

## Location reference recognition from texts: A survey and comparison

**XUKE HU\***, Institute of Data Science, German Aerospace Center (DLR), Germany

**ZHIYONG ZHOU<sup>†</sup>**, Department of Geography, University of Zurich, Switzerland

**HAO LI**, Department of Geography, University of Heidelberg, Germany

**YINGJIE HU**, Department of Geography, University at Buffalo, USA

**FUQIANG GU**, College of Computer Science, Chongqing University, China

**JENS KERSTEN**, Institute of Data Science, German Aerospace Center (DLR), Germany

**HONGCHAO FAN**, Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Norway

**FRIEDERIKE KLAN**, Institute of Data Science, German Aerospace Center (DLR), Germany

A vast amount of location information exists in unstructured texts, such as social media posts, news stories, scientific articles, web pages, travel blogs, and historical archives. Geoparsing refers to the process of recognizing location references from texts and identifying their geospatial representations. While geoparsing can benefit many domains, a summary of the specific applications is still missing. Further, there lacks a comprehensive review and comparison of existing approaches for location reference recognition, which is the first and a core step of geoparsing. To fill these research gaps, this review first summarizes seven typical application domains of geoparsing: geographic information retrieval, disaster management, disease surveillance, traffic management, spatial humanities, tourism management, and crime management. We then review existing approaches for location reference recognition by categorizing these approaches into four groups based on their underlying functional principle: rule-based, gazetteer matching-based, statistical learning-based, and hybrid approaches. Next, we thoroughly evaluate the correctness and computational efficiency of the 27 most widely used approaches for location reference recognition based on 26 public datasets with different types of texts (e.g., social media posts and news stories) containing 39,736 location references across the world. Results from this thorough evaluation can help inform future methodological developments for location reference recognition, and can help guide the selection of proper approaches based on application needs.

CCS Concepts: • Computing methodologies → Information extraction; • Information systems → Retrieval effectiveness; Geographic information systems.

Additional Key Words and Phrases: geoparsing, location reference recognition, machine learning, comparative review

---

\*Corresponding author

<sup>†</sup>Corresponding author

---

Authors' addresses: **Xuke Hu**, xuke.hu@dlr.de, Institute of Data Science, German Aerospace Center (DLR), Jena, Germany; **Zhiyong Zhou**, zhiyong.zhou@geo.uzh.ch, Department of Geography, University of Zurich, Zurich, Switzerland; **Hao Li**, hao.li@uni-heidelberg.de, Department of Geography, University of Heidelberg, Heidelberg, Germany; **Yingjie Hu**, yhu42@buffalo.edu, Department of Geography, University at Buffalo, Buffalo, USA; **Fuqiang Gu**, gufq@cqu.edu.cn, College of Computer Science, Chongqing University, Chongqing, China; **Jens Kersten**, Jens.Kersten@dlr.de, Institute of Data Science, German Aerospace Center (DLR), Jena, Germany; **Hongchao Fan**, hongchao.fan@ntnu.no, Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway; **Friederike Klan**, Friederike.Klan@dlr.de, Institute of Data Science, German Aerospace Center (DLR), Jena, Germany.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2022 Association for Computing Machinery.

Manuscript submitted to ACM

**ACM Reference Format:**

Xuke Hu, Zhiyong Zhou, Hao Li, Yingjie Hu, Fuqiang Gu, Jens Kersten, Hongchao Fan, and Friederike Klan. 2022. Location reference recognition from texts: A survey and comparison. 1, 1 (July 2022), 35 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

