How do people describe locations during a natural disaster: an analysis of tweets from Hurricane Harvey

Yingjie Hu
GeoAI Lab, Department of Geography, University at Buffalo, USA
yhu42@buffalo.edu

Jimin Wang
GeoAI Lab, Department of Geography, University at Buffalo, USA
jiminwan@buffalo.edu

Abstract

Social media platforms, such as Twitter, have been increasingly used by people during natural disasters to share information and request for help. Hurricane Harvey was a category 4 hurricane that devastated Houston, Texas, USA in August 2017 and caused catastrophic flooding in the Houston metropolitan area. Hurricane Harvey also witnessed the widespread use of social media by the general public in response to this major disaster, and geographic locations are key information pieces described in many of the social media messages. A geoparsing system, or a geoparser, can be utilized to automatically extract and locate the described locations, which can help first responders reach the people in need. While a number of geoparsers have already been developed, it is unclear how effective they are in recognizing and geo-locating the locations described by people during natural disasters. To fill this gap, this work seeks to understand how people describe locations during a natural disaster by analyzing a sample of tweets posted during Hurricane Harvey. We then identify the limitations of existing geoparsers in processing these tweets, and discuss possible approaches to overcoming these limitations.

2012 ACM Subject Classification Information systems → Content analysis and feature selection; Information systems → Retrieval effectiveness

Keywords and phrases Geoparsing, geographic informational retrieval, social media, tweet analysis, disaster response

Digital Object Identifier 10.4230/LIPIcs.GIScience.2020.23

1 Introduction

Hurricane Harvey was a Category 4 tropical storm which started on August 17, 2017 and ended on September 2, 2017 and made a landfall on Texas and Louisiana, USA. It dropped more than 1,300 mm of rain over the Houston metropolitan area and caused catastrophic flooding [44]. During the hurricane and the subsequent flooding, social media platforms, such as Twitter, were used by many residents in the city of Houston and the surrounding areas to share disaster-related information and send help requests.

The use of social media during natural disasters is not new. An early work by Longueville et al. [6] used Twitter to analyze a forest fire in the South of France back in July 2009. In the following years, many studies were conducted based on the social media data collected from disasters to understand the emergency situations on the ground and the reactions of the general public. Examples include the 2010 Pakistan flood [29], the 2011 earthquake on the East Coast of the US [4], Hurricane Sandy in 2012 [27], the 2014 wildfire of California [42], Hurricane Joaquin in 2015 [41], and Hurricane Irma in 2017 [43]. Social media data, such as tweets, provide near real-time information about what is happening in the disaster-affected area, and are suitable for applications in disaster response and situational awareness [25].

11th International Conference on Geographic Information Science (GIScience 2020).
Editors: Piotr Jankowski, Krzysztof Janowicz, and Zbigniew Zwoliński; Article No. 23; pp. 23:1–23:16
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Twitter, in particular, allows researchers to retrieve about 1% of the total number of public tweets for free via its API, and this ability enables various tweet-based disaster studies.

While social media has already been used in disasters and emergency situations, Hurricane Harvey was probably the first major disaster in which the use of social media was comparable or even surpassed the use of some traditional communication methods during a disaster. The National Public Radio (NPR) of the US published an article with the headline "Facebook, Twitter Replace 911 Calls For Stranded In Houston" [35], which described how social media platforms were widely used by Houston residents to request for help when 911 could not be reached. The fact that the storm took out over a dozen emergency call centers and that there were too many 911 calls during and after the hurricane were among the reasons responsible for the failure of the 911 system. Another article published in The Wall Street Journal was titled "Hurricane Harvey Victims Turn to Social Media for Assistance", which described similar stories in which people turned to social media for help after their 911 calls failed [34]. In addition, Hurricane Harvey was called by The Time Magazine as "The U.S.'s First Social Media Storm" [33]. Besides news articles, a survey was conducted by researchers [28] after Hurricane Harvey, which filtered through 2,082 people in Houston and the surrounding communities, and focused on 195 Twitter users. They found that about one-third of their respondents indicated that they used social media to request for help because they were unable to connect to 911.

With the ubiquity of smart mobile devices and the popularity of social media, it seems to be a natural choice for people to turn to Twitter, Facebook, or other social media platforms when their 911 calls fail. People are already familiar with the basic use of these social media platforms (e.g., how to create a post and how to upload a photo), and they can stay connected with their friends and family members online, follow the latest information from public figures (e.g., the Twitter account of the mayor of the affected city), authoritative agencies (e.g., FEMA), and voluntary organizations, and can "@" related people and organizations to send targeted messages. Indeed, a survey by Pourebrahima et al. [30] based on Hurricane Sandy in 2012 revealed that Twitter users received emergency information faster and from more sources than non-Twitter users. The survey by Mihunov et al. [28] found that about 76% of their respondents considered Twitter as "very useful" or "extremely useful" for seeking help during Hurricane Harvey, and roughly three quarters of their respondents indicated that Twitter and other social media were easy to use. Their survey also revealed some challenges in the use of Twitter during a natural disaster, such as not knowing whether volunteers received their requests or when they would send help. However, these situations could change in future disasters, as volunteers and relief organizations learn to collect the requests from social media. In addition to Twitter, other social media platforms were also used by people to seek help [22]. For example, an online group named "Hurricane Harvey 2017 - Together We Will Make It" was created on Facebook to enable victims to post messages about their situations during the flooding [35].

One major challenge in handling the help requests that people sent on social media platforms is to efficiently process the huge number of posts. As described by a disaster responding consultant during Hurricane Harvey [35], "It is literally trying to drink from a firehose". Disaster responders simply do not have the bandwidth and time to manually monitor the huge number of posts on social media and identify actionable information. In fact, there exist multiple challenges in effectively using the information from social media platforms, including verifying the veracity of the posted information, understanding the purpose of the posts (e.g., whether a post is about requesting rescue, reporting disaster situation, calling for donation, or praying for the affected people), and extracting critical information pieces (e.g., the locations of the people who need help). Much research has already been devoted to identifying true information from false information [13, 38], classifying the purposes of social media posts [15, 3], and extracting information from tweets [16, 32].

This paper focuses on the specific challenge of extracting locations from the tweets posted during a natural disaster. As a first step, we focus on understanding how people describe locations during a disaster by analyzing a sample of tweets randomly selected from over 7 million tweets posted during Hurricane Harvey. The contribution of this paper is twofold:

- We conduct an analysis on a sample of 1,000 randomly selected tweets to understand and categorize the ways people describe locations during a natural disaster.
- We identify the limitations of existing tools in extracting locations from these tweets and discuss possible approaches to overcoming these limitations.

