

LORE: a model for the detection of fine-grained locative references in tweets

Nicolás José Fernández-Martínez¹

Universidad Católica San Antonio de Murcia
Spain

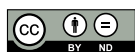
Carlos Perriñán-Pascual

Universitat Politècnica de València
Spain

ONOMÁZEIN 52 (June 2021): 195-225

DOI: 10.7764/onomazein.52.11

ISSN: 0718-5758



Nicolás José Fernández-Martínez: Department of Languages, Universidad Católica San Antonio de Murcia, Spain. | E-mail: njfernandezmartinez@gmail.com

Carlos Perriñán-Pascual: Applied Linguistics Department, Universitat Politècnica de València, Spain.

| E-mail: joepas3@upv.es

Received: November 2019

Accepted: April 2019

Abstract

Extracting geospatially rich knowledge from tweets is of utmost importance for location-based systems in emergency services to raise situational awareness about a given crisis-related incident, such as earthquakes, floods, car accidents, terrorist attacks, shooting attacks, etc. The problem is that the majority of tweets are not geotagged, so we need to resort to the messages in the search of geospatial evidence. In this context, we present LORE, a location-detection system for tweets that leverages the geographic database GeoNames together with linguistic knowledge through NLP techniques. One of the main contributions of this model is to capture fine-grained complex locative references, ranging from geopolitical entities and natural geographic references to points of interest and traffic ways. LORE outperforms state-of-the-art open-source location-extraction systems (i.e. Stanford NER, spaCy, NLTK and OpenNLP), achieving an unprecedented trade-off between precision and recall. Therefore, our model provides not only a quantitative advantage over other well-known systems in terms of performance but also a qualitative advantage in terms of the diversity and semantic granularity of the locative references extracted from the tweets.

Keywords: location detection; location extraction; geolocation; tweet; named entity recognition.

1 Corresponding author.

1. Introduction

Twitter is one of the most widely used and popular microblogging and social media sites, as well as one of the most investigated in event detection, sentiment analysis and geolocation, among many other research areas (Murthy, 2018). Applications of Twitter-based geolocation systems range from health surveillance and disease tracking (Dredze et al., 2013) or disaster management and tracking (Vieweg et al., 2010; Crooks et al., 2013; Imran et al., 2014; Jongman et al., 2015; Martínez-Rojas et al., 2018; Zhang et al., 2019) to traffic-incident detection and road traffic control (Gonzalez-Paule et al., 2019). In this regard, Twitter can play the role of a real-time social sensor system that provides multidirectional channels of communication in emergency and crisis events between the affected people and the disaster-management agencies (Martínez-Rojas et al., 2018; Zhang et al., 2019). A rapid understanding of these events can help competent authorities take immediate decisions and actions to allocate human and economic resources effectively. The problem is that geotagged tweets represent around 1% of tweets only (Middleton et al., 2014), which hinders the development of geolocation systems for Twitter. To overcome this problem, it becomes necessary to analyze other geographically rich information, such as location mentions in tweet messages, which are indeed much more frequent than geotagged data (Wallgrün et al., 2018).

In this context, the primary goal of this article is to provide detailed insight into the processing that takes place in LORE (LOCative Reference Extractor), a linguistically aware model that captures any type of locative reference in microtexts, ranging from geopolitical entities (e.g. towns, cities, states or countries), natural landforms (e.g. lakes, rivers, mountains, ridges or beaches) and points of interest or POIs (e.g. schools, churches, malls, museums or police stations) to traffic ways (e.g. streets, avenues, turnpikes, boulevards, highways or roads)². To the best of our knowledge, this is the first model that manages to extract fine-grained locative references in tweets, for which the model leverages GeoNames (Ahlers, 2013) and a gazetteer of location-indicative words extracted from WordNet (Miller, 1995; Fellbaum, 1998) in conjunction with an inventory of lexico-syntactic rules. LORE currently works for English tweets only, with a view to projecting its functionalities into other languages, such as French, Italian and Spanish. The result of this research is aimed at being integrated into CASPER, CAtegory- and Sentiment-based Problem FindER (Perriñán-Pascual & Arcas-Túnez, 2017, 2018, 2019), a multi-domain problem-detection system for tweets. The remainder of this article is organised as follows. Section 2 describes the state of the art in location detection. Section 3 provides an accurate account of our method of location detection, and section 4 evaluates the research. Finally, section 5 presents some conclusions.

2 LORE, which has been developed in C# with ASP.NET 4.6 and MySQL Database, is freely accessible from the FunGramKB website (<http://www.fungramkb.com/nlp.aspx>).

2. Background and related work

2.1. Terminology

Location detection in user-generated text content on social media, e.g. tweets, is a major focus of research in the field of geographic information retrieval (Purves et al., 2018), where areas such as computational linguistics, natural language processing (NLP) and knowledge engineering converge. Henceforth, the term “location detection”, also known as toponym recognition (Middleton et al., 2018) or location extraction (Dutt et al., 2018) in the scientific literature, will be used to refer to the identification and extraction of locative references from unstructured text. However, location detection should not be confused with other terms such as geocoding (Middleton et al., 2018) or geotagging (Gritta et al., 2018), which deal with the assignation of spatial coordinates to locative references after being disambiguated (Gritta et al., 2018). In this context, geoparsing (Leidner & Lieberman, 2011; Liu et al., 2014) usually consists of two phases, location detection and location disambiguation (Gelernter & Balaji, 2013; Purves et al., 2018; Wallgrün et al., 2018).

2.2. Location-detection models

On the one hand, the commonest approach to location detection is Named Entity Recognition (NER), a line of research in NLP that deals with the identification and classification of named entities, not only location names but also person and organization names *inter alia* (Barrière, 2016; Goyal et al., 2018). NER-based approaches applied to microblogging services such as Twitter perform reasonably well when confronting the challenges presented by the noisy nature of tweets (X. Liu et al., 2011; Karimzadeh et al., 2019). However, NER usually experiences performance drops when dealing with the non-standard spelling and typographical characteristics of social-media microtexts. According to Jurafsky & Martin (2020), there are three types of NER-based models:

- (a) feature-based NER, which employs machine learning (ML) algorithms such as Conditional Random Fields (Finkel et al., 2005; Han, Jimeno-Yepes et al., 2014) or Hidden Markov Models (Sarkar, 2015),
- (b) neural NER, which uses deep-learning (DL) techniques such as bidirectional Long Short Term Memory (biLSTM) (Gerguis et al., 2016; Limsopatham & Collier, 2016), usually in combination with Convolutional Neural Networks (CNN) (Dugas & Nichols, 2016; Aguilar et al., 2018), and
- (c) rule-based NER, which is based on hand-crafted lexico-syntactic rules typically using regular expressions (Malmasi & Dras, 2016; Dutt et al., 2018).

In feature-based NER, sentences are tokenized, where each token (or word) is taken as a vector representing a set of linguistic features (Leidner & Lieberman, 2011; Middleton et al., 2018).

Some of these linguistic features are capitalization, location-indicative word or POS tag. Two steps are essential to build any ML model: training and testing. First, we train an ML algorithm with a training corpus, which contains manually tagged data in the form of the different linguistic features. Second, we apply the algorithm to a test corpus to measure the performance of the model; in other words, we evaluate how well the algorithm was trained with the training corpus by making some predictions on the test corpus. For example, the Stanford NER tool uses this kind of model, achieving very high performance in the news genre (Finkel et al., 2005). However, performance considerably degrades with Twitter data, so the ML algorithm is usually retrained with tweets (Lingad et al., 2013; Hoang & Mothe, 2018). In this regard, Ritter et al. (2011) implemented a well-known feature-based Twitter-specific NER tool, which can detect named entities such as locations, person names and organization names.

Neural NER models achieve good performance in many NER tasks (Espinosa et al., 2016). These models rely on neural networks, which consist of an input layer, multiple hidden layers, and an output layer. These layers are nodes that transform real-world data into numerical values and process them to obtain an output that is then learned by the algorithm that performs feature extraction. With regard to Twitter-specific NER tasks, biLSTM and/or CNN have been successfully applied (Dugas & Nichols, 2016; Espinosa et al., 2016; Aguilar et al., 2018).

