these include a job notice and a report about hospitalization of clean-up crew members.

An additional kind of functionality called for by practitioners was access to photos and video. While plans exist (see below) to extend SensePlace2 to support processing of diverse information sources, as a step toward that goal we added the ability for URLs within tweets to be live links. The functionality has proved to be very useful in enabling a user to quickly build a comprehensive picture of an evolving situation.

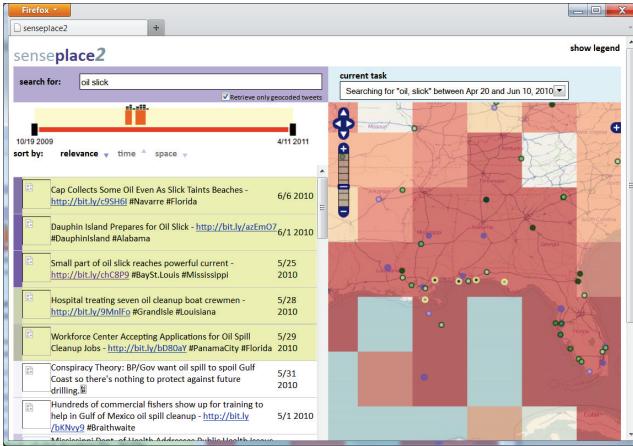


Fig. 12. Query restricted to tweets with geo-location; several coastal locations selected and tweets promoted in list.

As one example, we used SensePlace2 functionality to understand implications of the spill for birds and bird habitat (starting with the query “oil, birds”). One tweet link led to a FOX news story, “Oil Spill’s Long Reach,” which included a video with the lead scientist from the Exxon spill cleanup contending that the BP oil would most likely be consumed by bacteria and cause no real problems. This story was countered by a tweet linked to a web site with many photos of birds covered in oil (by an AP photographer) and another linked to a YouTube video showing the oil close to shore off Pensacola, FL and others with photos of oil on beaches there. These bits of information together with the many tweets about oil position and impacts provided context for a detailed story in the Natural Resources Defence Council Staff Blog (also linked from a tweet) that detailed the potential threat to breeding and migrating birds (Fig. 15). Overall, broad-based SA related to impact of the oil spill on birds was developed in approximately one hour of work. This brief use case provides a second proof-of-concept demonstration to complement the success demonstrated above and in the accompanying video in supporting the design scenario that guided SensePlace2 development.

## 6 DISCUSSION AND FUTURE RESEARCH

We have reported on Phase-I of ongoing research to develop geovisual analytics methods that leverage social media for situational assessment in crisis management and related domains where SA is a core activity. Results of our formal survey of emergency management professionals indicate a clear openness

among professionals in these domains to use social media as part of SA activities as well as for other aspects of emergency/crisis management. The initial version of SensePlace2 demonstrates the potential of using a place-time-concept framework to support SA in this domain. The scenario-based design approach applied to development provides an efficient strategy to refine design and a framework for subsequent user studies to inform ongoing design.

As noted in the introduction, SA support should not rely on Twitter alone. One reason is that Twitter is biased toward age ranges that overemphasize 25-54 year olds (<http://tiny.cc/f948m>). For crisis management SA, the demographics of those reporting events may be of less importance than when Twitter is analyzed as a source of opinion (e.g., about political events). However, even if user demographics are not critical, comprehensive SA requires diverse information sources. Thus, SensePlace2 has been designed to be extensible to ingest data from many other sources.

Planned next steps in the research (several of which respond directly or indirectly to feedback in the survey reported above) include addition of: (a) support for more kinds of social media including text from blogs, RSS feeds, and tagged photos from photo sharing sites, (b) a flexible mechanism to normalize tweet frequency for the heatmap based on any gridded data layer that an analyst considers appropriate, (c) a non-tag cloud method to represent salient concepts extracted from tweets (our original design included a tag cloud, but it was removed in response to participant reactions), (d) a *Storylines* tool that enables analysts to document and share fragments of information they consider important to overall SA along with report generation tools for any information synthesis they develop over time, and (e) cloud-based core document processing capabilities to enable live streaming and scaling to much larger data sets. For the latter, we have about 20 million tweets in our database currently but only 3 million for 2010 have been indexed so that queries are practical. We are now in the process of migrating from data collection using the Search API to a mix of data collection using both the Search and the Streaming API. Specifically the Search API will be used for seeding the system with past data while the Streaming API will be used for collecting new data. A cloud-based implementation is expected to allow real-time updates to the system so that it is possible to support SA for events as they happen.

## ACKNOWLEDGEMENT

This material is based upon work supported by the U.S. Department of Homeland Security under Award #: 2009-ST-061-CI0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security.