The remainder of this paper is organized as follows. Section 2 reviews related work in geoparsing and tweet analysis in the context of disasters. Section 3 describes the dataset from Hurricane Harvey. In Section 4, we analyze and classify location descriptions in the selected tweets. Section 5 reports the experiment results of using existing tools for processing the tweets. Finally, Section 6 summarizes this work and discusses future directions.

2 Related work

Locations in tweets can be extracted through *geoparsing*, a process of recognizing and geo-locating place names (or toponyms) from texts [8, 12, 40]. Geoparsing is often studied within the topic of geographic information retrieval (GIR) [17, 31]. A software tool developed for geoparsing is called a *geoparser*, which typically functions in two consecutive steps: *toponym recognition* and *toponym resolution*. The first step recognizes toponyms from texts, and the second step resolves any place name ambiguity and assigns suitable geographic coordinates. Figure 1 illustrates these two steps. It is worth noting that geoparsing can be applied to other types of texts in addition to social media messages, such as Web pages, news articles, organization documents, and others.

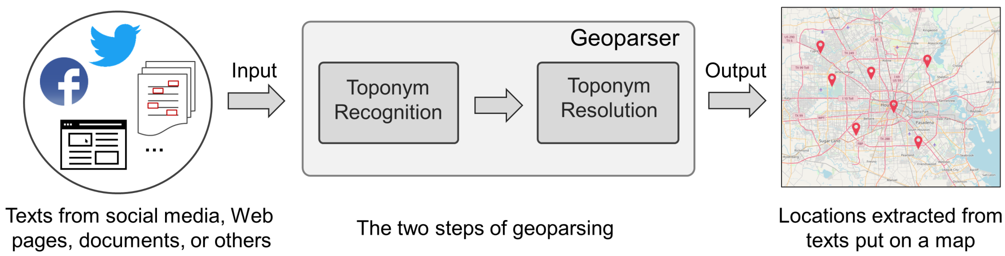


Figure 1 The typical process of geoparsing text to extract locations.

A number of geoparsers have already been developed by researchers. GeoTxt is an online geoparser developed by Karimzadeh et al. [19, 20], which uses the Stanford Named Entity Recognition (NER) tool and several other NER tools for toponym recognition and employs the GeoNames gazetteer[1] for toponym resolution. TopoCluster, developed by Delozier et al. [7], is a geoparser that uses the Stanford NER for toponym recognition and leverages a technique based on the geographic word profiles for toponym resolution. The Edinburgh Geoparser, developed by the Language Technology Group at Edinburgh

[1] https://www.geonames.org/

University [1], uses their own natural language processing (NLP) tool, called LT-TTT2, for toponym recognition, and a gazetteer (e.g., GeoNames) and pre-defined heuristics for toponym resolution. Cartographic Location And Vicinity INdexer (CLAVIN)[2] is a geoparser developed by Berico Technologies that employs the NER tool from the Apache OpenNLP library or the Stanford NER for toponym recognition, and utilizes a gazetteer and heuristics for toponym resolution. CamCoder is a toponym resolution model developed by Gritta et al. [11], which integrates a convolutional neural network and geographic vector representations. Gritta et al. further converted CamCoder into a geoparser by employing the spaCy NER tool for toponym recognition.

Twitter data were used in many previous studies on situational awareness and disaster response. Imran et al. [15] and Yu et al. [43] developed machine learning and text mining systems for automatically classifying tweets into topics, e.g., *caution and advice* and *casualty and damage*. Huang and Xiao [14] classified tweets into different disaster phases, such as preparedness, response, impact and recovery. Kryvasheyeu et al. [21] and Li et al. [23] used tweets for assessing the damages of disasters. Existing studies, however, often used only the geotagged locations of tweets [5, 42] or the locations in the profiles of Twitter users [45, 46], rather than the locations described in tweet content. Many geotagged locations were collected by the GPS receivers in smart mobile devices, and therefore are generally more accurate than the locations geoparsed from the content of tweets. This can be a reason that motivated researchers to use the geotagged locations of tweets. Meanwhile, geotagged locations reflect only the current locations of Twitter users, which may not be the same as the locations described in the content of tweets. In addition, only about 1% tweets were geotagged [36], and the number of geotagged tweets further decreased with Twitter's removal of precise geotagging in June 2019. By contrast, researchers found that over 10% tweets contain some location references in their content [25]. For the locations in the profiles of Twitter users, they may reflect neither the current locations of the users nor the locations described by the users, since the profile locations can be their birthplaces, work places, marriage places, or even imaginary places, and are not always updated.

Some research examined location extraction from the content of tweets. GeoTxt is a geoparser originally developed for processing tweets [20]; however, their testing experiments were based on a tweet corpus, GeoCopora [39], whose toponyms are mostly country names and major city names, rather than fine-grained place names in a disaster affected area (although GeoCopora does contain some fine-grained locations, such as school names). Gelernter and Balaji [9] geoparsed locations in the tweets from the 2011 earthquake in Christchurch, New Zealand, and Wang et al. [41] extracted locations from tweets for monitoring the flood during Hurricane Joaquin in 2015. However, both work focused on using a mixture of NLP techniques and packages (e.g., abbreviation expansion, spell correction, and NER tools) for location extraction, rather than a more detailed analysis on the characteristics of the location descriptions. This paper aims to fill such a gap by examining how people describe locations in tweets during a natural disaster, with the ultimate goal of helping design more effective geoparsers for assisting disaster response.

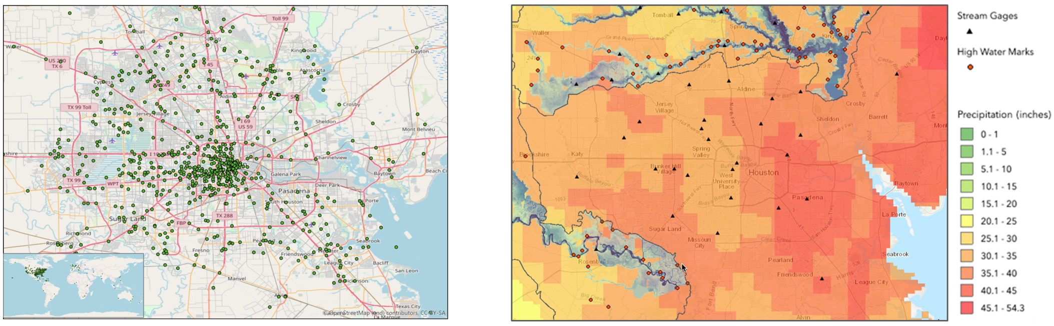
3 Dataset

The dataset used in this work is a set of 7,041,866 tweets collected during Hurricane Harvey and the subsequent flooding from August 18, 2017 to September 22, 2017. This dataset was

[2] https://clavin.bericotechnologies.com/Berico_CLAVIN.pdf

prepared by the University of North Texas Libraries, and the tweets were retrieved based on a set of hashtags and keywords, such as “#HurricaneHarvey”, “#HoustonFlood”, and “Hurricane Harvey”. The entire dataset is available from the library repository of North Texas University (NTU)[3], and it is in the public domain.

