Transactions in GIS

Transactions in GIS, 2011, 15(6): 753–773

Research Article

Geo-parsing Messages from Microtext

Judith Gelernter
Language Technologies Institute
School of Computer Science
Carnegie-Mellon University

Nikolai Mushegian
Language Technologies Institute
School of Computer Science
Carnegie-Mellon University

Abstract
Widespread use of social media during crises has become commonplace, as shown by the volume of messages during the Haiti earthquake of 2010 and Japan tsunami of 2011. Location mentions are particularly important in disaster messages as they can show emergency responders where problems have occurred. This article explores the sorts of locations that occur in disaster-related social messages, how well off-the-shelf software identifies those locations, and what is needed to improve automated location identification, called geo-parsing. To do this, we have sampled Twitter messages from the February 2011 earthquake in Christchurch, Canterbury, New Zealand. We annotated locations in messages manually to make a gold standard by which to measure locations identified by a Named Entity Recognition software. The Stanford NER software found some locations that were proper nouns, but did not identify locations that were not capitalized, local streets and buildings, or non-standard place abbreviations and mis-spellings that are plentiful in microtext. We review how these problems might be solved in software research, and model a readable crisis map that shows crisis location clusters via enlarged place labels.

1 Introduction

1.1 Why Social Media for Disaster Relief?

The use of social media during disasters was established by the time of the Haiti earthquake in January 2010, when local and international posts to Facebook, Flickr, YouTube and Twitter were voluminous (Gao et al. 2010). In America, too, the use of social media during disasters continues. A poll that asked Americans whom they would contact in an emergency found that 28% would turn to Twitter for help if unable to reach 911 (American Red Cross 2010). In response, the primary emergency response organi-

Address for correspondence: Judith Gelernter, Language Technologies Institute, School of Computer Science, Carnegie-Mellon University, 5000 Forbes Ave., Pittsburgh, PA 15213, USA. E-mail: gelern@cs.cmu.edu

doi: 10.1111/j.1467-9671.2011.01294.x

zation in the U.S., the American Red Cross, hosted emergency social media summits in 2010 and 2011 to convene government agencies, emergency management professionals, disaster response organizations, technology companies and citizens to address how to reply more effectively to digital cries for help. Our research does the same.

Twitter is one of many social media microblog platforms.[1] We experiment with Twitter because it makes a quantity of data freely available to researchers. We find disaster messages by following the crosshatches that accompany a terms (for example, #eqnz for earthquake New Zealand) on a tweet of that topic.

Twitter allows the freedom of writing about any topic on *micro* scale, in that the messages are limited to 140 characters (or about 24 words). Twitter messages, called tweets, are time-stamped automatically with Greenwich Mean Time. Messages are archived by the company. Anyone with or without a Twitter account can search through very recent messages at Twitter.com. Twitter users receive tweets posted in real time from people or organizations they are "Following," and their messages are shared with those who have signed up as their "Followers." Direct tweets may be sent to another user by beginning the message with @[person's name].

Twitter is particularly attractive for crisis communication because:

1. Twitter is easy to use, cell-phone compatible, and is becoming widely familiar.
2. Twitter handles high volume data and is robust during times of crisis. Note a message posted about 12 hours after the earthquake shook New Zealand:
   `RT @radioovermoscow: It seems every news site in the country is down. Twitter is now it for news on the quake. #eqnz`
3. Twitter is up-to-the minute. A crisis template has been circulated to make it easier to process and re-distribute information from a tweet. The template uses hashtags for #imok [name – "I am ok"], #contact, #city, #addy (address or cross street), #floor, and others (Starbird and Stamberger 2010). But the template has hardly been used by tweeters thus far, and information has at times been put into a more palatable format by "voluntweeters" (Starbird and Palen 2011).
4. Twitter allows two-way communication. Presently, getting information back to the man-on-the-street is not handled well in emergencies (Louise Comfort, Center for Disaster Management, University of Pittsburgh, pers. comm., 2010). But within the social media platform, authorities might hire someone to answer individual tweets, or respond to quell panic if they learn of a faux tweet.[2]

1.2 Purpose of Study: Location as Key to Crisis Mapping

It has been proposed that the presence of a location *within the text* of a crisis message makes that message more valuable than a message that does not have a location (Munro 2011). Munro's study using data from the Haiti earthquake of 2010 concerned messages also in Haitian Creole. But it has been found in other crises that Twitter users are more likely to pass along tweets with location and situational updates than other tweets (Vieweg et al. 2010), indicating that Twitter users themselves find location to be important. Note that a Twitter message may concern more than one location. For example, `"If you want to put supplies on a truck going from Wellington to Christchurch #eqnz details here http:\/\/bit.ly\/gzXCrv - Wigan St by 5pm Thurs".`[3]

Human recognition of city or street name is easy even when they might be unfamiliar – although impossible when the tweets fly in from around the world at a rate of over a thousand per second. Our research seeks to extract this location information automatically.

Currently, tweets are mapped from a user's location, as determined from his registration information which is appended as metadata to every tweet. Even though Twitter provides the option of having the positioning device in the user's cell phone attach a latitude and longitude to each tweet, we found that less than 1% of our sample included geo-specifics, so we could not use this parameter for the information. Others have tried to infer location based on words used in the tweets (Cheng et al. 2010, Eisenstein et al. 2010). The location of the user, however, is not essential for our work, as we are interested in the location of what the user is tweeting about.

Why put tweets on a map? Social media information about a disaster can be of significant value (Liu and Palen 2010), and maps are of proven utility for disaster response teams (Gunawan et al. 2011). Disaster management solutions using web-based geographic information systems could provide medical, police and fire relief to crisis victims (Raj and Sasipraba 2010). We advocate a simple map, along the lines of what is currently available from Ushahidi, Esri, or CrisisCommons, in which different layers of information can be added. Our map will be social media-enriched. Location clusters, or hotspots, have been used in other domains to predict crime and detect disease. In our case, hotspot crisis clusters show where problems (or at least complaints) are loudest (Shekhar et al. 2011).

This research describes one step in preparing tweets for disaster mapping. The process of creating the map includes attaching geographic locations to tweets, or geocoding, but this can be done by third-party software which is readily available.[4] Geography can then be used to cluster tweets, while individual tweets within a cluster could be presented according to timestamp.

1.3 Research Questions

We want to make citizen crisis tweets immediately usable by first responders and local authorities on an overview map. One of the steps to create such a map is to associate a specific location with words in each tweet.

> Research Question_1: Compared with manually geo-tagged tweets, how accurately does one Named Entity Recognition software identify locations?
>
> Research Question_2: What types of errors does the software make?