Both feature-based NER and neural NER are based on probabilistic models, whose performance largely depend on the coverage and quality of the training data (Purves et al., 2018). In contrast, rule-based NER is based on a symbolic model, which makes use of hand-crafted lexico-syntactic rules that help infer location terms. These rules usually take the form of regular expressions that aim to capture linguistic patterns in text strings from the knowledge provided by NLP tasks such as tokenization and POS tagging. For example, the presence of prepositions followed by proper nouns is usually taken as a strong linguistic cue to extract location entities (Hoang & Mothe, 2018). Overall, rule-based NER alone can achieve very high precision but low recall (Jurafsky & Martin, 2020).

On the other hand, another frequently used approach to location detection is Named Entity Matching (NEM) (Leidner & Lieberman, 2011; Middleton et al., 2018). It consists in the use of lists of location names (or gazetteers) retrieved from geographical databases (or geodatabases), such as GeoNames³ (Ahlers, 2013) or OpenStreetMap⁴ (Acheson et al., 2017), to identify locative references in the text (Middleton et al., 2014; Malmasi & Dras, 2016; de Bruijn et al., 2018). For example, GeoNames is one of the most widely used geodatabase, containing over 25 million place names in all countries. Most of the location types stored in this database are geopolitical entities, natural geographic references and a few POIs. Besides place names, it

3 www.geonames.org

4 <https://www.openstreetmap.org>

also contains some geographical features of populated places, such as population size and latitude-longitude coordinates, which are very helpful for location-disambiguation and geovisualization purposes (Purves et al., 2018). However, although including a high number of place names, GeoNames lacks location subtypes such as addresses, roads, buildings, etc. (Ahlers, 2013; Dutt et al., 2018). NEM-based systems for location detection in tweets seem to achieve greater performance than NER-based systems (Middleton et al., 2014). However, NEM presents several drawbacks. First, geodatabases are finite, so they might not capture the full range of existent place names (Purves et al., 2018). Second, these models cannot serve to disambiguate place names from person names, e.g. the city of *Paris* from *Paris Hilton* (Gritta et al., 2020), which results in a case of ambiguity.

Finally, Middleton et al. (2018) suggested that a hybrid approach, based on the combination of NER with NEM, can greatly reduce the number of errors. Although most location-detection algorithms perform relatively well with a few or even without linguistic features, it is our contention that they fail to fully exploit the linguistic knowledge that permeates natural-language texts, e.g. locative prepositions (e.g. *in*, *at*, *near*, etc.), location-indicative nouns (e.g. *avenue*, *city*, *province*, *road*, *school*, *street*, etc.) or locative markers (e.g. *south of*, *XX kms away from*, etc.) that signal the presence of place names (Hoang & Mothe, 2018). This view emphasizes the need to develop rule-based location-detection systems that could improve state-of-the-art performance without requiring the significant amount of processing time and computational resources involved in ML and DL techniques (Gelernter & Balaji, 2013; Malmasi & Dras, 2016; Dutt et al., 2018; Middleton et al., 2018). In this context, the contribution of our research lies in the fine granularity of the extracted locative references, unlike previous NER and/or NEM models.

2.3. Location detection in Twitter

In this section, we present a typology of location-detection systems for Twitter, which have been classified according to three criteria: data extracted, data processed and model.

2.3.1. Type of data extracted

According to the type of data extracted from the tweet, location-detection systems can discover location mentions in Twitter messages, the user's location or the tweet location.

On the one hand, many systems have been developed to identify and extract the locative references that are mentioned in Twitter messages (Lingad et al., 2013; Gelernter & Balaji, 2013; Ghahremanlou et al., 2014; Han, Cook, et al., 2014; Malmasi & Dras, 2016; Inkpen et al., 2017; Al-Olimat et al., 2018; Avvenuti et al., 2018; Middleton et al., 2018; de Bruijn et al., 2018; Dutt et al., 2018; Hoang & Mothe, 2018; Karimzadeh et al., 2019; Kumar & Singh, 2019; Di Rocco et al., 2019; Hernandez-Suarez et al., 2019). For example, Hernandez-Suarez et al. (2019) and Di Rocco et al. (2019) managed to detect and geocode sub-city level locative references

(e.g. street, building), where the former used a DL algorithm based on biLSTM with a CRF top layer, and the latter used a knowledge-driven algorithm based on LinkedGeoData and openStreetMap Facet Ontology. Kumar & Singh (2019) proposed a system for the location detection of earthquake events by means of a supervised DL-based approach using a CNN without linguistic-feature engineering.

On the other hand, some systems are intended to detect the user's location on the basis of the user's profile and tweet history, and/or the tweet metadata (Alex et al., 2016; Cheng et al., 2010; Han et al., 2014; Li et al., 2011; Miyazaki et al., 2018). For example, Han et al. (2014) designed a system to predict the user's location at city level through location-indicative words in tweets and information from the user profile. Miyazaki et al. (2018) devised a knowledge-based neural network framework for Twitter user geolocation that exploits the user's tweet history with semantic relations from the Yago3 knowledge base (e.g. isLocatedIn, livesIn, happenedIn, etc.).

Finally, some systems have been designed to detect the tweet location (i.e. the location where the tweet was posted) for event geolocation (i.e. the location where a particular event took place) by analysing the user's profile and geotagged metadata (Sakaki et al., 2010; Priedhorsky et al., 2014; Chong & Lim, 2018; Gonzalez-Paule et al., 2019; Khodabandeh-Shahraki et al., 2019). For example, Gonzalez-Paule et al. (2019) devised a model that focuses on non-geotagged tweets by exploiting similarity content of geotagged tweets for traffic-incident detection. Khodabandeh-Shahraki et al. (2019) designed a model for event geolocation that considers multiple variables, such as the tweet message, the user's profile, geotagged metadata and posting time. As they noted, locative references in tweet messages might not always be a reliable variable to predict the location of an event.

2.3.2. Type of data processed

According to the type of data processed in the tweet, location-detection systems can rely on (a) the message (Lingad et al., 2013; C. Li & Sun, 2014; Ghahremanlou et al., 2014; Han, Jimeno-Yepes et al., 2014; Malmasi & Dras, 2016; Ikawa et al., 2016; Inkpen et al., 2017; Avvenuti et al., 2018; Middleton et al., 2018; Miyazaki et al., 2018; Dutt et al., 2018; Hernandez-Suarez et al., 2019; Karimzadeh et al., 2019), (b) the geotagged metadata (Li et al., 2011), (c) the user's profile information and/or user's tweet history (Cheng et al., 2010; Alex et al., 2016; Chong & Lim, 2018) or (d) a combination of the previous types of data (Sakaki et al., 2010; Dredze et al., 2013; Han, Cook, et al., 2014; Yin et al., 2014; Gonzalez-Paule et al., 2019).

2.3.3. Type of model

According to the type of model, location-detection systems can be based on (a) probabilistic models, such as ML or DL (Cheng et al., 2010; Sakaki et al., 2010; Lingad et al., 2013; Yin et al., 2014; Ghahremanlou et al., 2014; Han, Cook et al., 2014; Han, Jimeno-Yepes et al., 2014; Inkpen et al., 2017; Avvenuti et al., 2018; Miyazaki et al., 2018; Chong & Lim, 2018; Gonzalez-Paule et al.,

2019; Hernandez-Suarez et al., 2019), (b) symbolic models (Malmasi & Dras, 2016; Al-Olimat et al., 2018; Dutt et al., 2018) or (c) a combination of both (Gelernter & Balaji, 2013; Hoang & Mothe, 2018; Middleton et al., 2018).

3. Materials and methods

In this section, we introduce the development corpus, which became the cornerstone of the location-detection model. At this stage of the research, we also required a clear definition of our notion of “locative reference”, from which we could discover all the locative references in our development corpus. With the development corpus and a typology of locative references, we managed to construct the pipeline of our processing model.