Among the over seven million tweets in the entire dataset, only 7,540 are geotagged with longitude and latitude coordinates. These geotagged tweets are distributed not only within the Houston area but also throughout the world, with most of the tweets located inside the United States. Figure 2(a) shows the locations of the geotagged tweets in the Houston area, and the locations of all the geotagged tweets are visualized in the overview map in the lower-left corner. The low percentage of geotagged tweets (about 0.1%) in this



(a) (b)

Figure 2 A comparison of the locations of geotagged tweets and the precipitation during Hurricane Harvey: (a) locations of the geotagged tweets; (b) precipitation in the Houston area from the USGS.

dataset and the fact that the geotagged tweets are distributed throughout the world can be attributed to the data collection process: the data were collected using a list of keywords and hashtags rather than focusing on a particular geographic area. We compare the locations of the geotagged tweets with the precipitation map[4] from the US Geological Survey (USGS) (Figure 2(b)). No clear relationship can be visually identified between the locations of the geotagged tweets and the severity of the precipitation in different areas. For example, the northwestern region received relatively less precipitation than the southeastern region, but there were more geotagged tweets in the former region.

In this work, we are particularly interested in the locations described in the content of tweets. While both the news and literature told us that people used Twitter and other social media platforms to request for help and share information, we still do not know how specifically people describe locations in social media messages during this natural disaster. Manually analyzing the 7,041,866 tweets is practically impossible. Thus, we use a simple regular expression to narrow down the target tweets to be analyzed. The regular expression contains about 70 location-related terms that are frequently observed in place names and location descriptions, such as “street”, “avenue”, “park”, “square”, “bridge”, “rd”, and “ave”. A full list of these terms and the constructed regular expression can be accessed at: `https://github.com/geoai-lab/HowDoPeopleDescribeLocations`. Running this regular expression against the 7 million tweets returns 15,834 tweets. A quick examination of these 15,834 tweets shows that many of them contain detailed location descriptions, such as house number addresses or school names. For curiosity, we also run the same regular expression

[3] https://digital.library.unt.edu/ark:/67531/metadc993940/
[4] https://webapps.usgs.gov/harvey/

against the 7,540 geotagged tweets. Only 203 tweets are returned. This result suggests that there are many tweets that contain location descriptions but are not geotagged. Thus, we will miss important information if we focus on geotagged locations only.

We randomly select 1,000 tweets from the 15,834 records returned by the regular expression. This selection is performed as follows: we first remove retweets to avoid duplication; we then index the remaining tweets, and generate 1,000 non-repeating random integers that are used as the indexes to retrieve the corresponding tweets. As we read through some of these tweets, we see vivid images of people seeking help and sharing information during Hurricane Harvey. Some examples are provided as below:

- *"12 Y/O BOY NEEDs RESCUED! 8100 Cypresswood Dr Spring TX 77379 They are trapped on second story! #houstonflood"*
- *"80 people stranded in a church!! 5547 Cavalcade St, Houston, TX 77026 #harveyrescue #hurricaneharvey"*
- *"Rescue needed: 2907 Trinity Drive, Pearland, Tx. Need boat rescue 3 people, 2 elderly one is 90 not steady in her feet & cant swim. #Harvey"*
- *"Community is responding at shelters in College Park High School and Magnolia High School #TheWoodlands #Harvey..."*
- *"#Houston #HoustonFlood the intersection of I-45 & N. Main Street"*

While the above tweets certainly do not represent all of those posted during Hurricane Harvey, they demonstrate the urgency of some requests. Effectively and efficiently extracting locations from these tweets can help responders and volunteers to reach the people at risk more quickly and can even save lives. In addition, these examples also show that some people were requesting help for others. Thus, even if their tweets were geotagged, it is necessary to focus on the locations described in the content rather than the geotagged locations.

4 Understanding the locations described in Harvey tweets

In this section, we examine and understand the ways people describe locations based on the 1,000 tweets. To do so, we carefully read through each of the tweets, identify and annotate the locations described in their content, and classify the location descriptions. It is worth noting that we focus on the descriptions that refer to specific geographic locations rather than general *locative expressions* [24], such as "this corner" or "that building". The data annotation is done in the following steps. First, the second author reads each tweet and annotates the location descriptions identified; second, the first author goes through the entire dataset, checking each location annotation and discussing with the second author to resolve any annotation difference; a preliminary list of location categories is also identified in this step; third, the first author goes through the entire dataset again, refines the list of categories, and classifies the location descriptions; fourth, the second author performs another round of checking to examine the classified location descriptions. The locations are annotated using the IOB model widely adopted in the CoNLL shared tasks [37]. In the process of annotating the data, we also find that some of the initial 1,000 tweets do not contain specific locations (e.g., a tweet may say: "My side street is now a rushing tributary"). We replace those tweets with others randomly selected from the rest of the data, so that each of the 1,000 tweets contains at least one specific location description. The annotated dataset is available at: `https://github.com/geoai-lab/HowDoPeopleDescribeLocations`.

Ten categories of location descriptions are identified based on the 1,000 Hurricane Harvey tweets (Table 1). The number of tweets in each category is also summarized in Table 1 in the column *Count*. It is worth noting that a tweet may contain more than one type of location descriptions, and therefore can be counted toward more than one category.

Table 1 Ten categories of location descriptions identified from the 1,000 Harvey tweets.