1.4 Plan of Work

The article unfolds by describing our disaster tweet set, and our method of finding locations in tweets. We first annotated locations manually by assigning geo-tags. We used a gold standard of the geo-tags to compare with individual annotators and with a Named Entity Recognition program that finds locations automatically. Analysis of locations that the software missed helped define the next stage of research: how to create a software that will find more locations in microtext, more accurately. Related directions of research are proposed as an invitation to others to pursue related work. We show how tweets might be visualized for easy map reading, and then review the literature before restating findings in conclusion.

2 Our Data

We selected a recent disaster in an English-speaking country so that we could test methods with mostly English tweets. On February 21, 2011, there was an earthquake on Tuesday, 12:51 p.m. local time in Christchurch in Canterbury, New Zealand.

Some people who use Twitter add indexing information to their tweet in the form of a hashtag (#). Not every tweet on a topic includes a hashtag. For our purposes, we used the earthquake New Zealand hashtag (#eqnz) to find tweets we knew to be relevant. A doctoral student at our university has been archiving tweets from 2009 at the low sample rate that the company makes publicly available, and we downloaded #eqnz tweets from this archive. Our data set came to 1,490 tweets, of which 1,407 were in English. Our tweets date mostly between February 22nd and 26th 2011. Each tweet includes user metadata from that person's registration, as well as tweet-related information of day, date and time. Our data thus represents a tiny subset of the Twitterverse from the days following the Christchurch quake.

3 Method

We needed to define a location in a tweet so that we could use that definition for consistency among annotators looking for locations manually. Separate annotations by three people were used to create an "answer sheet" of locations mentioned in the data set. These adjudicated annotations were used as a gold standard against which to measure the accuracy of individual people annotating the same data, and against which to measure of accuracy of a Named Entity Recognition software program doing the same. Our result analysis considers primarily the types of errors made by the software in identifying location so that in ensuing research we can create an algorithm that achieves even higher accuracy.

3.1 Our Definition of Location

Our first task was to define location as found in location words in a 300-tweet sample. Discrepancies among locations supplied by three annotators who geo-tagged tweets for this pilot study were discussed, to arrive at parameters of what is and what is not a location. Results of the pilot study appear below. Location words in tweets include:

- Country
  - `Australia's, Kenya`
- State, Region or City
  - `Canturbury [NZ]`
- Abbreviated place
  - `AKL`, `CBD` (the Central Business District), `JPN`
- Neighborhood or district
  - `Takapuwahia, Timaru, the four avenues, Tower Junction`
- Topographical or infrastructure feature
  - `Lyttelton Port`
- Clusters of buildings
  - `Cashel mall`

- Geo-locatable buildings, areas or organizations
  - `Diamond Harbour School, Hutt Library, Pyne Gould Building`
  - `Art Gallery bus stop`
- What + where
  - `BNZ in Riccarton, Shell Station in New Brighton`
- Street address
  - `75 Lyttelton Street`
- Multiple places
  - `welfare centres`
- Generic places (not easily geolocatable)
  - `cordoned off area, room, home, house, city`
  - BUT NOT: `outside, everywhere, on top, in the rubble, across debris`
- Places with hashtags
  - "Japan" in the tweet: `Parents & relatives #Japan ready to go there by all means #perplexity #eqnz`

Excluded from what we considered locations:

- Metonymy (places used to represent an authority)
  - `Christchurch appeal on Givealittle`
- Part of a word without spaces
  - `EarthquakeChCh`
- Twitter-specific references: URLs in a tweet, @location [Auckland in this example is not a location]
  - `@mr_orgue: RT @Auckland_UP: Hospital reported as being evacuated #eqnz`

4 Procedure

4.1 Creation of the Gold Standard

To create a list of objectively correct locations that correspond to those in the data, the two authors of this article and a colleague independently geo-tagged the 1,407 tweets. We produced the gold standard by adjudicating these three manual annotations. Then we use this standard as the yardstick by which to measure how thoroughly individuals found locations in the same data, and by which to measure how thoroughly a Natural Language Processing software could find the same locations.

4.2 Gold Standard Annotations: A Detailed Look at Locations in Crisis Tweets

It has been found that on-the-scene people during a crisis tend to refer to location in tweets (Vieweg et al. 2010). Not surprisingly, some locations they report are local buildings and urban structures (Table 2). People mention a site where there is a problem, or a place where they have seen damage or destruction. Some illustrate their tweets with reference to an article or website, or supply their own photo or video footage. Photo links get quite a lot of attention, if this randomly-selected cathedral photo is any indication (Figure 1), in part because re-tweeted links[5] to the same photo may have different URLs.

Table 1 Locations occurring four or more times within our 1,491-tweet sample. More than one location might occur within a single tweet, but this is relatively infrequent

Locations found in tweet sample	Number of occurrences
Christchurch / ChCh	633
New Zealand / NZ / new zealand	159
CTV building	18
Cathedral	17
Australia	14
Canterbury	12
Auckland	11
Wellington	10
CBD / Christchurch CBD	7
building	7
city	7
airport	6
U.S.	5
Lyttelton	5
hospital	4
Hornby	4

That is, the same person no doubt saw different URLs for this photo and viewed it more than once just as we did, believing it had not been previously seen (Figure 1).

The adjudicated annotations contain 1,207 locations, of which 253 were distinct, and only 15 of these were outside New Zealand. Factors that determine the number of locations found in messages in general include the type of disaster, the phase of the disaster at the time of tweeting, and the duration of the disaster (Vieweg et al. 2010). The most frequently occurring locations among the adjudicated annotations appear in Table 1. This table was created by a simple string matching algorithm. We took more specific places to be distinct from the more general case: for example, we did not count "Hornby mall" as an occurrence of "Hornby."

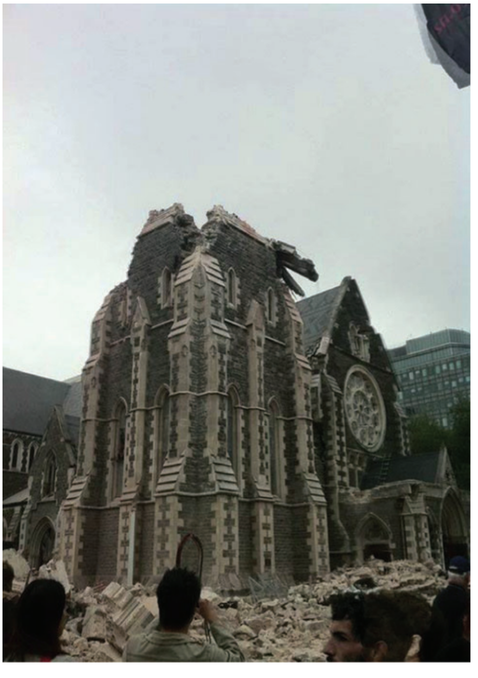
4.3 Manual Annotation

Each individual geo-tagged the tweets based on the definition from the pilot study of what constitutes a location. The gold standard was derived by adjudicating the annotations collected individually from three people. After this, the individual annotations were scored by comparison with the adjudicated annotations.