3.1. Definition of locative reference

We define a locative reference as a subtype of named entity that designates a specific physically locatable geographic reference, i.e. one that can be pinpointed on a map (F. Liu et al., 2014; Gritta et al., 2018). Locative references can linguistically take the form of full words, abbreviations, acronyms, alphanumeric codes or a combination of them. Semantically, locative references can be classified into four main categories: geopolitical entities (e.g. *Beverly Hills*), natural geographic places (e.g. *Grand Canyon National Park*), POIs (e.g. *Paramount Pictures Studio*) and traffic ways (e.g. *Rodeo Drive*). With respect to the lexical units found within the boundaries of the linguistic realization of locative references, we differentiate between simple and complex locative references. Whereas a simple locative reference takes the form of a proper noun, a complex locative reference is represented by means of a proper noun that can be preceded and/or followed by one or more location-indicative nouns that, in turn, can be preceded by one or more locative markers (i.e. directional, distance or temporal) in combination with some prepositions. The following examples serve to illustrate a sample of the diversity of linguistic structures that can form locative references:

- China, New York, Buenos Aires (proper noun)
- 35miles from New York (number + locative marker [distance] + preposition + proper noun)
- South of Madrid (locative marker [directional] + preposition + proper noun)
- 1h away from London, 25min out of Melbourne (number + locative marker [temporal] + prepositions + proper noun)
- 57km SW of Cantwell (number + locative marker [distance] + locative marker [directional] + preposition + proper noun)
- Hotel Park Villa, Mount Everest (location-indicative noun + proper noun)
- coast of NZ (location-indicative noun + preposition + proper noun)

- Dyckman Street Station, Fox Valley Animal Referral Center (proper noun + location-indicative nouns)
- 10mins away from Mansion House (number + locative marker [temporal] + prepositions + location-indicative noun + proper noun)
- I 95 NB (proper noun + locative marker [directional])
- 4kms from Narok Town (number + locative marker [distance] + preposition + proper noun + location-indicative noun)

When it comes to defining what is not a locative reference, we need to refer to commonplace or informal locative expressions (Herskovits, 1985; F. Liu et al., 2014). These are phrasal chunks in the clause that contain unspecific and vague geospatial information. They usually appear in the form of (a) noun phrases containing common noun words or pronouns (e.g. *at home*, *in the garden*, *in front of you*, *on the street*) or (b) coreferential adverbs (e.g. *here*, *there*). We did not consider them as locative references because, as they cannot be pinpointed on a map without any further contextual clue, they do not provide sufficiently precise information for emergency and crisis events. Other cases that were discarded are demonyms, i.e. adjectives denoting nationality (e.g. British, Spanish).

3.2. Language resources

3.2.1. Development corpus

We compiled a development corpus of English tweets using FireAnt (Anthony & Hardaker, 2017) from a list of 7 keywords related to emergency and crisis events, e.g. *bombing attack*, *car accident*, *earthquake*, *flood*, *incident*, *shooting attack* and *terrorist attack*. The corpus was pre-processed automatically by removing (a) newline characters in multi-line tweets, so that each line represented a single tweet, and (b) duplicate tweets, so that each tweet was unique. This resulted in a corpus of 500 English tweets.

Following the criteria in the previous section, we obtained a list of manually annotated locative references from our development corpus, where a sample is shown in Tables 1 and 2, which are related by the tweet ID (see the tables on the next page).

As can be seen, each locative reference was assigned the ID number of the tweet from which it was extracted. This list was actually used as a gold standard to test the results generated by our location-detection model. Table 3 presents the number of locative references in terms of n-grams and Table 4 offers some statistics related to the development corpus.

Finally, Table 5 presents the locative references whose number of occurrences in the development corpus is 3 or higher.

TABLE 1

Sample of the locative-reference dataset

TWEET ID	LOCATIVE REFERENCE
1	North of New Gretna Toll Plaza
1	Garden State parkway NB
2	off-sunset boulevard
2	street 13
2	Karachi
2	Karachi
3	California
4	M4 Westbound
4	J33
4	Capel Llanilltern
4	J34
4	Miskin
4	Wales

TABLE 2

Sample of the tweet dataset

TWEET ID	TWEET
1	Cleared: Incident on #GardenStateParkway NB at North of New Gretna Toll Plaza
2	RT @naqvi1966: Another incident of police harassment at street 13 of off-sunset boulevard Karachi. Reportedly the squad of AIG Karachi stop.
3	RT @california: California parties trash. The DJ just said make some noise if u got earthquake insurance
4	#M4: Westbound: J33 Capel Llanilltern to J34 Miskin: Incident: Accident: Lanes closed: Delays #TrafficWalesAlert

TABLE 3

Form and distribution of locative references in the development corpus

No. of unigrams	213
No. of bigrams	109
No. of trigrams	48
No. of n-grams where $n \geq 4$	13
Total	383

TABLE 4

Locative-reference statistics in the development corpus

NO. OF LOCATIVE REFERENCES	NO. OF TWEETS WITH LOCATIVE REFERENCES	AVERAGE OF LOCATIVE REFERENCES PER TWEET	AVERAGE OF LOCATIVE REFERENCES PER LOCATION-RICH TWEET
383	199	0.77	1.92

TABLE 5

Most frequent locative references in the development corpus

LOCATIVE REFERENCE	CATEGORY	OCCURRENCES #
Iran	Geopolitical entity (country)	41
Edison station	POI (station)	4
Indonesia	Geopolitical entity (country)	4
Trenton station	POI (station)	4
Auburn	Geopolitical entity (city)	3
EB I-84	Traffic way	3
Fort Lauderdale	Geopolitical entity (city)	3
Garda	POI (headquarters)	3
Halifax	Geopolitical entity (city)	3
Halifax library	POI (library)	3
I 5 NB	Traffic way	3
New Zealand	Geopolitical entity (country)	3
Syria	Geopolitical entity (country)	3
US	Geopolitical entity (country)	3

3.2.2. Datasets

Our location-detection model is mostly grounded on four lexical resources: place-name dataset, location-indicative noun dataset, locative-marker dataset and stopword dataset.

The place-name dataset was derived from GeoNames⁵. The first step was the automatic pre-processing of the list of geographical names in GeoNames. In particular, we conducted three consecutive tasks:

5 GeoNames dump files (approx. 1.92 GB) were downloaded from <https://www.geonames.org> on 13 February 2019.

- (i) Setting the language parameter for English location names only.
- (ii) Retrieving location names, consisting of 1- up to 8-grams, whose population size is greater than 100 inhabitants. The population-size filter served to dramatically decrease the rate of false positives, since it managed to discard place names that would by chance match common words. However, the application of this filter slightly increased the number of false negatives.
- (iii) Removing names of historical places that no longer exist, which are marked by the tag “historical” (e.g. ancient Roman provinces).

This pre-processing step generated a much smaller file of only 12.4 MB, which greatly contributed to speeding up the performance of our model.

The location-indicative noun dataset was automatically constructed from WordNet (Miller, 1995; Fellbaum, 1998) by taking all the lexical units linked to the senses “road.n.01”, “building.n.01”, “facility.n.01”, “junction.n.01”, “district.n.01”, “area.n.01”, “geological_formation.n.01”, “body_of_water.n.01”, “tract.n.01”, “way.n.06” and “beach.n.01”, or to any of their subordinate synsets. Duplicates and more-than-two-word lexical units were discarded. Then, the dataset was also filtered by removing (a) n-grams containing named entities (e.g. *Baltic state*, *French region*, etc.) and (b) n-grams whose locative meaning is not self-evident (e.g. *bed*, *melting pot*, *scene of action*, etc.). Whereas the filter (a) was automatically applied by means of a regular expression, the words in (b) were manually removed. In the end, the dataset, containing 1217 lexical items, was supplemented with a list of traffic-way and other place abbreviations obtained from the US postal service, thus making a total of 1766 items⁶. Table 6 shows a sample of location-indicative nouns classified according to the type of locative reference.

The locative-marker dataset was manually constructed. Locative markers can be classified into directional, distance or temporal markers, some of which are illustrated in Table 7.

The locative-marker dataset contains 56 directional markers, 4 distance markers and 2 temporal markers.

Finally, the stopword dataset contains the 5000 most frequent English words from the Corpus of Contemporary American English (COCA)⁷ together with a list of the 5500 most common names and surnames⁸.

6 The list can be found on http://cool.conservation-us.org/lex/abbr_suf.html.

7 The 5000 most frequent English words were retrieved from the COCA on <https://www.wordfrequency.info/>.

8 The names and surnames were compiled from <https://names.mongabay.com/> and <https://surname.sofeminine.co.uk/w/surnames/most-common-surnames-in-great-britain.html>. We filtered out the proper nouns that matched the names of cities and countries (e.g. *Nevada*, *Verona*, *Milan*, *Paris*, *Kenya*, *Valencia*, etc.).