Category	*Examples*	*Count*
C1: House number addresses	- *"Papa stranded in home. Water rising above waist. HELP* ***8111 Woodlyn Rd, 77028*** *#houstonflood"* - *"#HurricaneHarvey family needs rescuing at* ***11800 Grant Rd. Apt. 1009. Cypress, Texas 77429"***	257
C2: Street names	- *"#Harvey LIVE from San Antonio, TX. Fatal car accident at* ***Ingram Rd.****, Strong winds."* - *"****Allen Parkway****,* ***Memorial****,* ***Waugh overpass****, Spotts park and Buffalo Bayou park completely under water"*	571
C3: Highways	- *"9:00AM update video from Hogan St over White Oak Bayou,* ***I-10****,* ***I-45****: water down about 4' since last night..."* - *"Left Corpus bout to be in San Angelo #HurricaneHarvey Y'all be safe Avoided* ***highway 37*** *Took the back road"*	68
C4: Exits of highways	- *"Need trailers/trucks to move dogs from Park Location: Whites Park Pavillion off* ***I-10 exit 61*** *Anahuac TX"* - *"****TX 249 Northbound*** *at Chasewood Dr.* ***Louetta Rd. Exit****. #houstonflood"*	8
C5: Intersections of roads (rivers)	- *"Guys, this is* ***I-45*** *at* ***Main Street*** *in Houston. Crazy. #hurricane #harvey..."* - *"Major flooding at* ***Clay Rd*** *&* ***Queenston*** *in west Houston. Lots of rescues going on for ppl trapped..."*	109
C6: Natural features	- *"****Buffalo Bayou*** *holding steady at 10,000 cfs at the gage near Terry Hershey Park"* - *"Frontage Rd at the river #hurricaneHarvey #hurricaneharvey @* ***San Jacinto River****"*	77
C7: Other human-made features	- *"Houston's* ***Buffalo Bayou Park*** *- always among the first to flood. #Harvey"* - *"If you need a place to escape #HurricaneHarvey, The* ***Willie De Leon Civic Center****: 300 E. Main St in Uvalde is open as a shelter"*	219
C8: Local organizations	- *"#Harvey does anyone know about the flooding conditions around* ***Cypress Ridge High School****?! #HurricaneHarvey"* - *"Cleaning supply drive is underway. 9-11 am today at* ***Preston Hollow Presbyterian Church****"*	60
C9: Admin units	- *"#HurricaneHarvey INTENSE eye wall of category 4 Hurricane Harvey from* ***Rockport, TX****"* - *"Pictures of downed trees and damaged apartment building on Airline Road in* ***Corpus Christi****."*	644
C10: Multiple areas	- *"#HurricaneHarvey Anyone doing high water rescues in the* ***Pasadena/Deer Park*** *area? My daughter has been stranded in a parking lot all night"* - *"FYI to any of you in* ***NW Houston/Lakewood Forest****, Projections are showing Cypress Creek overflowing at Grant Rd"*	6

For category *C1*, we are surprised to see many tweets using the very standard U.S. postal office address format, with a house number, street name, city name, state name, and postal code. Those house number addresses, once effectively extracted from the text, can be located via a typical geocoder (although today's geocoders and geoparsers are developed as separate tools). Some addresses only contain a house number and a street name. Those addresses can be located by narrowing down to the area that is affected by the disaster, e.g., Houston or

Texas in the case of Hurricane Harvey.

Categories *C2* and *C3* cover location descriptions about roads and highways. These two categories could be merged into one. We separate them because our experiments later find that existing NER tools have difficulty in recognizing the US highway names, such as *I-45* and *Hwy 90*. Yet, those highway names are common in many geographic areas of the US and in the daily conversations of people. Thus, we believe that this category is worth to be highlighted from the perspective of developing better geoparsers.

Category *C4* covers highway exits. People can use an exit to provide a more precise location related to a highway. They may use the exit number, e.g., *"exit 61"*, or the street name of an exit, e.g., *"Louetta Rd. Exit"*. This may be related to the US culture since road signs on the US highways often provide both the exit numbers and the corresponding street names. One may also use two exits in one tweet to describe a segment of a highway, such as in *"My uncle is stuck in his truck on I-45 between Cypress Hill & Huffmeister exits"*.

Category *C5* covers location descriptions related to road (or river) intersections. We identify five ways used by people in tweets to describe road intersections: (1) Road A and Road B, (2) Road A & Road B, (3) Road A at Road B, (4) Road A @ Road B, (5) Road A / Road B. Besides, people often use abbreviations when describing intersections, e.g., they may write *"Mary Bates and Concho St"* instead of *"Mary Bates Blvd and Concho St"*. The intersections of two rivers, or a road and a river, are described in similar ways, such as in *"White Oak Bayou at Houston Avenue 1:00 pm Saturday #Houston"*. A tweet may contain more than one intersection, such as in *"Streets Flooded: Almeda Genoa Rd. from Windmill Lakes Blvd. to Rowlett Rd."* which uses two intersections to describe a road segment.

Categories *C6*, *C7*, and *C8* cover location descriptions related to natural features, other human-made features (excluding streets and highways), and local organizations. These location descriptions are generally in the form of place names, such as the name of a bayou, a church, a school, or a park. We find that many tweets also provide the exact address in addition to a place name, such as the second example of *C7*.

Category *C9* covers location descriptions related to towns, cities, and states. Examples include *Houston*, *Katy*, *Rockport*, *Corpus Christi*, *Texas*, and *TX*. This type of locations has limited value from a disaster response perspective, due to their coarse geospatial resolutions.

Category *C10* covers locations related to multiple areas. We find that people use this way to describe a geographic region that typically involves two smaller neighborhoods, towns, or cities, such as *"Pasadena"* and *"Deer Park"* in the first example.

In summary, we have identified ten categories of location descriptions based on the 1,000 tweets from Hurricane Harvey. Overall, people seem to describe their locations precisely by providing the exact house number addresses, road intersections, exit numbers of highways, or adding detailed address information to place names. One reason may be that people, when under emergency situations, may choose to describe locations in precise ways in order to be understood by others such as first responders and volunteers. While these categories are identified based on the 1,000 tweets from a particular disaster, they seem to be general and are likely to be used by people in future disasters in the US. Understanding these location descriptions is fundamental for designing effective geoparsers to support disaster response.