Individuals sometimes disagreed on what constitutes a location (is "rubble" a location in "sending email from the rubble"?). More often, individuals overlooked locations, perhaps from fatigue with the task, or from staring at an onscreen spreadsheet. Even when local places in the Canterbury, New Zealand area were unfamiliar to us American annotators, they were mostly recognizable as places, given the way the tweet was worded, although occasionally an annotator looked up an unfamiliar name on the Internet to determine whether it was in fact a place.

Table 2 Examples of local buildings and infrastructure in the data set

Locations within Christchurch in the tweet sample	
buildings	roads and bridges
CTV building (Canterbury Television building)	Hereford Street
Pyne Gould building	Hackthorne Road
Central Business District (CBD)	Hagley Park
RSA building	Ferrymead Bridge
Forsyth Barr building	85 Mays Road
Cashel mall	Evans Pass
Cowles Stadium	Colombo Street
Civil Defense Headquarters	4 avenues
Scorpio Books	Tower Junction



17 tweets mention Christchurch Cathedral
1 of these tweets links to a blog
1 of these tweets links to a news article
1 of these tweets recommends a television station's video footage of the cathedral
6 of these tweets include links to the photo shown[6]
1 of these tweets links to a photo that the Tweeter believed (incorrectly) was the cathedral

Photo of Christchurch Cathedral (left)
41,895: the number of times this photo was viewed during the disaster
21 persons commented on this photo

Copyright © Tess Woolcock, with permission, 2011

Figure 1 Statistics on tweets containing the location "cathedral" in our data sample

4.4 Automated Annotation

Two classes of software have been used to identify location in text: geo-referencing software and Named Entity Recognition (NER) software. Geo-referencing software identifies a location and then attaches geographic coordinates that it references in a gazetteer. Regions are given a coordinate pair of latitude and longitude that defines the region's geographic center, or centroid. In a broader task, Named Entity Recognition (NER) software identifies locations as well as other named entities such as persons, organizations or events. No coordinates are attached to locations. The NER learns how to analyze grammar and identify entities in unseen text based on statistical models learned from annotated text. We selected Named Entity Recognition software for tweet parsing because the software does not require a thorough list of places to find locations and so is potentially more generalizable and better-equipped to recognize locations that are local.

Highly regarded Named Entity Recognition (NER) softwares include the Named Entity Tagger of the University of Illinois at Urbana-Champaign, the AlchemyAPI by the Orchestr8 company, and LingPipe from the Alias-i group.[7] Different NER packages rely upon different algorithms to find named entities and classify them into different person/organization/location categories, such that different algorithms will achieve different, if roughly comparable, results parsing the same data set.

The many releases of the Stanford parser since 2002 make it a reliable program. We use the Stanford Named Entity Recognizer version 1.2.1. The Stanford NER uses Conditional Random Fields (CRF) to classify person, location, and organization entities. The CRF approach looks for dependencies among features. In natural language tasks, useful features include neighboring words and word bigrams, prefixes and suffixes, capitalization, membership in domain-specific lexicons, and semantic information from sources such as WordNet (Sutton and McCallum 2006). The parser can be trained to identify other categories, so that we potentially could train it to recognize streets, for example, although in this study we used it as is.

5 Results

5.1 Scoring Results

We compared geo-tags of individual annotators and of the NER software to those of the gold standard. Accuracy of each manual annotator is scored as a percentage. Statistical tests for intra-rater reliability were not used for several reasons. Reliability tests are optimized for continuous data rather than the sort of nominal data we have here (Gwet 2008). The adjudicated annotations were not determined by strict consensus among raters; rather, their separate annotations were used as a guide to create a standard.

Locations assigned by two individual annotators and by the software were scored in different categories depending upon whether a location appeared in the adjudicated annotations (column 2 in Table 3) or not (column 1 in Table 3). The scoring scale is graded so that an exact match with the adjudicated location scored two points, a partial match with truth as one point, and no match as zero points, and for non-locations in the adjudicated scoring (worth less, in our opinion, than the location scoring), an exact match was worth one point and a non-match as worth zero points. The results are presented in Table 3.

Notice in Table 3 that for tweets with no locations, human annotators and software agreed with high accuracy. Identifying a location in a tweet was more challenging for people and also for the software. The single most frequent error of the software not

Table 3 Comparison of geo-parsing accuracy of manual annotators and Named Entity Recognition program

	No location in tweets % accuracy	Location in tweets % accuracy
Annotator A	96.9%	72.8%
Annotator B	99.2%	65.5%
NER software	98.4%	34.4%
NER software (forgiving "chch" error)	not applicable	51.0%

recognizing abbreviations for the disaster city, Christchurch, decreased performance by 15%, so we produced a separate statistic that overlooked this error.

We expected that the Stanford Named Entity Recognizer would find locations that were named entities, but we did not know what proportion of the locations in the tweets would be unnamed entities, or what named entities it would be unable to identify. We found that the Stanford software performed well enough to use as a first pass, but is insufficient to geoparse locations in microtext.

Had we used a standard geo-locator program rather than a Named Entity Recognition program to parse the tweets, our accuracy would have been even lower. This is because most geo-locator programs are supported by gazetteers with city or region as the most specific level of the spatial hierarchy and local places in the category of neighborhood, street or building in a tweet that do not appear in a gazetteer would be missed, whereas some can be identified by NER software.

5.2 Analysis of Human Annotator Mistakes

The time it takes to manually annotate even clipped 140-character tweets makes the annotation task impossible at web scale. Hence, an analysis of human annotator mistakes is irrelevant to the crisis mapping task and is included pro forma to balance our exposition.

People are capable of identifying places correctly in microtext, in accordance with a given place definition. Here, most human errors stemmed from fatigue in the length of the task, or from unfamiliarity in presentation (such as "cbd," an abbreviation in all lower case letters), or unfamiliarity with local Christchurch areas ("4 avenues"). Other errors may have occurred from the eye not absorbing the entire location phrase, for example, "Lyttleton" rather than the full "Lyttleton Port" or "within 5 km of Lyttleton".