TABLE 6

Sample of location-indicative nouns

GEOPOLITICAL ENTITIES	NATURAL GEOGRAPHIC REFERENCES	POIS	TRAFFIC WAYS
barrio	beach	art school	alley
caliphate	canyon	bus station	avenue
city	gulf	café	boulevard
country	hill	castle	driveway
county	lake	cathedral	freeway
jurisdiction	mountain	embassy	highway
province	ridge	hospital	parkway
region	river	hotel	road
state	valley	residence	street
town	volcano	university	turnpike

TABLE 7

Sample of locative markers

DIRECTIONAL MARKERS	DISTANCE MARKERS	TEMPORAL MARKERS
North, N	kilometre, km	hour, hr, h
South, S	metre, m	minute, min
East-North-East, ENE	mile, mi	
Southwest, SW	yard, yd	
South-East, SE		
Eastbound, EB		
Western		

3.3. Location-detection model

The pipeline of our location-detection model consists of four main stages: (a) pre-processing, (b) tokenization and POS tagging, (c) place-name search and (d) linguistic processing.

3.3.1. Pre-processing

Several tasks were performed in the pre-processing stage: (a) removing emojis, genitive mark- and unwanted white spaces, (b) replacing user mentions and URLs by the tokens “user” and

“url”, respectively, and (c) replacing hashtags by the words contained therein (e.g. #GardenStateParkway was replaced by *Garden State Parkway*).

3.3.2. Tokenization and POS tagging

In this stage, the Stanford POS tagger was used to split the tweet into tokens and identify the word form and POS tag corresponding to each token. Thus, tweets are represented as a sequence of objects with attributes such as position, token, word form and POS.

3.3.3. Place-name search

The goal of this stage is to match n-grams of different size extracted from the tweets with the nouns in the place-name dataset. In particular, n-grams in the tweet are searched in order of decreasing size. Therefore, if no match is found with a given 8-gram, then the model traverses its embedded n-grams in depth-first search, until a match takes place or unigrams are reached. Before the search is performed, two tasks are carried out:

- In the case of unigrams, the system discards those that (a) have not been tagged as proper nouns, or (b) are found in the stopword dataset, the location-indicative noun dataset, or the locative-marker dataset.
- In the case of bigrams, the system discards those that do not match the following pattern: noun + proper noun (e.g. *the country, beautiful isle, nice airport*).

3.3.4. Linguistic processing

For the purpose of discovering new locative references or expanding locative references detected in the previous or current stage, the system relies on contextual clues to perform three consecutive tasks. Indeed, an exhaustive linguistic analysis of the development corpus resulted in the discovery of a variety of linguistic patterns. Subsequently, we managed to formulate regex-based rules from these patterns, so that they were able to retrieve most of the manually annotated locative references.

The first task consists in searching for proper nouns not included in the place-name dataset that are introduced by locative prepositions, which can be found as far as four positions to the left of the proper noun. For example:

- (1) [...] Our car skidded on loose gravel, lost control and flipped at Lorubuko.

It is noteworthy that this module can now retrieve acronyms and abbreviations of place names that were ignored in the place-name search:

- (2) So gay couple in severe car accident in TX [...]

The locative prepositions used in this stage are *@*, *across*, *along*, *at*, *in* and *near*. The reason why we do not include other prepositions that signal location and direction (e.g. *on*, *to*, *from*) is because these tend to generate many false positives, particularly when they introduce indirect objects (e.g. *John gave a present to Mary*) or oblique objects (e.g. *I received a present from John*).

The second task serves to expand existing locative references or discover new locative references with one or several location-indicative nouns to be found within a range of four positions to the left or right of the proper noun. As shown in the following examples, location-indicative nouns and proper nouns should be placed together:

- (3) Recent case of people jumping the shark is the Batavia High School incident. [...]
- (4) Fracking = Earthquakes? "The Changing shale gas block in the South Sichuan Basin has been the site of fracking operation."
- (5) 4.6 magnitude #earthquake. Central Mid-Atlantic Ridge

Alternatively, the preposition *of* can be placed between the location-indicative word and the proper noun, as in the following example:

- (6) Earthquake swarm similar to what preceded 2011 Japanese quake and tsunami happening off coast of NZ

Moreover, when an Arabic numeral appears in the left-most position within the window, the system stops searching, since it signals an address number:

- (7) *ACCIDENT: DAMAGE ONLY* - RALEIGH POLICE – 8800 GLENWOOD AVE

It should be noted that, if locative references detected in the place-name search stage were wrongly delimited, i.e. locative references that were shorter than expected (e.g. *High School* instead of *Batavian High School*, *Sichuan* instead of *South Sichuan Basin*, or *Glenwood* instead of *Glenwood Ave*, to name but a few), their boundaries would now be expanded. For example, this allows our model to detect locative references consisting of the name of a road route together with a directional marker on the left or right:

- (8) KAIMUKI: HPD on scene of accident on H1 WB [...]

In this regard, road routes are detected by means of a regular expression based on knowledge from Wikipedia entries containing the commonest naming conventions for roads and highways in English-speaking countries such as UK, US, Australia, Canada and India.

- (9) HAPPENING NOW: A vehicle crash on I-75 MM 377 has blocked all northbound lanes [...]

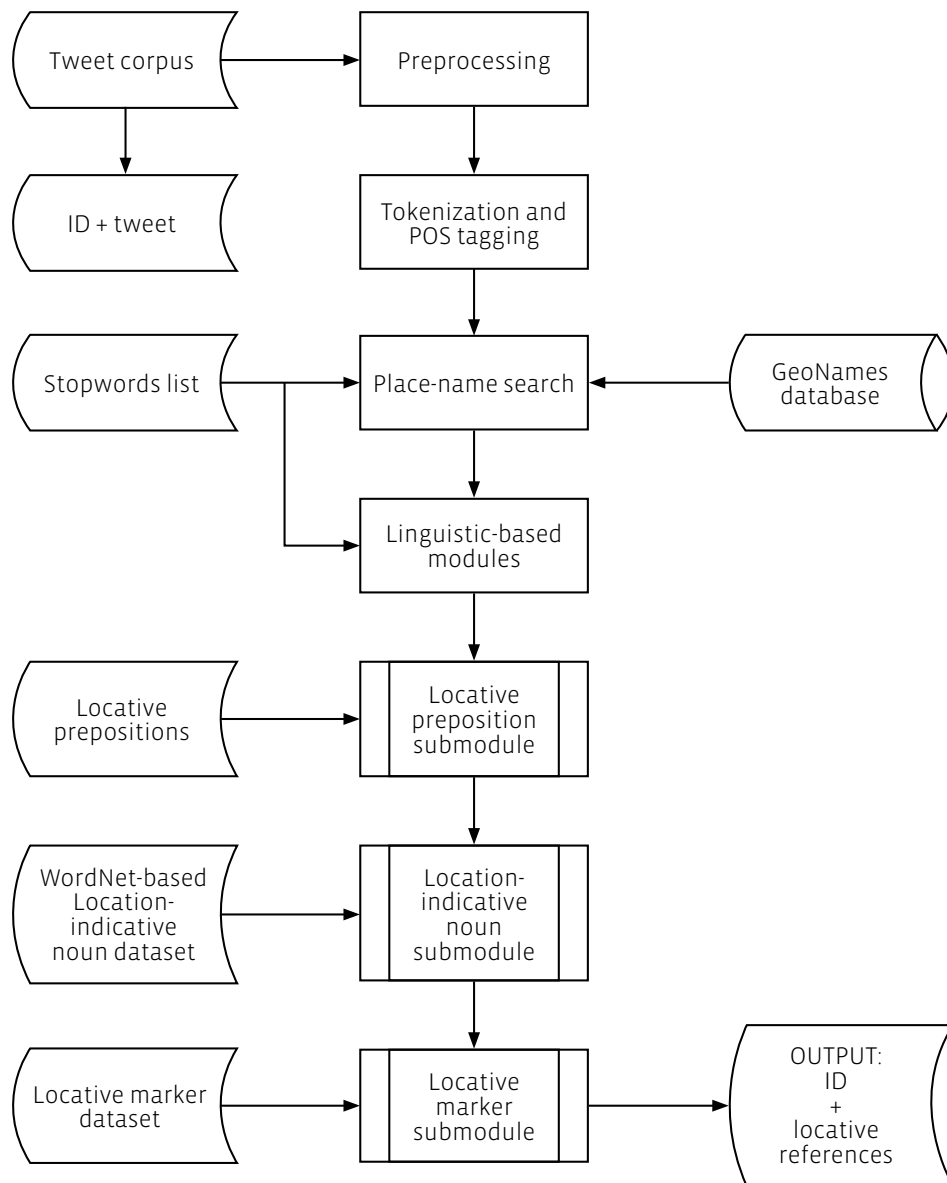
For the third task, we built lexico-syntactic rules to combine the various types of locative markers with the purpose of determining the full scope of complex locative references:

- (10) Incident on #GardenStateParkway NB at North of New Gretna Toll Plaza
- (11) #Earthquake Reported: M 5.1 - 66km NW of Kota Ternate, Indonesia [...]