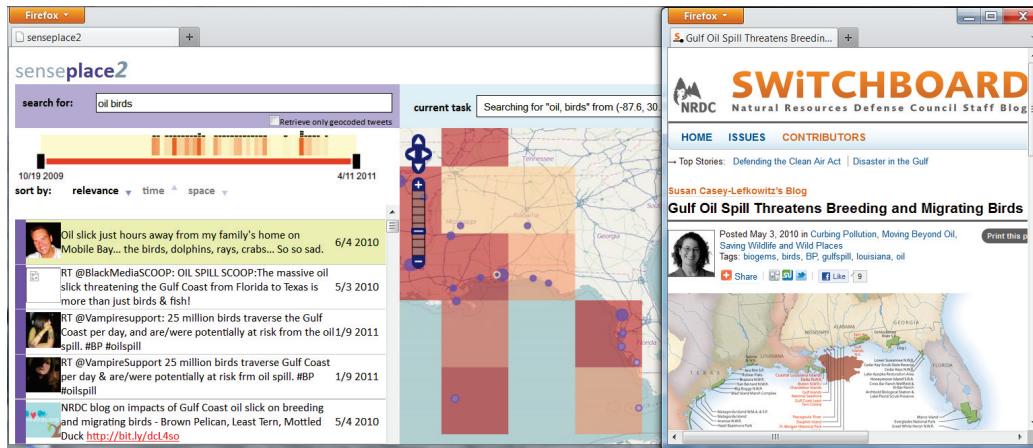


Fig. 13. Highlighted location of first-hand knowledge of slick position reported in tweet with geo-location (left) and results of following link on last tweet to story on bird impacts.