5.3 Analysis of Software Location-finding Mistakes

Different Named Entity Recognition software packages employ different classification algorithms, as mentioned above. The Stanford NER software used in this experiment uses Conditional Random Fields. Other NER software using different algorithms will necessarily produce different results.

We need only examine the parameters coded into the Stanford NER software to deduce what locations it would find correctly and what it would miss. Nonetheless, we have experimented by running our data through the software. We found that proper nouns as locations are identified unevenly, and place abbreviations, common nouns as locations and local buildings were not identified at all. For example, none of the following were identified as places by the Stanford NER software: Chriistchurch (note the misspelling), Chch, ChCh, CHCH, CC, CHC, C/church.

From the results, we can induce the types of locations it will find in microtext, as well as the types of locations it will get wrong, whether by omission in type 1 errors or by commission in type 2 errors. These categories of errors are presented in Table 4. Our errors analysis shows the sort of language processing goals an algorithm must address for higher geo-parsing accuracy.

5.4 How to Improve Automatic Place Recognition Results

Our exploration of this data suggests that without customization, the Stanford Named Entity Recognizer will be insufficient to identify many types of locations found in disaster tweets, with types as shown in Table 4. Results should improve if several Named Entity Recognition algorithms were configured to work together. A similar method was used by Shah et al. (2010) in their Named Entity Recognition system that was able to parse Swahili by combining the Conditional Random Field techniques of the Stanford NER and the Learning-Based Java Named Entity Tagger of the University of Illinois at Urbana-Champaign.

Place names can be identified with the help of a spatial lexicon, or gazetteer. GeoNames.org is the most extensive gazetteer available, at over eight million names worldwide.[8] Problems with using a gazetteer are the granularity of locations required and speed of processing required. The number of local places in the data indicates that the gazetteer would need to be enriched substantially with local listings. But if enriched for entries globally, the unwieldy size of the gazetteer file would make processing impossibly slow. If enriched for entries only for the local area of the disaster, the local listings portion might need to be added at the time of the disaster, which would be most inconvenient.

The Stanford NER could be trained to find more types of location data as shown in Table 4, or else one could find another Natural Language Processing package such as OpenCalais that would identify building names, for instance. A promising approach would be to use the data itself to locate and disambiguate place names found in the data. We are conducting further research on this.

5.5 Other Features our Place Recognition Algorithm Requires: Running in Real Time

Our algorithm can recognize new tweets associated with an event automatically in real time when related hashtags are input manually. If the set of hashtags related to an event must be detected automatically, the program would get more thorough results if it could delay up to two days after an event. This is because it has been found that tweet activity flares the day of an event, remains high the day after, and then returns to normal levels (Yom Tov 2011), and new, related hashtags could be introduced at any time during this period.

Table 4 General rules of locations this Stanford Named Entity Recognizer will find correctly in microtext, and the kinds of errors it is likely to make, with examples from actual tweets

Categories of locations NER software got correct (agreed with person) tweet examples	**Categories of type 2 NER error: locations included wrongly tweet examples**	**Categories of type 1 NER error: Locations omitted wrongly tweet examples**
most countries and cities that were capitalized Paris, New Zealand	looked for a second place name consecutive with a first Christchurch. Hang [in there]	cities capitalized but not in standard syntax ST JOHN 1481 [phone number]
standard abbreviations U.S., NZ	symbol confusion MT @	urban areas within a city Latimer Square, Bay of Plenty
after prepositions "in" and "at" and "near" in Sumner, near Oamaru	place name used as adjective Australia USAR teams	local building names Press Building
before "area" Linwood area	transitive verb + "to" Plz translate to Farsi.	common nouns downtown, cathedral
possessives New Zealand's		less common abbreviations Chch [Christchurch] "to" On my way to kaiapoi
		not before "region" Waikato region
		misspelling Chriistchurch
		full phrases including city name Lyttelton Port

6 Toward a Research Agenda using Twitter for Disaster Mapping

The work described above involves only the geo-parsing step in the process of getting tweets onto a map. Here we briefly describe what we have learned about the importance of collecting tweets related to an event, filtering for message accuracy, and prioritizing.

6.1 Collect Relevant Tweets

On-going research should consider how to amass a collection of tweets concerning a particular disaster event. Some Twitter users voluntarily include an index term that allows tweets to be grouped easily with others on the same topic. That indexing information is within tweet text in the form of a crosshatch character and a topic word. Together, the symbol and the topic word make what is called a hashtag. But not every tweet on a topic includes a hashtag. And a single topic may have more than one associated hashtag. For example, both #nzeq and #eqnz represented the 2011 New Zealand earthquake. To collect a thorough sampling of tweets about an event, therefore, requires knowledge of the related hashtags.

6.2 Filter Tweets for Information Accuracy

Achieving reliability in social media remains an unsolved, although studied problem (Mendoza et al. 2010, Castillo et al. 2011). A photo linked to a tweet for this reason has been preferred over tweet text because photo validity is higher, according to J. Stamberger, Associate Director, Disaster Management Initiative, Carnegie Mellon – Silicon Valley Campus. Acting upon faux tweets may hurt the relief effort if actions deflect valuable resources from where they are needed most. We found unreliable tweets in our data set regarding the local hospital being evacuated, as shown in Table 5.

Our ongoing research considers how to spot information that might be unreliable. In the example in Table 5, three of the five faux tweets were actually re-tweets. That is, information reiterated or slightly reconstituted by someone else. We will examine the semantic, syntactic, and network characteristics of faux tweets to see whether we can write an algorithm to predict which tweets might carry mis-information.

Table 5 Example of mis-information and information correction

Mis-information
RT @MareeAnderson: Airport damaged. Hospital being evacuated. Town Centre was packed with people. #eqnz
Omg the tv3 news pictures . . . Chaos. Hospital now evacuated. #eqnz
@mr_orgue: RT @Auckland_UP: Hospital reported as being evacuated #eqnz
RT @BR3NDA: Hospital being evacuated #eqnz
RT @flpatriot: RT @br3nda: Hospital being evacuated #eqnz
Information correction
RT @CEQgovtnz: Christchurch hospital is operational (contrary to some media reports) and one ward has been evacuated. #eqnz #chch http: . . .
RT @martinluff: All hospitals are operating (but phone a medical centre or your GP first unless it is an emergency) #eqnz

6.3 Prioritize Tweets

One of our initial goals was to filter messages by urgency. We learned, however, that responders may have difficulty discerning which of many phone-in victims is in the most danger, so it will be that much more difficult, if not impossible, to discern which messages are most urgent.