To summarize, Figure 1 shows the flowchart that describes our location-detection model.

FIGURE 1

Flowchart of the location-detection pipeline



4. Evaluation

4.1. Test corpus

For the evaluation of LORE, we collected a large and representative dataset with FireAnt from the same seven keywords used in the construction of the development corpus. After pre-processing (i.e. removal of newline characters and duplicate posts), we managed to have a test corpus of 800 tweets. Table 8 presents the distribution of locative references in terms of n-gram size in the test corpus and Table 9 shows some statistics regarding the nature of the test corpus.

TABLE 8

Form and distribution of locative references in the test corpus

No. of unigrams	264
No. of bigrams	190
No. of trigrams	60
No. of n-grams where $n \geq 4$	23
Total	383

TABLE 9

Locative-reference statistics in the test corpus

NO. OF LOCATIVE REFERENCES	NO. OF TWEETS WITH LOCATIVE REFERENCES	AVERAGE OF LOCATIVE REFERENCES PER TWEET	AVERAGE OF LOCATIVE REFERENCES PER LOCATION-RICH TWEET
537	259	0.67	2.07

To illustrate, Table 10 (on the next page) provides the most frequent locative references in the test corpus, particularly those whose number of occurrences is 4 or higher.

4.2. Results

For the evaluation of our model, we employed measures that are most widely used in information retrieval, i.e. precision (P), recall (R) and F1, which are computed as follows:

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN}$$

$$F1 = 2 * \frac{P * R}{P + R}$$

where TP, FP and FN refer to the number of true positives, false positives and false negatives, respectively. F1 is the harmonic mean of precision and recall. These measures generate a score that ranges from 0 to 1.

The evaluation process was performed on the test corpus following the metrics presented above. We present the evaluation results on both a per-entity basis and a per-token basis (Gritta et al., 2018). On the one hand, the entity-based evaluation only considers full locative references as TP, being the commonest and strictest evaluation method for benchmarking NER systems (Jurafsky & Martin, 2020). On the other hand, the token-based evaluation works more leniently, since partial or inexact matches count not only as TP but also as either FP or FN; for example, it is FP when the extracted instance exceeds the boundaries of the locative reference (e.g. *OFF EAST COAST OF HONSHU* for *EAST COAST OF HONSHU*) and FN when the boundaries of the instance fall short (e.g. *Camino* for *Camino Pablo*). Table 11 shows the P, R and F1 scores of both types of evaluation on the test corpus (see the table on the next page).

As can be noted, the evaluation is performed on the two main modules, i.e. place-name search and linguistic processing, individually and in combination to observe their contributions in the performance of the system. The best results for each type of evaluation are highlighted in bold.

TABLE 10

Most frequent locative references in the test corpus

LOCATIVE REFERENCE	CATEGORY	OCCURRENCES #
Iran	Geopolitical entity (country)	20
India	Geopolitical entity (country)	11
Pulwama	Geopolitical entity (city)	6
San Bernadino	Geopolitical entity (city)	5
J18	Traffic way	5
Japan	Geopolitical entity (country)	5
M74	Traffic way	5
New Zealand	Geopolitical entity (country)	4
Grapevine	Geopolitical entity (city)	4
Pakistan	Geopolitical entity (country)	4
Sr4 E	Traffic way	4
Kingston	Geopolitical entity (district)	4
Balakot	Geopolitical entity (town)	4

TABLE 11

Evaluation of LORE

	TOKEN-BASED EVALUATION			ENTITY-BASED EVALUATION		
	P	R	F1	P	R	F1
Only with place-name search	0.84	0.48	0.61	0.59	0.40	0.47
Only with linguistic processing	0.90	0.56	0.69	0.86	0.52	0.65
Place-name search + linguistic processing	0.85	0.83	0.84	0.81	0.81	0.81

We also compared the performance of our model against well-known open-source human language technology tools (e.g. Stanford CoreNLP, NLTK, spaCy and OpenNLP). For this experiment, we employed the same test corpus processed with the same computer hardware (i.e. i5-6200U @ 2.30 GHz with 2 cores and 8GB RAM). Before presenting the results of the performance tests, we provide a brief, technical description of the different tools:

- Stanford CoreNLP 3.9.2 is a Java suite of NLP tools, including the POS tagger, the NER and the parser, among others. The Stanford NER, which implements a probabilistic algorithm based on a Conditional Random Field linear classifier (Finkel et al., 2005)⁹, makes use of the labels PERSON, LOCATION and ORGANIZATION for named entities. The Stanford NER for English is trained on news corpora from CoNLL 2003, MUC 6 and MUC 7, ACE 2002 and additional data. In our experiment, we only considered LOCATION-labelled entities for the extraction of locative references.
- Natural Language Toolkit (NLTK) 3.4.4 is a Python library for a wide variety of NLP tasks, such as tokenization, lemmatization, POS tagging, chunking, NER, semantic tagging and parsing, among others (Bird, 2006)¹⁰. The NER module, which is based on a Maximum Entropy algorithm trained on the ACE corpus, employs the labels ORGANIZATION, PERSON, LOCATION, DATE, TIME, MONEY, PERCENT, FACILITY and GPE for named entities. For our experiment, the only relevant categories were FACILITY (for POIs), GPE (for geopolitical entities) and LOCATION for the remaining locative references.
- spaCy 2.1.6. is a widely used Python library for many advanced NLP tasks for software production (Honnibal & Montani, 2017)¹¹. The NER component for English makes use of a deep-learning algorithm (i.e. Convolutional Neural Networks) trained on OntoNotes 5.0, a large corpus comprising various genres of text, e.g. news, telephone

9 <https://sergey-tihon.github.io/Stanford.NLP.NET/>

10 <http://nltk.org/>

11 <https://spacy.io/>

speech, talk shows, etc. spaCy can recognize many named-entity types, of which we selected GPE (for geopolitical entities), FAC (for POIs) and LOC for the remaining locative references.

- OpenNLP is a C# tool for basic NLP tasks such as sentence splitting, tokenization, POS tagging, chunking or NE¹². The NER component is based on a Maximum Entropy model trained on a variety of corpora, such as MUC6, MUC7, ACE, CONLL 2002 and CONLL 2003. The location types are DATE, LOCATION, MONEY, ORGANIZATION, PERCENTAGE, PERSON and TIME, of which LOCATION was the only one considered in our experiment.

Table 12 and Table 13 display the results of the performance tests in terms of token- and entity-based evaluation (i.e. P, R and F1) and processing speed, respectively.

TABLE 12

Evaluation of NER systems

NER SYSTEM	TOKEN-BASED EVALUATION			ENTITY-BASED EVALUATION		
	P	R	F1	P	R	F1
LORE	0.85	0.83	0.84	0.81	0.81	0.81
Stanford NER	0.89	0.42	0.57	0.79	0.37	0.50
NLTK	0.55	0.29	0.38	0.43	0.24	0.31
spaCy	0.75	0.33	0.46	0.66	0.28	0.39
OpenNLP	0.73	0.27	0.40	0.56	0.21	0.30

TABLE 13

Processing speed of NER systems

NER SYSTEM	PROCESSING SPEED (MM:SS.SSS)
LORE	00:08.695
Stanford NER	00:09.822
NLTK	00:10.886
spaCy	00:12.151
OpenNLP	03:35.100

12 <https://github.com/AlexPoint/OpenNlp>

4.3. Discussion

LORE performs better when the place-name search and linguistic processing modules are integrated (Table 11), which also contributes to outperform other NER systems (Table 12). On the one hand, the place-name search module helped improve R, although it had a slight impact on the number of FPs. On the other hand, the regex-based lexico-syntactic rules with the help of the location-indicative noun dataset helped identify not only complex locative references but also specific POIs and traffic ways that typically go unnoticed, as illustrated in (12), (13) and (14).