## 1 INTRODUCTION

"Location matters, and not just for real estate" [177]. With the rapid development of the Global Navigation Satellite System (GNSS), sensor-rich (e.g., inertial sensors, Wi-Fi module, and cameras) smart devices, and ubiquitous communication infrastructure (e.g., cellular and 4G networks and Wi-Fi access points), our capability of obtaining location information of moving objects and events in both indoor and outdoor spaces has been dramatically improved [159]. This increased our ability to better understand geo-spatial processes and to support decision making in all contexts from business, entertainment, to crisis management [177]. Apart from sensor equipment, natural language texts (e.g., social media posts, web pages, and news stories) are another important source that contains much geospatial information in the form of *location references*. These *location references* embedded in texts can be in the form of simple *place names* (or *toponyms*), and can also be in the form of *location descriptions* that contain both place names and additional spatial modifiers (e.g., direction, distance, and spatial relationship) [175]. *Geoparsing* refers to the process of recognizing location references from texts and identifying their geospatial representations. *Geoparsing* is an ongoing research problem that has been studied over the past two decades [9, 15, 80, 90, 162]. It consists of two steps: (1) toponym recognition, which is also called location reference recognition (2) toponym resolution, which is also called geocoding that disambiguates toponyms and identifies their geographic coordinates. Figure 1 illustrates the workflow of geoparsing.

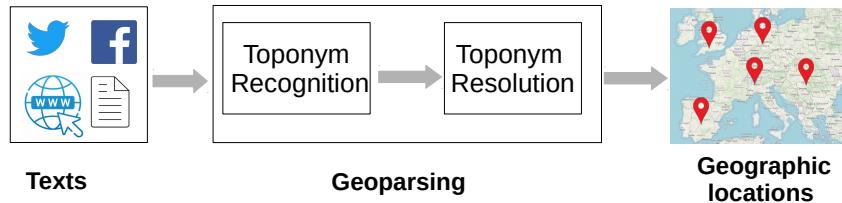


Fig. 1. The general workflow of geoparsing and its two steps.

Geoparsing has traditionally been used in formal texts for location extraction, such as web pages, news, scientific articles, travel blogs, and historical archives [15, 177]. However, the drastically increased importance of social media data (SMD) in various domains such as social science, policy making, and humanitarian relief [18, 36, 75, 170] has facilitated efforts to extend geoparsing to informal texts [177]. According to Statista<sup>1</sup>, the number of worldwide social network users will reach 4.4 billion by 2025. On average, 500 million tweets<sup>2</sup> and 4.75 billion Facebook posts<sup>3</sup> are sent each day. Formal texts normally do not have location-related metadata, while informal texts, such as tweets, can be geotagged, i.e., a Twitter user can select a location and attach that location to the posted message. However, geotagged tweets are rare, and according to Cheng et al. [31], Morstatter et al. [134], and Kumar et al. [100], only 0.42%, 3.17%, and

<sup>1</sup><https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users/>

<sup>2</sup><https://www.dsayce.com/social-media/tweets-day/>

<sup>3</sup><https://blog.wishpond.com/post/115675435109/40-up-to-date-facebook-facts-and-stats>

7.90% of the total number of tweets contain geotags, respectively. In addition, Twitter removed their precise geotagging feature in June 2019, showing only a rough location, e.g., the bounding box of a tagged place rather than a pair of latitude and longitude coordinates. This change could lead to a further decrease in the number of geotagged tweets [85]. The geotagged locations of tweets are not always the same as the locations described in their tweet content either [110]. In a nutshell, it is often necessary to extract location references from unstructured texts. Notably, informal texts, such as tweets, are short, have few or no formatting or grammatical requirements, and can have uncommon abbreviations, slang, and misspellings, which pose additional challenges for geoparsing [178].

While there exist quite some studies on geoparsing [67, 145], we identify two gaps in the literature that motivate this current review paper. First, the many possible applications of geoparsing are scattered in individual papers [1, 15, 57, 62] or are only partially reviewed [66, 82], and there lacks a systematic and more comprehensive summary of these applications. Consequently, it is difficult for researchers who are new to geoparsing to have a quick view of these many possible applications. Second, existing review papers on geoparsing, such as [67, 124, 133, 179], focused on the entire workflow of geoparsing (i.e., both of the two steps) rather than location reference recognition alone (i.e., the first step only). While providing more comprehensive coverage on the topic of geoparsing, existing efforts reviewed only some approaches for the step of location reference recognition. In recent years, many new approaches for location reference recognition have been developed, such as Flair NER [4], NeuroTPR [180], nLORE [51], and GazPNE2 [81]. Given the high importance of location reference recognition in geoparsing (i.e., only those references that are correctly recognized can be geo-located), it is necessary to have a review that specifically focuses on the possible and recent approaches for location reference recognition.