5 Extracting locations from Harvey tweets using existing tools

With the 1,000 Harvey tweets annotated, we examine the performance of existing tools on extracting locations from these tweets. While this seems to be a straightforward task, there are limitations in existing geoparsers that prevent their direct application. First, none of the five geoparsers that we discussed previously, namely GeoTxt, TopoCluster, CLAVIN, the Edinburgh Geoparser, and CamCoder, have the capability of geocoding house number addresses which are an important type of location descriptions (the category of *C1*). Second, none of the five geoparsers have the capability of geo-locating local street names and highway names (the categories of *C2* and *C3*) at a sub-city level (largely due to their use of the GeoNames gazetteer which focuses on the names of cities and upper-level administrative units), let alone road intersections and highway exits (the categories of *C4* and *C5*). It is worth noting that these limitations do not suggest that existing geoparsers are not well designed; instead, they suggest that there is a gap between the demand of processing disaster-related tweets focusing on a local area and the expected application of the existing geoparsers for extracting city- and upper-level toponyms throughout the world (the category of *C9*). Such an application fits well with one of the important objectives of GIR research, namely to geographically index documents such as Web pages [2]. Although we cannot directly apply existing geoparsers to the Harvey tweets, we can examine their components on toponym recognition and resolution respectively.

5.1 Toponym recognition

Existing geoparsers typically use off-the-shelf NER tools for the step of toponym recognition rather than designing their own models. A rationale of doing so is that toponym recognition, to some extent, can be considered as a subtask of named entity recognition. Indeed, many NER tools can recognize multiple types of entities from text, such as persons, companies, locations, genes, music albums, and others. Thus, one can use an NER tool for toponym recognition by keeping only *locations* in the output, and save the effort of developing a model from scratch. How would the NER tools used in existing geoparsers perform on the Hurricane Harvey tweets? In the following, we conduct experiments to answer this question.

The NER tools to be tested in our experiments are the Stanford NER and the spaCy NER, both of which are used in existing geoparsers. Particularly, the Stanford NER has been used in GeoTxt, TopoCluster, and CLAVIN, and the spaCy NER has been used in CamCoder. The Stanford NER has both a default version, which is sensitive to upper and lower letter cases, and a caseless version. Considering that the content of tweets may not have regular capitalization as in well-formatted text, we test both the default case-sensitive Stanford NER and the caseless version. With the typically used 3-class model, both case-sensitive and caseless Stanford NER have three classes in their output: *Person*, *Organization*, and *Location*. Given the names of the three classes, one might choose to keep *Location* only in the output. However, doing so will miss schools and churches described in the tweets, which are often used as shelters during a disaster, because the Stanford NER considers schools and churches as *Organization*. An alternative choice is to keep both *Location* and *Organization* in the output. However, such a design choice will include false positives. For example, in the sentence "The Red Cross has provided recovery assistance to more than 46,000 households affected by Hurricane Harvey", *"Red Cross"* will be included in the output since it is an *Organization*; this adds a false positive into the toponym recognition result. The spaCy NER has a similar issue, whose output includes multiple classes related to geography. These classes are *Facility* (e.g., buildings, airports, and highways), *Organization* (e.g., companies, agencies, and institutions), *GPE* (Geo-Political Entity; e.g., countries, cities, and states), and *Location* (e.g., non-GPE locations, mountain ranges, and bodies of water). Again, one might choose to keep *Location* only given the names of these classes, and a direct consequence is that the spaCy NER will only recognize natural features, such as rivers and mountains, and will miss all other valid toponyms. On the other hand, keeping all the classes can introduce false positives into the output of the spaCy NER. In this work, we test these different design choices for the Stanford NER and the spaCy NER. Specifically, we examine the performances of the Stanford NER when only *Location* is kept in the output (we call it *"narrow"* version) and when both *Organization* and *Location* are kept (*"broad"* version). For the spaCy NER, we examine its performances when only *Location* is kept (*"narrow"*) and when all location-related entities are kept (*"broad"*). In total, we test six versions of the NER tools: the default Stanford NER (narrow and broad), the caseless Stanford NER (narrow and broad), and the spaCy NER (narrow and broad).

In the first set of experiments, we evaluate the performances of these NER tools on recognizing all locations regardless of their categories from the 1,000 Hurricane Harvey tweets. The metrics used are *precision*, *recall*, and *F-score* (Equations 1-3).

$$Precision = \frac{tp}{tp + fp} \tag{1}$$

$$Recall = \frac{tp}{tp + fn} \tag{2}$$

$$F\text{-}score = 2 \cdot \frac{Precision \times Recall}{Precision + Recall} \tag{3}$$

Precision measures the percentage of correctly recognized locations (true positives or *tp*) among all the locations that are recognized by the model (both *tp* and false positives (*fp*)). *Recall* measures the percentage of correctly recognized locations among all the annotated locations which include *tp* and false negatives (*fn*). *F-score* is the harmonic mean of the precision and the recall. F-score is high only when both precision and recall are fairly high, and is low if either of the two is low.

The performances of the six versions of NER tools are reported in Table 2. Overall, the

Table 2 Performances of the NER tools on the 1,000 Hurricane Harvey tweets.

NER tool	***Precision***	***Recall***	***F-score***
Stanford default (*Narrow*)	**0.829**	0.400	0.540
Stanford default (*Broad*)	0.733	**0.441**	**0.551**
Stanford caseless (*Narrow*)	0.804	0.321	0.458
Stanford caseless (*Broad*)	0.723	0.337	0.460
spaCy NER (*Narrow*)	0.575	0.024	0.046
spaCy NER (*Broad*)	0.463	0.305	0.367

performances of all four versions of the Stanford NER dominate the spaCy NER. This result suggests the effectiveness of this classic and open source NER tool developed by the Stanford Natural Language Processing Group [26]. The default Stanford NER with a *narrow* output (i.e., keeping *Location* only) achieves the highest precision, while the default Stanford NER with a *broad* output (i.e., keeping both *Location* and *Organization*) achieves the highest recall and F-score. This result can be explained by the capability of the *broad* Stanford NER in recognizing schools, churches, and other organizations that are also locations in these Hurricane Harvey tweets. The lower precision of the *broad* Stanford NER compared with the *narrow* Stanford NER is explained by the included false positives of the *broad* version.

Another interesting observation from the result is that the default Stanford NER overall performs better than the caseless Stanford NER. Since tweets are user-generated content that may not follow the regular upper and lower cases, we may be tempted to use the caseless version of the Stanford NER. While there do exist tweets with ill-capitalized words, we find that a large percentage of the analyzed tweets (over 85%) still use correct capitalization. Thus, using a caseless version of the Stanford NER, which completely ignores letter cases in the text, will miss the information contained in the correct capitalization used by many tweets. On the other hand, if one expects that most capitalization in the text is incorrect or the text is not capitalized at all, then the caseless version is likely to be a better choice.