If we cannot prioritize tweets concerning the same event according to urgency, can we prioritize them by the time within the tweet? Maps presently list individual messages according to timestamp. Data mining by time mentioned in the tweet text would represent a new avenue of research. The February 23rd and 24th 2011 tweets in Table 6 show that no matter whether people refer to an event in the past, present or future, the tweeted report is often in the present or present perfect tense.[9] The present perfect tense in English may refer to an action that is still going on or has stopped recently. That is, people tend to tweet about events in the near past or near future, in a narrow band around "now". More work should be done here.

7 Our Disaster Map

7.1 How Others Map Social Media: Clusters and Individual Messages

Our New Zealand earthquake crisis had its own set of maps (Figure 2, for instance). The volunteer organization CrisisCommons engineered an earthquake crisis map to which people could contribute.[10] This CrisisCommons map is no longer available over the web, but the site claims to have provided information after the crisis to 100,000 people. The Esri map in Figures 2 and 3 provided by the Canterbury Earthquake Emergency website is a mashup combining data from different sources. It showed locations in categories such as road closures, traffic and bus routes, water tankers, recovery assistance centers and community run hubs.[11] Categories are listed in the leftmost panel, with each category an optional data layer that the user can reveal or hide. Each social media posting appears on the map as an individual icon. Many icons in an area of the map may overlay one another partially. To view individual postings requires clicking the associated icon. The maps in Figures 2 and 3 have icons for the photo sharing site, Flickr. Many Flickr photos come with geographic coordinates associated with the photo by the photo-capturing device. GPS-location precision is unavailable to the vast majority of present tweets, as mentioned above.

7.2 How we Map Social Media: Tweet Clusters and Individual Messages

We geo-locate crisis tweets by recognizing place names in each tweet automatically. In preparation for mapping, geographic coordinates need to be associated with each location. Then we could make a map of tweets associated with Christchurch and its surrounds, as should in our version of a disaster map in Figures 4 and 5. Each cluster radius is governed by the locations mined from the tweets. Our clusters are shown by enlarged labels. Figure 4, for example, shows that many tweets concerned the airport, Riccarton, Upper Riccarton, the hospital and the Central Business District. If a tweet incorporated two locations, we would map it to the more precise of the locations. If a tweet incorporated two locations at the same hierarchical level such as two streets, however, we would map it twice, perhaps with a symbol to signify that the same message appears twice on the map.

Table 6 These samples from 23rd and 24th February show how people represent past, present and future in messages that pertain to location

Past	Present	Future
Another big aftershock reported in Christchurch	hi does anyone at Chch Hospital know if the Chch Dietetics Students are ok? #eqnzContact	Should Chch be rebuilt in Hornby? Very interesting article I read on 6 Feb
#eqnz Air ambos have transported patients to other centres & other South Island hosps are clearing capacity to take casualties if needed	Okay – trying to find bridge access to NB\/Burwood – anyone know which bridge is still open???! #eqnz	#eqnz The Northern Motorway out of Chch appears not to have sustained major damage – but on-ramp at Chaney's Rd closed until further notice
@ NZStuff @robynmalcolm's mother has been pulled out of a collapsed building in the central city. Injured but OK #eqnz	Rescue underway at Forsyth Barr Building – 24 ppl on 17th floor trapped	Those able or willing to offer temporary accommodation on the South Island for #eqnz evacuees pls see our f\u2026 (cont) http:\/\/deck.ly\/~NXw3G
Search and rescue teams from the Bay of Plenty\/Waikato region have been dispatched to Christchurch. #eqnz	\u201c@MaryTV: *URGENT* RT @JoyReidTVNZ Anyone know layout of ctv building? Rescuers urgently want to know #eqnz\u201d yes 027 2290505	Not business as usual in Christchurch. Businesses and schools closed for the next three days. State of emergency in place. #eqnz

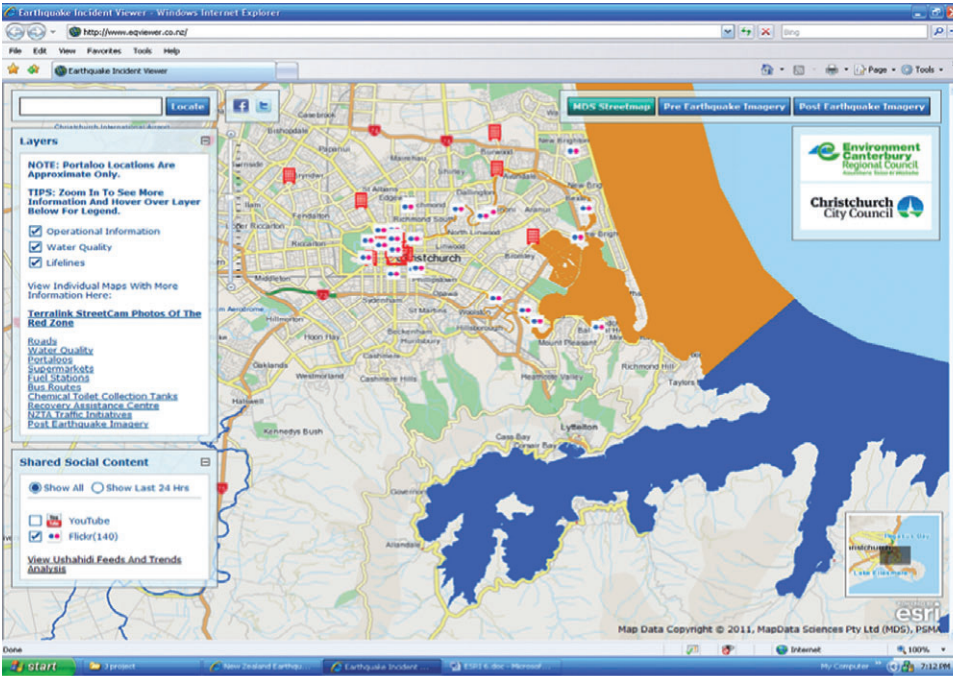


Figure 2 Christchurch, NZ map by Esri showing groups of social media posts (and additional layers that may be brought up), at http://www.eqviewer.co.nz/
Map Data Copyright © 2011, MapData Sciences Pty Ltd (MDS), PSMA

On our map, to read individual tweets within a cluster would require hovering the cursor over that label to bring up a box with the associated tweets (Figure 5). Tweets in our example are ordered in chronological reverse, with the most recent at the top. An even more precise ordering of tweets might result from automatic understanding of verb tenses within tweets as in Table 6, so that tweets describing future events at a particular location would remain on top of the list despite the timestamp order, and tweets describing the past would remain at the bottom. A human moderator of the map could keep the top messages current and remove messages as needed. Alternatively, messages in each geo-cluster might be grouped by topic.