This work aims at filling the two research gaps discussed above. First, we summarize seven typical application domains of geoparsing, which are geographical information retrieval (GIR) [56, 145], disaster management [110, 161], disease surveillance [63, 158, 171], traffic management [76, 114, 128], spatial humanities [62, 153, 171], tourism management [25, 34, 35], and crime management [17, 40, 176]. Second, we review existing approaches for location reference recognition by categorizing the approaches into four groups: rule-based, gazetteer matching-based, statistical learning-based, and hybrid approaches. Noticing that many existing approaches were not cross compared on the same datasets, we also conduct experiments to compare and evaluate the reviewed 27 existing approaches on 26 public datasets. We examine multiple characteristics of the existing approaches, including their performance on formal and informal texts, their performance on different types of locations (e.g., admin units and traffic ways), and their computational efficiency.

The remainder of this paper is structured as follows: In Section 2, we summarize seven typical application domains of geoparsing. In Section 3, we review existing approaches for location reference recognition. We evaluate existing approaches on the same public datasets in Section 4. Finally, we conclude the paper in Section 5 and discuss some potential future directions.

## 2 SEVEN APPLICATION DOMAINS OF GEOPARSING

Geoparsing has many possible applications. In this section, we summarize seven typical application domains of geoparsing, which are most discussed in literatures. Figure 2 provides an illustration of these domains.

**GIR:** One of the primary applications of geoparsing is geographic information retrieval. Historically, documents have been indexed by subject, author, title, and document type. However, a diverse and large group of information system users (e.g., readers, nature resources managers, scientists, historians, journalists, and tourists) desire geographically-oriented access to document collections, such as by retrieving interesting contents about specific geographic locations [24, 56, 107, 129, 144, 172, 189]. For instance, resources in digital libraries can be indexed by locations contained in

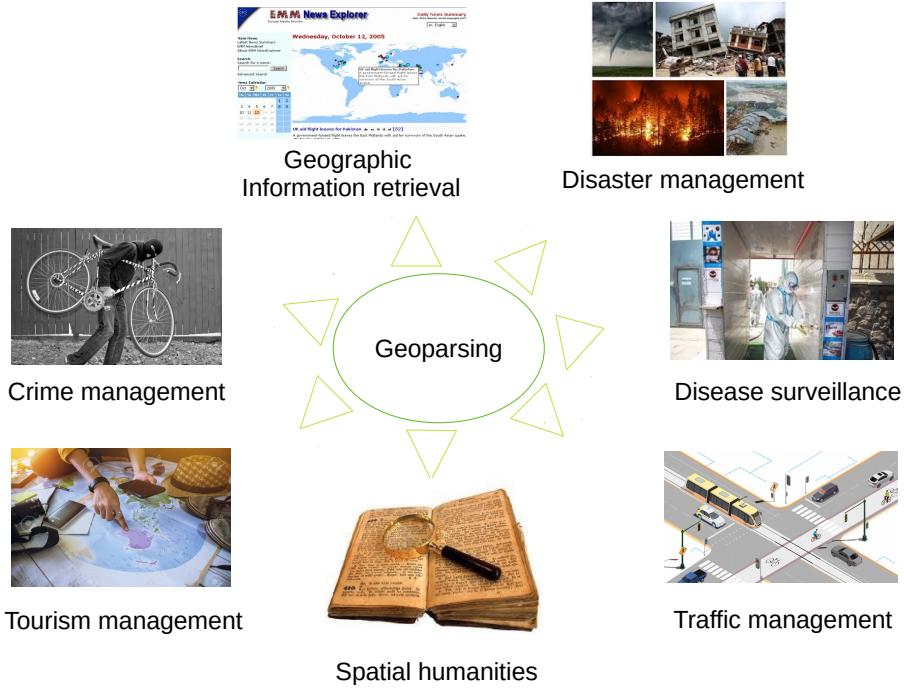


Fig. 2. Seven application domains of geoparsing.

descriptive metadata records associated with the resources, thereby improving users' experience in searching for their needed resources [56]. People are looking for web pages containing useful information about everyday tasks, such as local merchants, services, and news [24]. The public can consume up-to-date information related to COVID-19 (e.g., disease prevention, disease transmission, and death reports) on Twitter by locations [129].