- (12) Cleared: Motor Vehicle Accident - HARTFORD #I84 West 0.02 miles before Exit 51 (I-91NB) at 4/11/2019 10:56:03 AM
- (13) Extremist Pleads Guilty to Planning Mass Shooting Attack at Texas Mall
- (14) South LA 13219 S Penrose Ave **Hit and Run No Injuries**

It is noteworthy that the linguistic-processing stage alone achieved the highest P but the lowest R, since it cannot retrieve locative references if not signaled by markers. In fact, this behavior is in line with the results typically achieved by rule-based NER systems. For example, most geopolitical entities in hashtags did not provide any location-signaling clue for regex-based rules, so only the place-name search module could detect them, as in (15), (16) and (17).

- (15) #Incident #Ottawa #HWY417 WB at Metcalfe St (IC 119A), 2 left lanes [...]
- (16) CLEAR - #BCHwy10 EB vehicle incident at #BCHwy91 overpass. #DeltaBC
- (17) Accident with injury in #EastBatonRouge on Airline SB at I 12 #traffic

Only with the integration of both modules, the system achieved a trade-off between P and R.

Finally, with respect to the evaluation results of the other NER systems (Table 12 and Table 13), we can conclude that:

- LORE became one of the best systems in relation to P, being the best in the entity-based evaluation;
- LORE clearly outperformed in R and F1, and
- the processing speed of LORE surpassed the others by a few seconds.

There was clear evidence that the low scores in R and F1 provided by the other NER systems was due to a lack of granularity which was not able to adequately address the full semantics of locative references. In other words, although they excelled in the identification of place names such as geopolitical entities (mainly towns, cities and countries), many could not sat-

isfactorily (a) detect most natural geographic references, POIs and traffic ways, nor (b) cope with the complexity of wide-range locative references.

4.4. Analysis of errors

We also analyzed the commonest sources of error in our model (i.e. place-name search together with linguistic processing), providing an explanation of their occurrence as well as some possible solutions and alternatives that we leave for future research. The automated process of locative-reference detection made errors of omission and errors of commission. An error of omission occurred every time the model failed to detect a true locative reference, whereas an error of commission occurred every time the model retrieved an instance that was actually not a locative reference.

4.4.1. Errors of omission

After applying the population-size filter, the place-name dataset became a simplified version of the GeoNames database. However, this decision increased the number of FNs, particularly when lexico-syntactic rules were not effective in the linguistic processing stage. This was the case of (18), where *Indinapuram* was not detected as an instance of geopolitical entity.

(18) @MORTHRoadSafety Pls consider asking the #NHA1 to close the central verge on #NH24 between #Indirapuram and [...]

Obviously, making use of the whole GeoNames database could have decreased the number of FNs and, therefore, recall could have increased, but at the expense of many more cases of FP, which could have dramatically affected precision. Considering the benefit-cost ratio, we opted for the population-size restriction.

Another source of error has to do with misspellings or lack of capitalization of proper nouns, e.g. in (19).

(19) 00:36 Magpie Swamp Rd/mingbool Rd, Pleasant Park - Tree Down going (one appliance, CFS region 5)

In these cases, proper nouns were labelled as common nouns by the Stanford POS tagger, but a key feature of all locative references is that they must contain at least one proper noun. Therefore, the system could not avoid these cases of FNs, which are highly dependent on the performance of third-party POS taggers. Additionally, we tried to use SymSpell for text normalization¹³, which a priori could enhance the performance of the POS tagger and

13 <https://github.com/wolfgarbe/SymSpell>

thus avoid missing locative references. However, we soon realized that processing became 3 times slower. Therefore, considering again the benefit-cost ratio, we preferred not to perform text normalization. Perhaps by using a Twitter-specific POS tagger, we could reduce the number of FNs with no impact on the processing speed of the model, but this is an issue for future research.

Finally, despite not being present in our test corpus, we are aware of the existence of complex locative expressions other than those containing locative markers and/or location-indicative nouns. For example, this is the case of coordinated locative references (e.g. *in the US and the UK, between Madrid and Barcelona*, etc.) and other more complex locative expressions (e.g. *close to London but not far away from Croydon*). In fact, this is a quite challenging issue. However, it presents problems to the current model, because the linguistic patterns that underlie such locative references are so obscure that rule formalization does not seem to be a manageable task. A syntactic parser helps delimit the phrasal boundaries of locative references, but we realized that it dramatically slowed down the system and did not offer much improvement. For now, locative references found in such complex expressions can be identified individually.

4.4.2. Errors of commission

On the one hand, the place-name dataset is prone to overmatching. For example, in (20) the pet name *Nemo* was wrongly identified as a locative reference, since it is also the name of some populated places in Liberia, Mozambique and USA¹⁴.

(20) Happy #NationalPetDay and I really miss my Nemo (cat) passed away in car accident [...]

Despite our efforts to mitigate FPs by leveraging stopword lists and different lexico-syntactic rules in either of the two modules, we blame most cases of FP on the performance of the Stanford POS tagger, which sometimes considered common nouns and other parts of speech to be proper nouns because of wrong capitalization patterns. This is illustrated in examples (21) and (22).

(21) y'ALL THIS AU IS SPOT ON!!! IT'S SO BEAUTIFULLY MADE I CRIED OMGGGGGGGGGGGGG

(22) @MckarloFernan Honestly go awf fam ???? if the put cheese in by accident next time call me and I'll eat it for you LOL

14 <https://www.geonames.org/2274775/nemo.html>, <https://www.geonames.org/1032172/nemo.html> and <https://www.geonames.org/4714030/nemo.html>, respectively.

MADE and LOL were mislabelled as locative references in (21) and (22), respectively, because Made is a Dutch and also an Indonesian village¹⁵ and Lol is one of the states of South Sudan¹⁶. The false locative references in (21) and (22) can be explained by the confluence of three consecutive factors: (a) they were first captured in the place-name dataset, (b) they were tagged as proper nouns and finally (c) they bypassed the regex-based rules and the stopword filtering.

On the other hand, the location-indicative noun dataset also led to the extraction of false locative references, as in (23), (24) and (25).

(23) @BrianBLevinson I like how they list Lief Green next to Jim Greenleaf. No way that was by accident.

(24) @manishinsha93: #RoadSafetyInitiativeByDSS Saint Dr. MSG has come up with the initiative to tie reflector belts on the stray animals

(25) CLEARED HUDSON VALLEY: Slow traffic [...]

In (23), the location-indicative noun *green*, which denotes an area of land covered with grass, mismatched *Green* in the tweet and, since the system found that the previous word was a proper noun, *Lief Green* was extracted as a locative reference. In (24), *dr* mismatched the abbreviation of the location-indicative noun *drive*, which is the same as that of *doctor*, and all the preceding words were wrongly taken as proper nouns. In (25), the location boundary was wrongly delimited, because *CLEARED* was tagged as proper noun. Therefore, POS mistagging raised this type of errors.

Finally, other source of error had to do with the lexico-syntactic rules in the linguistic-processing module, which caused the wrong retrieval of sequences containing address numbers, as illustrated in (26).

(26) @thatpreacha: @TalbertSwan The 1st church burned, everyone thought it could have been an accident. After the 2nd church burned, deacons.

5. Conclusions

The extraction of geolocation information from Twitter is a key part in intelligent systems for emergency and crisis management services. For example, the location dimension is crit-

15 <https://www.geonames.org/2751272/made.html> and <https://www.geonames.org/6407244/made.html>, respectively.

16 <https://www.geonames.org/11550548/lol.html>

ical for raising situation awareness of disaster events and understanding their impact, i.e. where the incident happened, what areas were affected and where people are in need of help. Extracting such critical information from tweets could help emergency responders allocate material and human resources more effectively.

With respect to geographic information retrieval, and particularly to geoparsing, this article presented LORE, a model that exploits rich linguistic knowledge together with different NLP techniques to detect locative references in tweets. In particular, the integration of lexically rich datasets with text-processing tasks (e.g. tokenization and POS tagging) and regex-based rules helped identify coarse-grained locations such as cities, towns and countries as well as fine-grained locations such as addresses, buildings, roads, landforms, etc. LORE can also extract complex locative references, made up of any location-indicative word(s) and/or locative marker(s) accompanying a given place name. This semantic granularity constitutes in itself a great qualitative advantage over current NER models. In addition, our experiments demonstrated that LORE outperforms state-of-the-art approaches not only in precision but also in recall.