In the second set of experiments, we evaluate the performances of the NER tools on the different categories of location descriptions reported in Table 1. Here, we cannot use the same *Precision*, *Recall*, and *F-score* as the evaluation metrics. This is because these NER tools do not differentiate the ten categories of locations (e.g., the Stanford NER considers all of the entities as *Location* or *Organization*, while the spaCy NER does not differentiate streets, highways, and other human-made features). Thus, we use the metric of *Accuracy* that has been used in previous studies, such as [10, 18, 12, 40]. It is calculated using the equation below:

$$Accuracy_c = \frac{|Recognized \cap Annotated_c|}{|Annotated_c|} \tag{4}$$

where $Accuracy_c$ represents the *Accuracy* of a model on the location category c; *Recognized* represents the set of all locations recognized by the model; and $Annotated_c$ represents the set of annotated locations in the category c.

In addition, an NER tool cannot recognize a location that consists of multiple entities. For example, a house number address like *"5547 Cavalcade St, Houston, TX 77026"* (category *C1*) consists of a door number, a street name, a city name, a state name, and a zip code, which are typically recognized as separate entities by an NER. Similar situation applies to road intersections (category *C5*) and multiple areas (category *C10*). These three categories are thus not included in the experiments. The performances of the NER tools on the other seven location categories are shown in Figure 3.

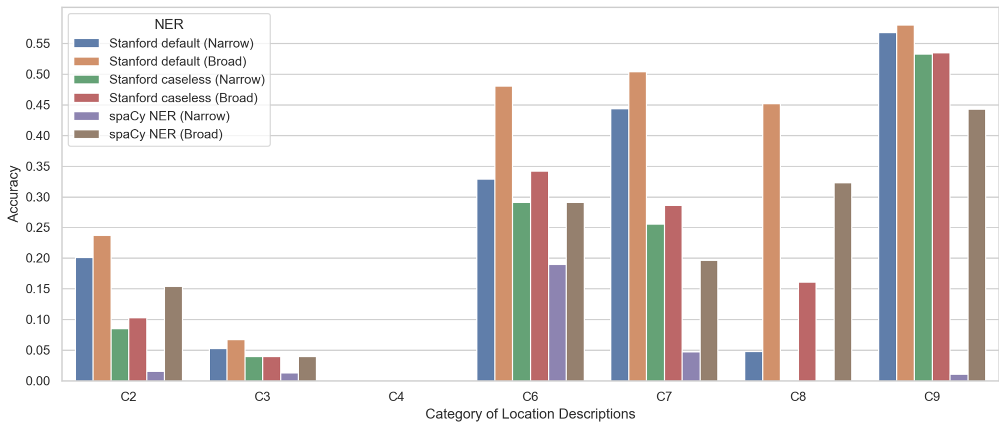


Figure 3 Performances of the NER tools on the different categories of location descriptions.

A number of observations can be obtained from the result. First, all six versions of the NER tools fail on the category *C4: Exits of highways*. This suggests a major limitation of using these off-the-shelf NER tools for toponym recognition: they will miss all the rescue requests whose locations are in the form of highway exits. Second, the broad version of the default Stanford NER has the highest *accuracy* across different categories of location descriptions. However, the broad version likely sacrifices *precision* for *recall* (which cannot be directly measured for each individual category), given its lower overall *precision* compared with the narrow version reported in Table 2. As can be seen in Figure 3, the broad version of the Stanford NER shows a major gain in recognizing organizations (*C8*), since it includes entities in the type of *Organization* in the output. While the broad version also recognizes more locations in other categories, this is often because those locations are considered as *Organization* by the Stanford NER in general. For example, centers, such as *"Walnut Hill Rec Center"* and *"Delco Center"*, in our category *C7* are considered as *Organization* by the Stanford NER. Third, five out of the six NER tools recognize fair percentages of administrative unit names (the category of *C9*), such as *"Houston"* and *"Texas"*. The only exception is the narrow version of the spaCy NER, since it only recognizes the names of natural geographic features. Despite the fair performances of the NER tools, this category of locations has limited value for disaster rescue purposes. Fourth, the performances of the NER tools on street names (*C2*) and highway names (*C2*) are low, but these location description are usually critical for locating the people who need help. A more detailed examination of the result shows that these NER tools often miss the street names that contain numbers, such as *26th St* and *31st Ave*. Similarly, they miss the highway names, such as *I-10* and *Hwy 90*, in which numbers are used even more frequently than in street names. Finally, these NER tools have only low to fair performances on natural features (e.g., rivers and bayous; *C6*) and other human-made features (e.g., parks; *C7*).

In sum, the experiment results suggest that existing NER tools have limited performance in recognizing locations, especially sub-city level locations, from disaster-related tweets. They do not have the capability of recognizing location descriptions that consist of multiple entities, such as house number addresses, road intersections, and multiple areas, and largely fail on highways, highway exits, and the street names that contain numbers. As a result, there is a need for developing more effective toponym recognition models that can recognize these location descriptions from tweets.

5.2 Toponym resolution

The toponym resolution components of existing geoparsers use a variety of strategies to resolve ambiguity and geo-locate place names. These strategies include heuristics based on the population of cities (e.g., a toponym is resolved to the place with the highest population), the co-occurrences of related place names (e.g., the names of higher administrative units), and others [1, 20]. There are also methods that create a grid tessellation covering the surface of the Earth and calculate the probability of a place name to be located in each grid [7, 11]. However, existing toponym resolution components focus more on the task of disambiguating and geo-locating place names at a world scale, such as understanding which *"Washington"* the place name is referring to, given the many places named *"Washington"* in the world.

By contrast, the task of resolving locations described in disaster related tweets has different characteristics. First, these locations are generally at sub-city level, such as roads and house number addresses. Unlike cities, these fine-grained locations are often not associated with populations. This makes it difficult to apply existing toponym resolution heuristics based on population. Second, given these location descriptions are about a disaster-affected local area, the task of toponym disambiguation becomes easier. While there can still be roads having the same name within the same city, the number of places that share the same name decreases largely (e.g., there is no need to disambiguate over 80 different *"Washington"*s when we focus on a local area). Third, point-based location representations typically returned by existing geoparsers become insufficient. We may need lines or polygons, in addition to points, to provide more accurate representation for the described locations.