**Disaster management:** News stories and SMD contain a large volume of historical and real-time disaster information. Location-enabled SMD can be very helpful to timely map the situational information, such as rescue requests [163, 196], resource (e.g., food, clothing, water, medical treatment, and shelter) needs and availability [21, 48], and facility status (e.g., building collapse, road closure, pipe broken, and power outage) [22, 50, 120, 156] in the aftermath of disasters. With a crisis map, first responders can track the unfolding situation and identify stricken locations that require prioritized intervention [19] and realize optimized real-time resource allocation [163], government agencies can conduct the damage assessment of the disasters in a faster manner [190], and the public can search for the locations where they can obtain needed resources. By extracting spatiotemporal, environmental, and other information about disaster events from news stories, flood-prone areas can be identified [192], the responsibility of atmospheric phenomena for floods can be understood [20], the spatial and temporal distributions of natural disasters during a long period can be analyzed [113], and the evolution of disasters (e.g., the phases of preparedness, impact, response, and recovery) can be tracked [86, 181, 182].

**Disease surveillance:** Scientific articles, historical archives, news reports, and social media contain detailed information of disease events, such as where the disease was first reported and how it spread spatiotemporally. Mining geographic locations and other related information of disease events can help track diseases [32, 63, 135, 139, 158, 171], Manuscript submitted to ACM

perform early warning and quick response [95], and understand the mechanisms underlying the emergence of diseases [12, 91]. For example, geoparsing historical archives (e.g., The annual US Patent Office Reports 1840-1850 and Registrar General's Reports) can help track the spread of potato disease 'late blight' in the 19th-century in the United States [171] and understand the relationship between cholera related disease and place names during Victorian times [135]. Scientific articles were geoparsed to analyze the demographic, environmental, and biological correlation of the occurrence of emerging infectious diseases at a global scale [12, 91]. Social media can also reflect the movement of the public and their feelings during pandemics through geotags or mentioned locations in texts. Location-enabled tweets were applied to analyze the mental health status of the public after the occurrence of COVID-19 [79, 195], to track and visualize the spread and diffusion of COVID-19 [16], and to reveal human mobility patterns [87, 89].

**Traffic management:** Twitter users report near-real-time information about traffic events (e.g., crashes and congestion). Detecting traffic events, their precise locations, and other related information from tweets [3, 13, 60, 70, 160, 167] is important for an effective transportation management system. The detected traffic events can also support urban policy making [38], such as to help drivers to avoid risk zones and choose the fastest and safest routes [10], to help the transportation management sector reduce fatalities and restore traffic flow as quickly as possible [10], to predict future traffic jams [11], and to improve road safety by recognizing high-risk areas [128]. By doing so, Twitter users acting as social sensors can complement existing physical transport infrastructure (e.g., video cameras and loop detectors) in a cost-effective manner, which is especially important to developing countries where resources are limited.

**Spatial humanities:** 'Spatial turn' was used to describe a general movement, observed since the end of the 1990s, emphasizing the reinsertion of place and space in humanities [183]. Digitizing and geoparsing large historical textual collections, such as books, reports, and novels create new ways for research in humanities (e.g., Archaeology, History, and Literature) [47, 61, 62, 68, 77, 130, 135, 171], such as to understand the historical geographies of nineteenth-century Britain and its relationships with the wider world [61], to identify the significance of specific commodities in relation to particular places and time [77], to analyze a correspondence between eighteenth-century aesthetic theory and the use of the terms beautiful, picturesque, sublime and majestic in contemporaneous and later accounts of the Lakes region [47], and to reveal the spatial structure of a narrative in fictional novels [130].

**Tourism management:** According to the prediction of Statista in 2016, there were going to be 32 million active bloggers in the US alone by 2020. Among all the active blogs, travel is rated as the top 5 topics shared by bloggers<sup>4</sup>. Travel blogs contain a wealth of information about visited places organized as bloggers' experiences and insights as well as their perceptions of these places [74]. These narratives reflect the blogger's behavior and interaction with places and also the relationships among the places. Geoparsing travel blogs is helpful for understanding places [73], such as to find their features and related activities, and can help describe a place with tourism attributes to support tour planning [73, 74, 97, 194]. Applications include helping travelers choose preferred places and visit them in an appropriate order at a proper time, and supporting wayfinding given the spatial relation of places [74].