8 Related Work

Geo-parsing entails identifying location words, also known as toponyms. Identifying locations is a sub-problem of identifying all named entities, and so extracting location is often discussed in the context of Named Entity Recognition (NER). NER evaluation is typically in terms of recall and precision. Some systems allow the recall–precision spectrum to be shifted toward one end of the spectrum or the other, since setting one factor high tends to sacrifice the other. Lieberman et al. (2010) offers a survey of recent text geolocation methods.

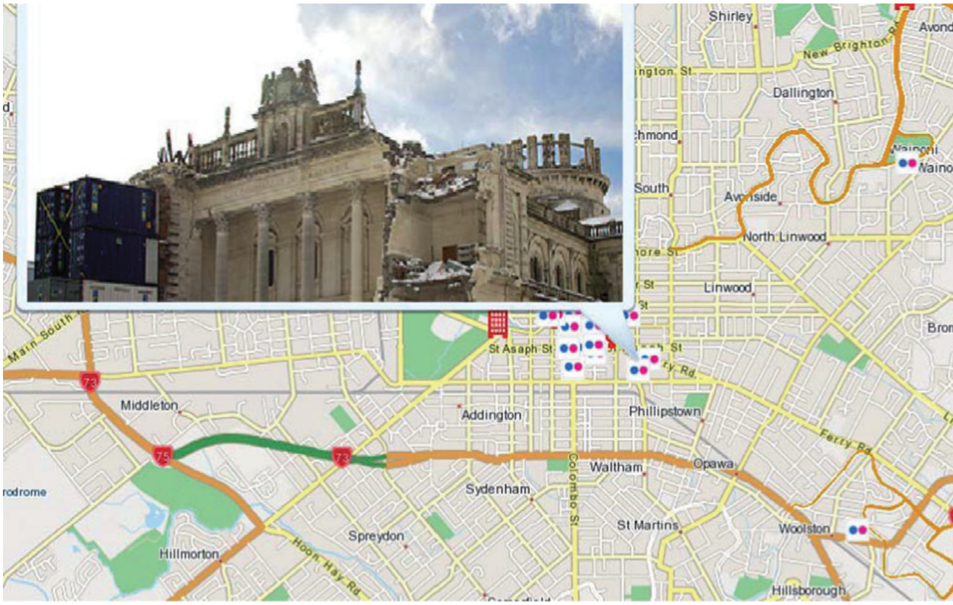


Figure 3 Christchurch, NZ map by Esri showing individual tweets http://www.eqviewer.co.nz/, map data copyright © 2011, MapData Sciences Pty Ltd (MDS), PSMA

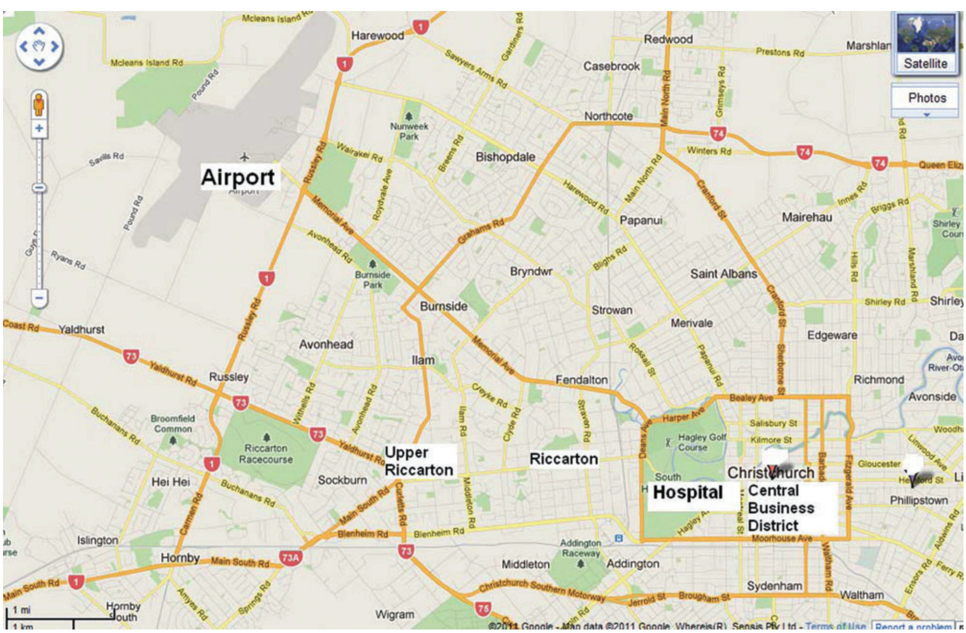


Figure 4 Christchurch, NZ disaster map in our version has enlarged labels for places with numerous tweets, here atop a Google Map base © 2011 Google

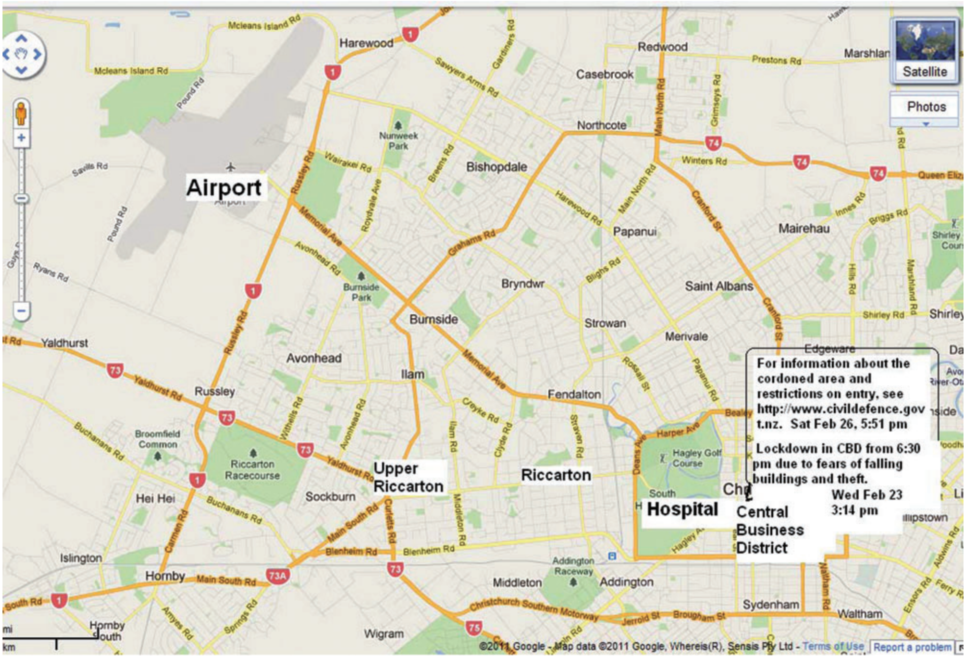


Figure 5 Christchurch, NZ disaster map in our modified version shows individual tweets per location listed in inverse time order © 2011 Google

Locations occurring in tweets resemble locations occurring in other sources, such as web queries (Mandl et al. 2008), with some additional characteristics due to the nature of social media. For example, web queries are not as rich in abbreviations, as searchers know instinctively that these abbreviations will be hard to find. On the other hand, place abbreviations are common in the space-limited Twitter.