The architecture of LORE, which consists of two primary modules (i.e. place-name search and linguistic processing), is able to provide similar performance with any collection of tweets about any domain. Moreover, its modularity facilitates the adaptation of our model to other languages, making it ideal for multilingual contexts. In fact, we are currently conducting research to provide support not only to English but also to French, Italian and Spanish, constructing language-specific components (i.e. lexical datasets and rules) by means of semi-automatic methods. Finally, another issue for future research is to enhance POS tagging. Since our model relies on proper nouns to delimit locative references, a Twitter-specific POS tagger could help not only recognize many locative references that go unnoticed due to wrongly assigned POS tags but also discard words that are mislabelled as proper nouns.

6. Acknowledgements

Financial support for this research has been provided by the Spanish Ministry of Science, Innovation and Universities [grant number RTC 2017-6389-5], and the European Union's Horizon 2020 research and innovation program [grant number 101017861: project SMARTLAGOON]. We also thank Universidad de Granada for their financial support to the first author through the *Becas de Iniciación para estudiantes de Máster 2018 del Plan Propio de la UGR*.

7. Bibliographic references

ANTHONY, Laurence, and Claire HARDAKER, 2017: FireAnt (version 1.1.4) [Computer Software], Tokyo: Waseda University [<https://www.laurenceanthony.net/software>].

ACHESON, Elise, Stefano DE SABBATA and Ross S. PURVES, 2017: "A quantitative analysis of global gazetteers: Patterns of coverage for common feature types", *Computers, Environment and Urban Systems* 64, 309-320.

AGUILAR, Gustavo, Suraj MAHARJAN, Adrian Pastor LÓPEZ MONROY and Tamar SOLORIO, 2018: "A multi-task approach for named entity recognition in social media data" in *Proceedings of the 3rd Workshop on Noisy User-generated Text*, Copenhagen: Association for Computational Linguistics, 148-153.

AHLERS, Dirk, 2013: "Assessment of the accuracy of GeoNames gazetteer data" in *Proceedings of the 7th Workshop on Geographic Information Retrieval - GIR '13*, 74-81.

AL-OLIMAT, Hussein S., Krishnaprasad THIRUNARAYAN, Valerie SHALIN and Amit SHETH, 2018: "Location name extraction from targeted text streams using gazetteer-based statistical language models" in *Proceedings of the 27th International Conference on Computational Linguistics*, Association for Computational Linguistics, 1986-1997.

ALEX, Beatrice, Clare LLEWELLYN, Claire GROVER, Jon OBERLANDER and Richard TOBIN, 2016: "Homing in on Twitter users: Evaluating an enhanced geoparser for user profile locations" in *Proceedings of the 2016 International Conference on Language Resources and Evaluation*, 3936-3944.

AVVENUTI, Marco, Stefano CRESCHI, Leonardo NIZZOLI and Maurizio TESCONI, 2018: "Geoparsing and geotagging with machine learning on top of linked data" in *Proceedings of the 15th Extended Semantic Web Conference*, 17-32.

BARRIÈRE, Caroline, 2016: *Natural Language Understanding in a Semantic Web Context*, Cham: Springer, 23-38.

BIRD, Steven, 2006: "NLTK: The Natural Language Toolkit" in *Proceedings of the COLING/ACL 2006 Interactive Presentation Sessions*, Sidney: Association for Computational Linguistics, 69-72.

CHENG, Zhiyuan, James CAVERLEE and Kyumin LEE, 2010: "You are where you tweet: A content-based approach to geo-locating Twitter users" in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, 759-768.

CHONG, Wen-Haw, and Ee-Peng LIM, 2018: "Exploiting user and venue characteristics for fine-grained tweet geolocation", *ACM Transactions on Information Systems* 36 (3), 1-34.

CROOKS, Andrew, Arie CROITORU, Anthony STEFANIDIS and Jacek RADZIKOWSKI, 2013: "#Earthquake: Twitter as a distributed sensor system", *Transactions in GIS* 17 (1), 124-147.

DE BRUIJN, Jens A., Hans DE MOEL, Brenden JONGMAN, Jurjen WAGEMAKER and Jeroen C.J.H. AERTS, 2018: "TAGGS: Grouping tweets to improve global geoparsing for disaster response", *Journal of Geovisualization and Spatial Analysis* 2 (2), 1-14.

DI ROCCO, Laura, Davide BUSCALDI, Michela BERTOLOTTI, Barbara CATANIA and Giovanna GUERRINI, 2019: "The role of geographic knowledge in sub-city level geolocation" in *Proceedings of the 34th ACM/SIGAPP Symposium on Applied Computing*, Limassol, 1-7.

DREDZE, Mark, Michael J. PAUL, Shane BERGSMAN and Hieu TRAN, 2013: "Carmen: A twitter geolocation system with applications to public health" in *Proceedings of the AAAI 2013 Workshop on Expanding the Boundaries of Health Informatics Using Artificial Intelligence*, 20-24.

DUGAS, Fabrice, and Eric NICHOLS, 2016: "DeepNNER: Applying BLSTM-CNNs and extended lexicons to named entity recognition in tweets" in *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, Osaka, 178-187.

DUTT, Ritam, Kaustubh HIWARE, Avijit GHOSH and Rameshwar BHASKARAN, 2018: "SAVITR: A system for real-time location extraction from microblogs during emergencies" in *Proceedings of the 2018 Web Conference Companion*, Lyon, 1643-1649.

ESPINOSA, Kurt Junshean, Riza BATISTA-NAVARRO and Sophia ANANIADOU, 2016: "Learning to recognise named entities in tweets by exploiting weakly labelled data" in *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, Osaka, 153-163.

FELLBAUM, Christiane (ed.), 1998: *WordNet: An electronic lexical database*, Cambridge, MA: MIT Press.

FINKEL, Jenny Rose, Trond GRENAGER and Christopher MANNING, 2005: "Incorporating non-local information into information extraction systems by Gibbs sampling", *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL '05 (1995)*, 363-370.

GELERNTER, Judith, and Shilpa BALAJI, 2013: "An algorithm for local geoparsing of microtext", *Geoinformatica* 17 (4), 635-667.

GERGUIS, Michel Naim, Cherif SALAMA and M. WATHEQ EL-KHARASHI, 2016: "ASU: An experimental study on applying deep learning in twitter named entity recognition" in *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, Osaka, 188-196.

GHAHREMANLOU, Lida, Wanita SHERCHAN and James A. THOM, 2014: "Geotagging Twitter messages in crisis management", *Computer Journal* 58 (9), 1937-1954.

GONZALEZ-PAULE, Jorge David, Yeran SUN and Yashar H, 2019: "On fine-grained geolocalisation of tweets and real-time traffic incident detection", *Information Processing and Management* 56 (3), 1-14.

GOYAL, Archana, Vishal GUPTA and Manish KUMAR, 2018: "Recent named entity recognition and classification techniques: A systematic review", *Computer Science Review* 29, 21-43.

GRITTA, Milan, Mohammad Taher PILEHVAR and Nigel COLLIER, 2020: "A pragmatic guide to geoparsing evaluation", *Language Resources and Evaluation* 54 (3), 683-712

GRITTA, Milan, Mohammad Taher PILEHVAR, Nut LIMSOPATHAM and Nigel COLLIER, 2018: "What's missing in geographical parsing?", *Language Resources and Evaluation* 52 (2), 603-623.

HAN, Bo, Paul COOK and Timothy BALDWIN, 2014: "Text-based Twitter user geolocation prediction", *Journal of Artificial Intelligence Research* 49, 451-500.

HAN, Bo, Antonio JIMENO-YEPES, Andrew MACKINLAY and Qiang CHEN, 2014: "Identifying Twitter location mentions" in *Proceedings of the Australasian Language Technology Association Workshop 2014*, Melbourne, 157-162.

HERNANDEZ-SUAREZ, Aldo, Gabriel SANCHEZ-PEREZ, Karina TOSCANO-MEDINA, Hector PEREZ-MEANA, JOSE PORTILLO-PORTILLO, Victor SANCHEZ-AND-LUIS and Luis Javier GARCIA-VILLALBA, 2019: "Using Twitter data to monitor natural disaster social dynamics: A recurrent neural network approach with word embeddings and kernel density estimation", *Sensors* 19 (7), 1-22.