**Crime management:** Many countries do not make crime data available to their citizens [17] or just provided coarse-grained details<sup>5</sup>, such as the total number of thefts in a district or a province. According to the Crime Information Need Survey [17], around 78.3% of respondents in Indonesia agreed that crime information should be available to the public. The needed information includes crime type, perpetrator, victim, time, and very importantly, location. Meanwhile, information related to crime is often scattered across news and social media. Mining and gathering crime-related information from these text-based sources can be useful for informing the public and may even help predict and prevent

<sup>4</sup><https://www.statista.com/statistics/187267/number-of-bloggers-in-usa/>

<sup>5</sup><http://liiofindia.org/>

some crimes [14, 39, 40, 149, 155, 165]. In particular, geoparsing can help extract location information of crimes, which can help residents to choose places to live and help travelers to avoid certain unsafe places [17].

Different applications have distinct requirements for the approaches for location reference recognition. For instance, informal texts (e.g., tweets) are the main source for emergency response, from which the real-time location of sub events caused by disasters can be extracted, while scientific articles are the main source for analysing the mechanisms underlying the emergence of diseases, from which the spatiotemporal evolution of diseases across the globe can be extracted. GIR just needs coarse-grained geospatial information, such as a city, while traffic management requires the fine-grained location (e.g., a street) of traffic events; Geoparsing historical documents that contain billions of words requires a fast processing workflow. Therefore, to guide the selection of proper approaches for location reference recognition based on application needs, examining the characteristics of existing approaches is necessary, which will be introduced in Section 4.

### 3 A SURVEY OF EXISTING APPROACHES

In this section, we review existing approaches for location reference recognition. In subsection 3.1, we review individual approaches by categorizing them into four groups, and in subsection 3.2, we review existing comparative studies and differentiate our current review from the existing studies.

#### 3.1 Approaches for location reference recognition

In the existing literature, Leidner and Lieberman [104], Monteiro et al. [133], Purves et al. [145] identified three types of approaches for location reference recognition, which are rule-based, gazetteer matching-based, and statistical learning-based. However, many studies, such as [57, 80, 105, 123], used a combination of different approaches to compensate the shortcomings of each other. Therefore, in this review, we add a fourth type, *hybrid approaches*, which combine two or all three types of approaches, and we use these four types to organize our review on location reference recognition. We show this classification schema in Figure 3.

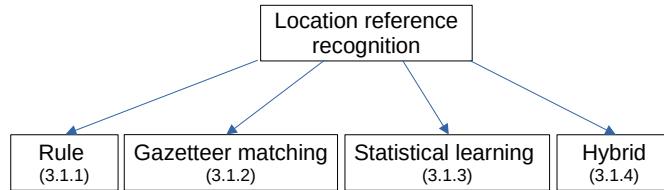


Fig. 3. Classification of existing approaches for location reference recognition.

**3.1.1 Rule-based approaches.** Location references in texts often have certain lexical, syntactic, and orthographic features. A set of rules, such as regular expressions (REs) and context-free grammars (CFGs), can be defined to decide if an n-gram of texts is a location reference or not [104]. n-grams are linear sequences of  $n$  words in texts. For example, given a text  $T = \{w_0 w_1 w_2 \dots w_n\}$ , its unigrams or n-grams of size  $n = 1$  include  $\{w_0\}, \{w_1\}, \{w_2\} \dots \{w_n\}$ . Its bigrams or n-grams of size  $n = 2$  include  $\{w_0 w_1\}, \{w_1 w_2\} \dots, \{w_{n-1} w_n\}$ .