Given that existing toponym resolution strategies are not applicable to the task of resolving location descriptions in disaster-related tweets, we discuss what are needed if we are going to develop a toponym resolution model for handling this task. First, it is necessary to have a local gazetteer that focuses on the disaster affected area and has detailed geometric representation (i.e., points, lines, and polygons) of the geographic features. Compared with the typically used GeoNames gazetteer, a local gazetteer serves two roles: (1) it reduces place name ambiguity by limiting place names to the disaster-affected area; and (2) it provides detailed spatial footprints for representing fine-grained locations. Such a local gazetteer could be constructed by conflating OpenStreetMap data, the GeoNames data within the local region, and authoritative geospatial data from mapping agencies. Second, we need a geocoder embedded in the toponym resolution model to handle house number addresses. Successfully embedding such a geocoder also requires the local gazetteer to contain house number data along with the roads and streets. Third, additional natural language processing methods are necessary to identify the spatial relations among the multiple locations described in the same tweet. This is especially important for location descriptions in Categories *C4*, *C5*, and *C10* when we need to locate the intersection of two roads (or a road and a river), the exit of a highway, or a combination of two regions. In addition, the NLP methods can help the toponym resolution model determine which geometric representation to use. Consider two possible tweets *"Both Allen Parkway and Memorial Dr are flooded"* and *"Flooding at the intersection of Allen Parkway and Memorial Dr"*. While the same roads are described in these two tweets, the ideal geometric representation for them should be different.

6 Conclusions and future work

Hurricane Harvey is a major natural disaster that devastated the Houston metropolitan area in 2017. Hurricane Harvey also witnessed the wide use of social media, such as Twitter, by the disaster-affected people to seek help and share information. Given the increasing popularity of social media among the general public, they are likely to be used in future disasters. One challenge in using social media messages for supporting disaster response is automatically and accurately extracting locations from these messages. In this work, we examine a sample of tweets sent out during Hurricane Harvey in order to understand how people describe locations in the context of a natural disaster. We identify ten categories of location descriptions, ranging from house number addresses and highway exits to human-made features and multiple regions. We find that under emergency situations people tend to describe their locations precisely by providing exact house numbers or clear road intersection information. We further conduct experiments to measure the performances of existing tools for geoparsing these Harvey tweets. Limitations of these tools are identified, and we discuss possible approaches to developing more effective models. In addition to social media messages, other approaches, such as *what3words* (`what3words.com`), could also be promoted to help people communicate their locations in emergency situations. *What3words* could be especially useful in geographic areas that lack standard addresses; meanwhile, it will also require people to have some familiarity with the system and install the relevant app.

A number of research topics can be pursued in the near future. First, while we have gained some understanding on how people describe locations during a natural disaster, it is limited to English language and within the culture of the United States. People speaking other languages or in other countries and cultures are likely to describe locations in different ways that need further investigation. Second, we can move forward and experiment possible approaches to developing models for recognizing and geo-locating the location descriptions in tweets posted during disasters. Examples include toponym recognition models that can correctly recognize highways and streets whose names contain numbers, and toponym resolution models that can correctly interpret the spatial relations of the multiple locations described in the same tweet. Finally, location extraction is only one part (although an important part) of the whole pipeline for deriving useful information from social media messages. Future research can integrate location extraction with other methods, such as those for verifying information veracity and classifying message purposes, to help disaster responders and volunteer organizations make more effective use of social media and reach the people in need.

References

1. Beatrice Alex, Kate Byrne, Claire Grover, and Richard Tobin. Adapting the Edinburgh geoparser for historical georeferencing. *International Journal of Humanities and Arts Computing*, 9(1):15–35, 2015.
2. Einat Amitay, Nadav Har'El, Ron Sivan, and Aya Soffer. Web-a-where: geotagging web content. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 273–280, New York, NY, USA, 2004. ACM.
3. Zahra Ashktorab, Christopher Brown, Manojit Nandi, and Aron Culotta. Tweedr: Mining twitter to inform disaster response. In *ISCRAM*, 2014.
4. Andrew Crooks, Arie Croitoru, Anthony Stefanidis, and Jacek Radzikowski. # earthquake: Twitter as a distributed sensor system. *Transactions in GIS*, 17(1):124–147, 2013.
5. Joao Porto De Albuquerque, Benjamin Herfort, Alexander Brenning, and Alexander Zipf. A geographic approach for combining social media and authoritative data towards identifying useful information for disaster management. *International Journal of Geographical Information Science*, 29(4):667–689, 2015.
6. Bertrand De Longueville, Robin S Smith, and Gianluca Luraschi. "OMG, from here, I can see the flames!" a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *Proceedings of the 2009 International Workshop on Location Based Social Networks*, pages 73–80, 2009.
7. Grant DeLozier, Jason Baldridge, and Loretta London. Gazetteer-independent toponym resolution using geographic word profiles. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 2382–2388, Palo Alto, CA, USA, 2015. AAAI Press.
8. Nuno Freire, José Borbinha, Pável Calado, and Bruno Martins. A metadata geoparsing system for place name recognition and resolution in metadata records. In *Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries*, pages 339–348, New York, NY, USA, 2011. ACM.
9. Judith Gelernter and Shilpa Balaji. An algorithm for local geoparsing of microtext. *GeoInformatica*, 17(4):635–667, 2013.
10. Judith Gelernter and Nikolai Mushegian. Geo-parsing messages from microtext. *Transactions in GIS*, 15(6):753–773, 2011.
11. Milan Gritta, Mohammad Taher Pilehvar, and Nigel Collier. Which melbourne? augmenting geocoding with maps. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, volume 1, pages 1285–1296, 2018.
12. Milan Gritta, Mohammad Taher Pilehvar, Nut Limsopatham, and Nigel Collier. What's missing in geographical parsing? *Language Resources and Evaluation*, 52(2):603–623, 2018.
13. Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22nd international conference on World Wide Web*, pages 729–736, 2013.

14 Qunying Huang and Yu Xiao. Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS International Journal of Geo-Information*, 4(3):1549–1568, 2015.

15 Muhammad Imran, Carlos Castillo, Ji Lucas, Patrick Meier, and Sarah Vieweg. AIDR: Artificial intelligence for disaster response. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 159–162, 2014.

16 Muhammad Imran, Shady Elbassuoni, Carlos Castillo, Fernando Diaz, and Patrick Meier. Extracting information nuggets from disaster-related messages in social media. In *ISCRAM*, 2013.

17 Christopher B. Jones and Ross S. Purves. Geographical information retrieval. *International Journal of Geographical Information Science*, 22(3):219–228, 2008.

18 Morteza Karimzadeh. Performance evaluation measures for toponym resolution. In *Proceedings of the 10th Workshop on Geographic Information Retrieval*, page 8, New York, NY, USA, 2016. ACM.