Finding the set of tweets relating to an event can be accomplished by keyword or hashtag. Present services include the TagWalk which shows a set of tags related topically (as for example, http://tagwalk.com/tag/nzeq). Their tags drop in relatedness after the first few tags, just as our preliminary algorithm to find related tags has done.

Programs to extract location words from text are available commercially or in open source versions. Web-a-Where from the IBM WebFountain system is an early example (Amitay et al. 2004). MetaCarta is proprietary (MetaCarta 2011); Yahoo! Placemaker and the Edinburgh Geoparser are open to any user.[12] Results surpassing commercial algorithms have been achieved in geo-locator algorithms through the combination of gazetteer match and heuristics (Gelernter and Carley 2011), and with the combination of gazetteer match and machine learning (Freire et al. 2011). Some approaches use context around the extracted location to improve performance (Qin et al. 2010), or local as well as global gazetteers, with the local gazetteer extracted from the domain text itself (Lieberman et al. 2010), or the local gazetteer enriched with volunteered geographic information (Keßler et al. 2009).

Geo-parsing of Twitter presents its own difficulties owing to the nature of microtext. Paradesi (2011) in her TwitterTagger combined NER and gazetteer methods. The system first assigned part-of-speech tags to find proper nouns, and then compared noun phrases per tweet to the U.S. Geological Survey gazetteer to identify locations. The system identified nouns that seem to be places by looking for a spatial indicator, usually a preposition, before a location name. Her research does not consider what sorts of places are found in tweets, however, as we have done for our set of disaster-related tweets, and therefore does not seem to account for abbreviations and common nouns as places that are found frequently.

Abbreviation identification represents a major sub-problem in social media geo-parsing because the abbreviations may be standard or non-standard, upper case or lower case. These steps include identifying the abbreviation. Determining the full word given an abbreviation is a special case of Word Sense Disambiguation (WSD). Researchers to solve this problem have looked for documents that contain the full word as well as the abbreviation, or have used a separate list of abbreviations (Ammar et al. 2011). In lieu of a single document, we can look for full words to disambiguate abbreviations using data indexed with the same hashtag, posted within a given time interval, or passed among social network friends. The alignment of letters between the full word(s) and its abbreviation is one way the abbreviation has been disambiguated (Okizawa et al. 2008).

Auto-recognition errors in geolocation may be higher in false negatives, that is missing location words in a text, than in false positives, that is selecting a word as a location that is not a location actually (Guillén 2008). Other auto-recognition errors have been found higher in false positives than negatives (Gelernter & Carley 2011). Some factors that influence results include effective use of machine learning or heuristics to identify locations, external resources to identify locations, and the nature of the data set.

9 Conclusions

The present standard to map social media messages is to map the location of the user, because the messages themselves rarely have geographic coordinates attached. Here we present research to show that a key step in creating crisis maps fortified by social media is to geo-parse the messages themselves so as to map the events referred to in the message. The contributions of this article include defining the types of locations found in a set of crisis-related messages and categorizing the types of mistakes that an off-the-shelf Named Entity Recognition software makes in identifying those types of locations in microtext to improve future tweet geo-parse algorithms. These results should be verified by testing tweets from a different geographical area. We have presented a map visualization that clusters according to places mentioned in tweets. Other lines of this research would be to analyze temporal characteristics of tweets, which could further enrich information processing in times of crisis.

Acknowledgments

Thanks are due to Dr. Jacob Eisenstein for his insights into bringing an overly ambitious task into focus, and making available the Twitter data sample. We could not have completed this study without the generosity of Brendan O'Connor, who made a subset of his downloaded tweet archive available for our research. We thank also Dr.

Jeannie Stamberger, Associate Director of the Carnegie Mellon Silicon Valley Disaster Management Initiative, and her Carnegie Mellon Silicon Valley colleague Dr. Ian Lane. We are grateful for the continuing support provided by Dr. Jaime Carbonell of the Language Technologies Institute of Carnegie Mellon University. The resulting article has been improved greatly owing to the thoughtful comments of our three reviewers.

Notes

1 Other microblogs besides Twitter, not all with the same space constrictions, are Orkut, Jaiku, Pownce, Yammer, Plurk, Tumblr, and Facebook.
2 The Common Alerting Protocol (CAP) suggests that false alarms should be answered for the general good. OASIS: Common Alerting Protocol v .1.1 http://www.oasis-open.org/committees/download.php/14759/emergency-CAPv1.1.pdf Section 2.3.5 Repudiating a False Alarm.
3 This URL web address is shortened to fit into microtext format, and acts as a redirect to the actual URL.
4 The Google Maps API might be useful for attaching geographic coordinates to place names, for example, although we have not yet experimented with it formally.
5 A re-tweet is a re-send of a tweet sent earlier. A re-tweet does not necessarily repeat the original verbatim. In this case, re-tweets sent by different people may use different URL shorteners. Re-tweets are popular because Twitter is real time, so that anyone who missed the tweet may have another opportunity to see the message.
6 Tweets IDs in our data set that link to this photo 13759, 13760, 15400, 11391, 12529, 12530.
7 As of September 18, 2011, the Illinois Named Entity Tagger is at http://cogcomp.cs.illinois.edu/page/software_view/4. LingPipe is at http://alias-i.com/lingpipe/. The Stanford parser is at http://nlp.stanford.edu/software/lex-parser.shtml; the AlchemyAPI is at http://www.alchemyapi.com/.
8 See http://www.geonames.org/ for additional details.
9 Present perfect = has written; Past perfect = has been written.
10 See http://crisiscommons.org/2011/02/26/project-update-for-the-christchurch-recovery-map/ for additional details.
11 See http://canterburyearthquake.org.nz/earthquake-map-overview/ for additional details.
12 As of July 26, 2011, the Yahoo Placemaker web service application is at http://lerdorf.com/pl.php, and the Edinburgh Geoparser is at http://unlock.edina.ac.uk