Table 1 lists some RE and grammar rules used in previous studies [59, 60, 104, 123]. Each row in the table indicates a rule. The former twelve rules are REs, using Part-of-speech (POS) tags and/or keywords. We use the standard meta characters (i.e., '?', '+', and '\*') of REs. The '?' sign indicates the presence of a tag zero or one time, the '+' sign indicates

the presence of a tag at least one time, and the ‘\*’ sign indicates the presence of a tag any times (zero or more). Numbers indicate different types of words. 1 represents street indicators, such as ‘street’, ‘highway’, ‘road’, ‘sh’, and ‘beltway’. 2 represents words that specify direction or a distance in measurable terms, such as ‘10’, ‘away’, ‘from’, ‘miles’, ‘km’, ‘south’, and ‘northbound’. 3 represents place category words, such as ‘city’, ‘str’, ‘avenue’, ‘rd’, and ‘village’. Used POS tags include Nouns (NN), Proper Nouns (NNP), Determiners (DT), Adjectives (JJ), Cardinal Numbers (CD), and Conjunctions (CC). The last two rules are grammar rules. X denotes candidate n-grams. 4 represents place category words that are often used with ‘of’, such as ‘city’, ‘town’, ‘gulf’, and ‘river’. 5 represents spatial prepositions that normally appear before a location, such as ‘in’, ‘around’, ‘on’, ‘near’, and ‘between’.

Table 1. Examples of RE and grammar rules for location reference recognition.

Regular Expression	Examples
<NN>+	tiburon blvd; san manteo
<NNP>+	Heidelberg; San Francisco
<DT>?<JJ>?<NN>+	the golden gate bridge; long island
<CD>?<DT>?<JJ>?<NN>+	third street; 11th avenue
<DT>?<JJ>?<NN>+<CD>?	freeway 91; highway 12
<DT>?<JJ><NN>(1)	the high cotton lane; high star drive
(1)<CD>	beltway 10; sh 73
(2)+<NNP>+(3)*	south Northumbria bridge road; northeast Munich
<NNP>+(3)*(2)+	Camanche Avenue east; Heidelberg North
(2)+(3)*(of)?<NNP>+	25 miles sw of San Francisco; 25min away from New York State
(3)*(of)?<NNP>+(2)+	town of San Francisco; district of Columbia
<A-Z><a-z>*berg	Heidelberg; Freiberg
(4) (of) X -><LOC>	city of beaumont; Gulf of Mexico;
(5) X -><LOC>	this overturned tanker in marin has created a huge jam on <b>wb 580</b>

There are several studies that used only rules to extract location references. For instance, Giridhar et al. [60] used road-traffic-related tweets to detect and locate point events, such as car accidents. Specifically, a set of REs were defined according to the composition of nouns, determiners, adjectives, cardinal numbers, conjunctions, and possessive endings. Furthermore, to decrease false positives, grammar-based rules were implemented based on spatial prepositions, such as *in*, *at*, *between*, and *near*. Zou et al. [196] analyzed the rescue request on Twitter during Hurricane Harvey. The authors first manually annotated the tweets that were asking for help. Then, they used a rule-based method to recognize location references from tweets. More specifically, they assumed that the formal description of an address in the United States is in the form of [Street Number, Street Name, Apartment Number (optional), City, State, Zip Code]. Since all rescue request tweets in their study contain zip codes, the full address in each tweet can be extracted by locating the zip code as the ending point and searching for the starting point based on several conditional criteria.

Although many studies classify rule-based approaches as one category [6, 104, 133], pure rules-based approaches are rare. All the rule-based approaches discussed in [133] are, in fact, hybrid approaches. This is likely because the approaches that rely on linguistic patterns only are ineffective [162]. It remains a challenge to define rules in a complete and robust manner that can explain all possible occurrences of location references in texts, especially in microblogs with dramatic variation of writing styles and weak grammar [151]. However, a set of simple rules can be used to enhance gazetteer matching and statistical learning-based approaches, which will be introduced in the following subsections.

**3.1.2 Gazetteer matching-based approaches.** A gazetteer is a dictionary of place names associated with geospatial information (e.g., place types and geographic coordinates) and some additional information such as population size, administrative level, and alternative names. Gazetteers play important roles in location reference recognition in many

studies. GeoNames<sup>6</sup> is a most widely used gazetteer, and OpenStreetMap (OSM)<sup>7</sup>, in a broad sense, can be considered as a gazetteer as well. There are 12,255,028<sup>8</sup> and 23,876,956<sup>9</sup> places in GeoNames and OSM, respectively. Figure 4 illustrates the point density map of the places in OSM and GeoNames.