19 Morteza Karimzadeh, Wenyi Huang, Siddhartha Banerjee, Jan Oliver Wallgrün, Frank Hardisty, Scott Pezanowski, Prasenjit Mitra, and Alan M MacEachren. Geotxt: a web api to leverage place references in text. In *Proceedings of the 7th workshop on geographic information retrieval*, pages 72–73, New York, NY, USA, 2013. ACM.

20 Morteza Karimzadeh, Scott Pezanowski, Alan M MacEachren, and Jan O Wallgrün. Geotxt: A scalable geoparsing system for unstructured text geolocation. *Transactions in GIS*, 23(1):118–136, 2019.

21 Yury Kryvasheyeu, Haohui Chen, Nick Obradovich, Esteban Moro, Pascal Van Hentenryck, James Fowler, and Manuel Cebrian. Rapid assessment of disaster damage using social media activity. *Science advances*, 2(3):e1500779, 2016.

22 Jing Li, Keri K Stephens, Yaguang Zhu, and Dhiraj Murthy. Using social media to call for help in Hurricane Harvey: Bonding emotion, culture, and community relationships. *International Journal of Disaster Risk Reduction*, 38:101212, 2019.

23 Zhenlong Li, Cuizhen Wang, Christopher T Emrich, and Diansheng Guo. A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 south carolina floods. *Cartography and Geographic Information Science*, 45(2):97–110, 2018.

24 Fei Liu, Maria Vasardani, and Timothy Baldwin. Automatic identification of locative expressions from social media text: A comparative analysis. In *Proceedings of the 4th International Workshop on Location and the Web*, pages 9–16, 2014.

25 Alan M MacEachren, Anuj Jaiswal, Anthony C Robinson, Scott Pezanowski, Alexander Savelyev, Prasenjit Mitra, Xiao Zhang, and Justine Blanford. Senseplace2: Geotwitter analytics support for situational awareness. In *Visual analytics science and technology (VAST), 2011 IEEE conference on*, pages 181–190. IEEE, 2011.

26 Christopher Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven Bethard, and David McClosky. The stanford corenlp natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*, pages 55–60, 2014.

27 Stuart E Middleton, Lee Middleton, and Stefano Modafferi. Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, 29(2):9–17, 2013.

28 Volodymyr V Mihunov, Nina SN Lam, Lei Zou, Zheye Wang, and Kejin Wang. Use of twitter in disaster rescue: lessons learned from Hurricane Harvey. *International Journal of Digital Earth*, pages 1–13, 2020.

29 Dhiraj Murthy and Scott A Longwell. Twitter and disasters: The uses of twitter during the 2010 pakistan floods. *Information, Communication & Society*, 16(6):837–855, 2013.

30 Nastaran Pourebrahim, Selima Sultana, John Edwards, Amanda Gochanour, and Somya Mohanty. Understanding communication dynamics on twitter during natural disasters: A case study of hurricane sandy. *International journal of disaster risk reduction*, 37:101176, 2019.

31 Ross S Purves, Paul Clough, Christopher B Jones, Mark H Hall, Vanessa Murdock, et al. Geographic information retrieval: Progress and challenges in spatial search of text. *Foundations and Trends® in Information Retrieval*, 12(2-3):164–318, 2018.

32 J Rexiline Ragini, PM Rubesh Anand, and Vidhyacharan Bhaskar. Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management*, 42:13–24, 2018.

33 Maya Rhodan. Hurricane Harvey: The U.S.'s first social media storm. *Time Magazine*, 2017. URL: https://time.com/4921961/hurricane-harvey-twitter-facebook-social-/.

34 Deepa Seetharaman and Georgia Wells. Hurricane Harvey victims turn to social media for assistance. *The Wall Street Journal*, 2017. URL: https://www.wsj.com/articles/hurricane-harvey-victims-turn-to-social-media-for-assistance-1503999001.

35 Lauren Silverman. Facebook, twitter replace 911 calls for stranded in houston. *National Public Radio*, 2017. URL: https://www.npr.org/sections/alltechconsidered/2017/08/28/546831780/texas-police-and-residents-turn-to-social-media-to-communicate-amid-harvey.

36 Luke Sloan, Jeffrey Morgan, William Housley, Matthew Williams, Adam Edwards, Pete Burnap, and Omer Rana. Knowing the tweeters: Deriving sociologically relevant demographics from twitter. *Sociological research online*, 18(3):1–11, 2013.

37 Erik F Tjong Kim Sang and Fien De Meulder. Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4*, pages 142–147. Association for Computational Linguistics, 2003.

38 Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

39 Jan Oliver Wallgrün, Morteza Karimzadeh, Alan M. MacEachren, and Scott Pezanowski. Geocorpora: building a corpus to test and train microblog geoparsers. *International Journal of Geographical Information Science*, 32(1):1–29, 2018.

40 Jimin Wang and Yingjie Hu. Enhancing spatial and textual analysis with eupeg: An extensible and unified platform for evaluating geoparsers. *Transactions in GIS*, 23(6):1393–1419, 2019.

41 Ruo-Qian Wang, Huina Mao, Yuan Wang, Chris Rae, and Wesley Shaw. Hyper-resolution monitoring of urban flooding with social media and crowdsourcing data. *Computers & Geosciences*, 111:139–147, 2018.

42 Zheye Wang, Xinyue Ye, and Ming-Hsiang Tsou. Spatial, temporal, and content analysis of twitter for wildfire hazards. *Natural Hazards*, 83(1):523–540, 2016.

43 Manzhu Yu, Qunying Huang, Han Qin, Chris Scheele, and Chaowei Yang. Deep learning for real-time social media text classification for situation awareness–using Hurricanes Sandy, Harvey, and Irma as case studies. *International Journal of Digital Earth*, pages 1–18, 2019.

44 Wei Zhang, Gabriele Villarini, Gabriel A Vecchi, and James A Smith. Urbanization exacerbated the rainfall and flooding caused by Hurricane Harvey in Houston. *Nature*, 563(7731):384–388, 2018.

45 Lei Zou, Nina SN Lam, Heng Cai, and Yi Qiang. Mining twitter data for improved understanding of disaster resilience. *Annals of the American Association of Geographers*, 108(5):1422–1441, 2018.

46 Lei Zou, Nina SN Lam, Shayan Shams, Heng Cai, Michelle A Meyer, Seungwon Yang, Kisung Lee, Seung-Jong Park, and Margaret A Reams. Social and geographical disparities in twitter use during Hurricane Harvey. *International Journal of Digital Earth*, 12(11):1300–1318, 2019.