HERSKOVITS, Annette, 1985: "Semantics and pragmatics of locative expressions", *Cognitive Science* 9 (3), 341-378.

HOANG, Thi Bich Ngoc, and Josiane MOTHE, 2018: "Location extraction from tweets", *Information Processing and Management* 54 (2), 129-144.

HONNIBAL, Matthew, and Ines MONTANI, 2017: spaCy 2 [Computer Software], [<https://spacy.io>].

IKAWA, Yohei, Maja VUKOVIC, Jakob ROGSTADIUS and Akiko MURAKAMI, 2016: "Location-based insights from the social web" in *Proceedings of the 22nd International Conference on World Wide Web*, Rio de Janeiro, 1013-1016.

IMRAN, Muhammad, Carlos CASTILLO, Fernando DIAZ and Sarah VIEWEG, 2014: "Processing social media messages in mass emergency: A survey", *ACM Computing Surveys* 47 (4), 1-38.

INKPEN, Diana, Ji LIU, Atefeh FARZINDAR, Farzaneh KAZEMI and Diman GHAZI, 2017: "Location detection and disambiguation from twitter messages", *Journal of Intelligent Information Systems* 49 (2), 237-253.

JONGMAN, Brenden, Jurjen WAGEMAKER, Beatriz ROMERO and Erin DE PEREZ, 2015: "Early flood detection for rapid humanitarian response: Harnessing near real-time satellite and Twitter signals", *ISPRS International Journal of Geo-Information* 4 (4), 2246-2266.

JURAFSKY, Daniel, and James H. MARTIN (3rd ed. draft), 2020: *Speech and Language Processing* [https://web.stanford.edu/~jurafsky/slp3/].

KARIMZADEH, Morteza, Scott PEZANOWSKI, Alan M. MAC EACHREN and Jan O. WALLGRÜN, 2019: "GeoTxt: A scalable geoparsing system for unstructured text geolocation", *Transactions in GIS* 23 (1), 118-136.

KHODABANDEH-SHAHRAKI, Zahra, Afsaneh FATEMI, and Hadi TABATABAEE MALAZI, 2019: "Evidential fine-grained event localization using Twitter", *Information Processing and Management* 56 (6), 102045.

KUMAR, Abhinav, and Jyoti Prakash SINGH, 2019: "Location reference identification from tweets during emergencies: A deep learning approach", *International Journal of Disaster Risk Reduction* 33, 365-375.

LEIDNER, Jochen, and Michael LIEBERMAN, 2011: "Detecting geographical references in the form of place names and associated spatial natural language", *SIGSPATIAL Special* 3 (2), 5-11.

LI, Chenliang, and Aixin SUN, 2014: "Fine-grained location extraction from tweets with temporal awareness" in *Proceedings of the 37th International ACM SIGIR Conference on Research & Development In Information Retrieval*, 43-52.

LI, Wen, Pavel SERDYUKOV, Arjen P. DE VRIES, Carsten EICKHOFF and Martha LARSON, 2011: "The where in the tweet" in *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 2473-2476.

LIMSOPATHAM, Nut, and Nigel COLLIER, 2016: "Bidirectional LSTM for named entity recognition in Twitter messages" in *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*, Osaka, 145-152.

LINGAD, John, Sarvnaz KARIMI and Jie YIN, 2013: "Location extraction from disaster-related microblogs" in *Proceedings of the 22nd International Conference on World Wide Web*, 1017-1020.

LIU, Fei, Maria VASARDANI and Timothy BALDWIN, 2014: "Automatic identification of locative expressions from social media text" in *Proceedings of the 4th International Workshop on Location and the Web*, Shangai, 9-16.

LIU, Xiaohua, Shaodian ZHANG, Furu WEI and Ming ZHOU, 2011: "Recognizing named entities in tweets" in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, Portland, 359-367.

MALMASI, Shervin, and Mark DRAS, 2016: "Location mention detection in tweets and microblogs" in Kôiti HASIDA and Ayu PURWARIANTI (eds.): *Communications in Computer and Information Science* (vol. 593), Singapore: Springer, 123-134.

MARTÍNEZ-ROJAS, María, María del Carmen PARDO-FERREIRA and Juan Carlos RUBIO-ROMERO, 2018: "Twitter as a tool for the management and analysis of emergency situations: A systematic literature review", *International Journal of Information Management* 43, 196-208.

MIDDLETON, Stuart E., Giorgos KORDOPATIS-ZILOS, Symeon PAPADOPOULOS and Yiannis KOMPATSIARIS, 2018: "Location extraction from social media", *ACM Transactions on Information Systems* 36 (4), 1-27.

MIDDLETON, Stuart E., Lee MIDDLETON and Stefano MODAFFERI, 2014: "Real-time crisis mapping of natural disasters using social media", *IEEE Intelligent Systems* 29 (2), 9-17.

MILLER, George A., 1995: "WordNet: A lexical database for English", *Communications of the ACM* 38 (11), 39-41.

MIYAZAKI, Taro, Trevor COHN and Timothy BALDWIN, 2018: "Twitter geolocation using knowledge-based methods" in *Proceedings of the 2018 EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text*, 7-16.

MURTHY, Dhiraj, 2018: *Twitter: Social Communication in the Twitter Age* (2nd ed.), Cambridge: Polity Press.

PERIÁN-PASCUAL, Carlos, and Francisco ARCAS-TÚNEZ, 2017: "A knowledge-based approach to social sensors for environmentally-related problems" in Cesar ANALIDE and Pankoo KIM (eds.): *Intelligent Environments 2017*, Amsterdam: IOS Press, 49-58.

PERIÁN-PASCUAL, Carlos, and Francisco ARCAS-TÚNEZ, 2018: "The analysis of tweets to detect natural hazards" in Ioannis CHATZIGIANNAKIS, Yoshito TOBE, Paulo NOVAIS and Oliver AMFT (eds.): *Intelligent Environments 2018*, Amsterdam: IOS Press, 87-96.

PERIÁN-PASCUAL, Carlos, and Francisco ARCAS-TÚNEZ, 2019: "Detecting environmentally-related problems on Twitter", *Biosystems Engineering* 177, 31-48.

PRIEDHORSKY, Reid, Aron CULOTTA and Sara Y. DEL VALLE, 2014: "Inferring the origin locations of tweets with quantitative confidence" in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, 1523-1536.

PURVES, Ross S., Paul CLOUGH, Christopher B. JONES, Mark H. HALL and Vanessa MURDOCK, 2018: "Geographic information retrieval: Progress and challenges in spatial search of text", *Foundations and Trends in Information Retrieval* 12 (2-3), 164-318.

RITTER, Alan, Sam CLARK, Mausam ETZIONI and Oren ETZIONI, 2011: "Named entity recognition in tweets: An experimental study" in *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, 1524-1534.

SAKAKI, Takeshi, Makoto OKAZAKI and Yutaka MATSUO, 2010: "Earthquake shakes Twitter users: Real-time event detection by social sensors" in *Proceedings of the 19th International Conference on World Wide Web*, 851-860.

SARKAR, Kamal, 2015: "A hidden Markov model based system for entity extraction from social media english text" in *CEUR Workshop Proceedings* (vol. 1587), 89-95.

VIEWEG, Sarah, Amanda L. HUGHES, Kate STARBIRD and Leysia PALEN, 2010: "Microblogging during two natural hazards events: What Twitter may contribute to situational awareness" in *Proceedings of the 28th International SIGCHI Conference on Human Factors in Computing Systems*, 1079-1088.

WALLGRÜN, Jan Oliver, Morteza KARIMZADEH, Alan M. MACEachREN and Scott PEZANOWSKI, 2018: "Geo-Corpora: Building a corpus to test and train microblog geoparsers", *International Journal of Geographical Information Science* 32 (1), 1-29.

YIN, Jie, Sarvnaz KARIMI and John LINGAD, 2014: "Pinpointing locational focus in microblogs" in *Proceedings of the 2014 Australasian Document Computing Symposium*, 66-72.

ZHANG, Cheng, Chao FAN, Wenlin YAO, Xia HU and Ali MOSTAFAVI, 2019: "Social media for intelligent public information and warning in disasters: An interdisciplinary review", *International Journal of Information Management* 49, 190-207.