References

American Red Cross 2010 Web Users Increasingly Rely on Social Media to Seek Help in a Disaster. WWW document, http://www.redcross.org/portal/site/en/menuitem.94aae335470e233f6cf911df43181aa0/?vgnextoid=6bb5a96d0a94a210VgnVCM10000089f0870aRCRD

Amitay E, Har'El N, Sivan R, and Soffer A 2004 Web-a-Where: Geotagging web content. In *Proceedings of the Twenty-seventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '04)*, Sheffield, United Kingdom: 273–80

Ammar W, Darwish K, El Kahki A, and Hafez K 2011 ICE-TEA: In-Context Expansion and Translation of English Abbreviations. In Gelbukh A (ed) *Computational Linguistics and Intelligent Text Processing*. Berlin, Springer Lecture Notes in Computer Science Vol. 6609: 41–54

Castillo C, Mendoza M, and Poblete B 2011 Information credibility on Twitter. In *Proceedings of the International World Wide Web Conference*, Hyderabad, India: 675–84

Cheng Z, Caverlee J, and Lee K 2010 You are where you tweet: A content-based approach to geo-locating Twitter users. In *Proceedings of the Nineteenth ACM Conference on Information and Knowledge Management (CIKM '10)*, Toronto, Ontario: 759–68

Eisenstein J, O'Connor B, Smith N A, and Xing E P 2010 A latent variable model for geographic lexical variation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, Cambridge, Massachusetts: 1277–87

Freire N, Borbinha J, Calado P, and Martins B 2011 A metadata geoparsing system for place name recognition and resolution in metadata records. In *Proceedings of the Eleventh International ACM/IEEE Joint Conference on Digital Libraries*, Ottawa, Ontario: 339–48

Gao H and Barbier G 2010 Harnessing the crowdsourcing power of social media for disaster relief. *IEEE Intelligent Systems* 26(3): 10–14

Gelernter J and Carley C 2011 Location data mining and time sequencing for social network visualization. *International Journal of Applied Geospatial Research* 2: in press

Guillén R 2008 GeoParsing web queries. In Peters C, Jijkoun V, Mandl T, Müller H, Oard D W, Peñas A, Petras V, and Santos D (eds) *Advances in Multilingual and Multimodal Information Retrieval*. Berlin, Springer Lecture Notes in Computer Science Vol. 5152: 781–85

Gunawan L T, Alers H, Brinkman W-P, and Neerincx M A 2011 Distributed collaborative situation-map making for disaster response. *Interacting with Computers* 23: 308–16

Gwet K L 2008 Intrarater reliability. In D'Agostino R B, Sullivan L, and Massaro J (eds) *Wiley Encyclopedia of Clinical Trials* (4 volume set). New York, John Wiley and Sons (available at http://www.agreestat.com/research_papers/wiley_encyclopedia2008_eoct631.pdf)

Keßler C, Janowicz L, and Bishr M 2009 An agenda for the next generation gazetteer: Geographic information contribution and retrieval. In *Proceedings of the Seventeenth ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS '09)*, Seattle, Washington: 91–100

Lieberman M D, Samet H, and Sankaranarayanan J 2010 Geotagging with local lexicons to build indexes for textually-specified spatial data. In *Proceedings of the Twenty-sixth International IEEE Conference on Data Engineering (ICDE)*, Long Beach, California: 201–12

Liu S B and Palen L 2010 The new cartographers: Crisis map mashups and the emergence of neogeographic practice. *Cartography and Geographic Information Science* 37: 69–90

Mandl T, Gey F, Di Nunzio G, Ferro N, Sanderson M, Santos D, and Womser-Hacker C 2008 An evaluation resource for geographic information retrieval. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC '08)*, Marrakech, Morocco: 6–12

Mendoza M, Poblete B, and Castillo C 2010 Twitter under crisis: Can we trust what we RT? In *Proceedings of the First Workshop on Social Media Analytics (SOMA '10)*, Washington, D.C.

MetaCarta 2011 Geo-Parsing. WWW document, http://www.metacarta.com/products-platform-queryparser.htm

Munro R 2011 Sub-word and spatiotemporal models for identifying actionable information in Haitian Kreyol. In *Proceedings of the Fifteenth Conference on Computational Natural Language Learning, (CoNLL-2011)*. Portland, Oregon

Okizawa N, Ananiandou S, and Tsujii J 2008 A discriminative alignment model for abbreviation recognition. In *Proceedings of the Twenty-second International Conference on Computational Linguistics*, Manchester, United Kingdom: 657–64

Paradesi S 2011 Geotagging tweets using their content. In *Proceedings of the Twenty-fourth International Florida Artificial Intelligence Research Society Conference*, Palm Beach, Florida: 355–56

Qin T, Xiao R, Fang L, Xie X, and Zhang L 2010 An efficient location extraction algorithm by leveraging web contextual information. In *Proceedings of the Eighteenth ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS '10)*, San Jose, California: 53–60

Raj R J R and Sasipraba T 2010 Disaster management system based on GIS web services. In *Proceedings of the International Conference on Recent Advances in Space Technology Services and Climate Change (RSTSCC)*, Chennai, India: 252–61

Shah R, Lin B, Gershman A, and Frederking R 2010 SYNERGY: A named entity recognition system for resource-scarce languages such as Swahili using online machine translation. In *Proceedings of the Second Workshop on African Language Technology (AFLAT 2010)*, Valletta, Malta (available at http://www.cs.cmu.edu/~encore/synergy.pdf)

Shekhar S, Evans M R, Kang J M, and Mohan P 2011 Identifying patterns in spatial information: A survey of methods. *Data Mining and Knowledge Discovery* 1: 193–214

Starbird K and Palen L 2011 Voluntweeters: Self-organizing by digital volunteers in times of crisis. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (CHI '11)*, Vancouver, British Columbia: 1071–80

Starbird K and Stamberger J 2010 Tweak the Tweet: Leveraging microblogging proliferation with a prescriptive syntax to support citizen reporting. In *Proceedings of the Seventh International Conference on Information Systems for Crisis Response and Management* (*ISCRAM 2010)*, Seattle, Washington

Sutton C and McCallum A 2006 An introduction to conditional random fields for relational learning. In Getoor L and Taskar B (eds) *Introduction to Statistical Relational Learning*. Cambridge, MA, MIT Press: 93–128

Vieweg S, Hughes A L, Starbird K, Palen L 2010 Microblogging during two natural hazards events: What Twitter may contribute to situational awareness. In *Proceedings of the 2010 Annual Conference on Human Factors in Computing Systems (CHI 2010)*, Atlanta, Georgia: 1079–88

Yom Tov E 2011 Social and Physical Detachment on Information Need. Unpublished Lecture at Carnegie Mellon University, 5 April, 2011