## REFERENCES

- [1] A. Nelson, I. Sigal, and D. Zambrano, "Media, Information System and Communities: Lessons from HAITI," InterNews, Communications with Disaster Affected Communities, Knight Foundation, Miami, FL Jan. 10, 2011.
- [2] S. Vieweg, A. Hughes, K. Starbird, and L. Palen, "Microblogging during two natural hazards events: what twitter may contribute to situational awareness," in *Proc. of the 28th Inter. Conference on Human Factors in Computing Systems*, 2010, pp. 1079-1088.
- [3] M. B. Rosson and J. M. Carroll, "Scenario-based design," in *Chapter 53 in The Human-Computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications*, J. Jacko and A. Sears, Eds.: Lawrence Erlbaum Associates, 2002, pp. 1032-1050.
- [4] J. M. Carroll and M. B. Rosson, "Getting around the task-artifact cycle: how to make claims and design by scenario," *ACM Transactions on Information Systems*, vol. 10, pp. 181-212, 1992.
- [5] M. R. Endsley, "Toward a theory of situation awareness in dynamic systems," *Human Factors*, vol. 37, pp. 32-64, 1995.
- [6] G. Klein, B. Moon, and R. R. Hoffman, "Making Sense of Sensemaking 1: Alternative Perspectives," *Intelligent Systems, IEEE*, vol. 21, pp. 70-73, 2006.
- [7] P. Pirolli and S. Card, "The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis," in *2005 Inter conf on intel. analysis* McLean, VA, 2005.
- [8] F. L. Greitzer, A. Schur, M. Paget, and R. T. Gutstromson, "A sensemaking perspective on situation awareness in power grid operations," in *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008 IEEE*, 2008, pp. 1-6.
- [9] D. Sinton, "The inherent structure of information as a constraint to analysis," in *Harvard Papers on Geographic Information Systems*, D. Dutton, Ed. Reading: Addison Wesley, 1978.
- [10] D. Peuquet, "Its About Time - a Conceptual-Framework for the Representation of Temporal Dynamics in Geographic Information-Systems," *Annals of the AAG*, vol. 84, pp. 441-461, Sep 1994.
- [11] D. Peuquet, *Representations of Space and Time*. NY: Guilford, 2002.
- [12] N. Andrienko, G. Andrienko, and P. Gatalsky, "Exploratory spatio-temporal visualization: an analytical review," *Journal of Visual Languages & Computing*, vol. 14, pp. 503-541, 2003/12 2003.
- [13] N. Andrienko and G. Andrienko, *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Berlin: Springer-Verlag, 2005.
- [14] Y. Livnat, J. Agutter, S. Moon, and S. Foresti, "Visual Correlation for Situational Awareness," in *Proceedings, IEEE Symposium on Information Visualization*, Minneapolis, MN, 2005, pp. 95-102.
- [15] B. M. Tomaszewski, A. C. Robinson, C. Weaver, M. Stryker, and A. M. MacEachren, "GeoVisual Analytics and Crisis Management," in *ISCRAM Conference*, Delft, the Netherlands, 2007, pp. 173-179.
- [16] B. Tomaszewski, "Situation awareness and virtual globes: Applications for disaster management," *Computers & Geosciences*, vol. 37, pp. 86-92, 2011.
- [17] B. Tomaszewski, J. Blanford, K. Ross, S. Pezanowski, and A. MacEachren, "Supporting geographically-aware web document foraging and sensemaking," *Computers, Environment and Urban Systems*, vol. 35, pp. 192-207, 2011.
- [18] S. Kim, Y. Jang, A. Mellem, D. S. Ebert, and T. Collins, "Visual Analytics on Mobile Devices for Emergency Response," in *VAST 2007*, Sacramento, CA, 2007, pp. 35-42.
- [19] J. Kohlhammer, T. May, and M. Hoffmann, "Visual Analytics for the Strategic Decision Making Process," in *GeoSpatial Visual Analytics*, 2009, pp. 299-310.
- [20] M. R. Endsley, "Theoretical underpinnings of situation awareness: A critical review," in *Situation awareness analysis and measurement*, M. R. Endsley and D. J. Garland, Eds. Mahwah, NJ: Lawrence Erlbaum Associates, 2000, pp. 3-32.
- [21] A. Livinski and L. D. Passman, "Web 2.0 for Emergency Response: Emerging Opportunities for Greater Situational Awareness," in *slides from presentation at: Annual Meeting of the Medical Library Association (MLA)*, Washington, DC, 2010.
- [22] T. Sakaki, M. Okazaki, and Y. Matsuo, "Earthquake shakes Twitter users: real-time event detection by social sensors," in *Proc. of the 19th Inter. Conference on World Wide Web*, 2010, pp. 851-860.
- [23] L. Palen, S. Vieweg, and K. M. Anderson, "Supporting "Everyday Analysts" in Safety-and Time-Critical Situations," *The Information Society*, vol. 27, pp. 52-62, 2011.
- [24] A. Mills, R. Chen, J. Lee, and H. R. Rao, "Web 2.0 Emergency Applications: How Useful can Twitter be for Emergency Response," *Journal of Information Privacy & Security*, vol. 5, pp. 3-26, 2009.
- [25] A. L. Hughes and L. Palen, "Twitter adoption and use in mass convergence and emergency events," *International Journal of Emergency Management*, vol. 6, pp. 248-260, 2009.
- [26] S. Sinnappan, C. Farrell, and E. Stewart, "Priceless Tweets! A Study on Twitter Messages Posted During Crisis: Black Saturday," in *21st Australasian Conference on Information Systems Twitter messages posted during Black Saturday*, Brisbane, 2010.
- [27] O. Oh, M. Agrawal, and H. Rao, "Information control and terrorism: Tracking the Mumbai terrorist attack through twitter," *Information Systems Frontiers*, pp. 1-11, 2010.
- [28] M. Zook, M. Graham, T. Shelton, and S. Gorman, "Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake," *World Medical & Health Policy*, vol. 2, p. Article 2, 2010.
- [29] M. Dörk, D. Gruen, C. Williamson, and S. Carpendale, "A Visual Backchannel for Large-Scale Events," *IEEE Transaction on Visualization & Computer Graphics*, vol. 16, pp. 1129-1138 2010.
- [30] D. Shamma, C. Santa Clara, L. Kennedy, and E. Churchill, "Tweetgeist: Can the Twitter Timeline Reveal the Structure of Broadcast Events?", in *CSCW, 2010*, Savannah, GA, 2010.
- [31] N. Diakopoulos, M. Naaman, and F. Kivran-Swaine, "Diamonds in the Rough: Social Media Visual Analytics for Journalistic Inquiry," in *IEEE Conference on Visual Analytics Science and Technology (IEEE VAST 2010)*, Salt Lake City, Utah, USA, 2010, pp. 115-122.
- [32] G. Andrienko, N. Andrienko, M. Mladenov, M. Mock, and C. Pöltz, "Discovering Bits of Place Histories from People's Activity Traces," in *IEEE Conference on Visual Analytics Science and Technology (IEEE VAST 2010)*, Salt Lake City, Utah, USA, 2010, pp. 59-66.
- [33] N. Andrienko and G. Andrienko, "Spatial generalisation and aggregation of massive movement data," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, pp. 205-219, 2011.
- [34] K. Field and J. O'Brien, "Cartobigraphy: Experiments in Using and Organising the Spatial Context of Micro-blogging," *Transactions in GIS*, vol. 14, pp. 5-23, 2010.
- [35] K. Starbird and L. Palen, "Tweak the Tweet: Leveraging Microblogging Proliferation with a Prescriptive Grammar to Support Citizen Reporting," in *Proceedings of the 7th International ISCRAM Conference (Short Paper)*, Seattle, WA, 2010.
- [36] D. H. Cunningham, D. D. Maynard, D. K. Bontcheva, and M. V. Tablan, "GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications," 2002.
- [37] C. D. Manning, H. Schütze, and MITCogNet, *Foundations of statistical natural language processing* Cambridge, MA: MIT Press, 1999.

<sup>i</sup> SensePlace2 was introduced in a short paper that sketched the crisis management motivation for the tools and described initial functionality for a partial implementation of the query and filtering tools that are developed and described more fully here. (MacEachren, A.M., Robinson, A.C., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J. and Mitra, P. in press: Geo-Twitter Analytics: Applications in Crisis Management. *25th International Cartographic Conference*, Paris, France.)