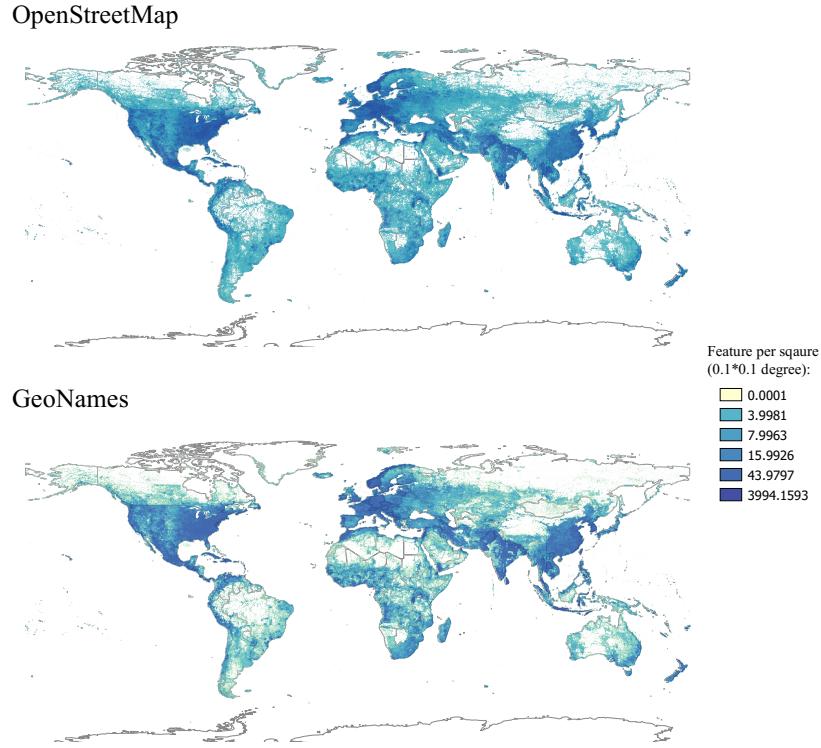


Fig. 4. Point density maps of the places contained in OpenStreetMap and GeoNames.

In gazetteer matching-based approaches, the n-grams of a text are first matched against a gazetteer, which are then filtered or disambiguated with a couple of heuristics. Gazetteer matching-based approaches are still faced with two main challenges. The first is that many location references appearing in texts are missing from gazetteers due to various reasons, such as name variation (e.g., ‘South rd’ for ‘South road’ and ‘Frankfurt airport’ for ‘Frankfurt international airport’) and data incompleteness (e.g., the missing of ‘Hidden Valley Church of Christ’ from a gazetteer) [57]. Second, gazetteer matching-based approaches often run into ambiguity issues. For instance, the names ‘Washington’, ‘MO’, ‘South Wind’, and ‘1 ft’ all exist in gazetteers, but can also refer to other types of entities. This is called geo/non-geo ambiguities, while geo/geo ambiguities refer to the situation that different spatial locations use the same name, such as Manchester, NH, USA versus Manchester, UK. For simplicity, we use *ambiguities* and *ambiguous* to refer to geo/non-geo

<sup>6</sup><http://www.geonames.org/>

<sup>7</sup><https://www.openstreetmap.org/>

<sup>8</sup>Retrieved from the official web of GeoNames on 2022.02.25

<sup>9</sup>Retrieved from OSMNames on 2022.02.25

ambiguities by default. We will use the full name *geo/geo ambiguities* to refer to the second situation. The main focus of gazetteer-based approaches is often to overcome the two mentioned challenges by using heuristics to perform disambiguation (to increase precision) and by including place name variants to expand the used gazetteer (to increase recall).

Table 2 summarizes the commonly used heuristics for disambiguation. The first four heuristics are used to reduce the number of candidate place names matched in gazetteers, thereby decreasing the number of ambiguous place names. The 5th heuristic uses common words. The 6th and 7th heuristics leverage the external and internal cues of candidate n-grams. The 8th heuristic leverages the POS tags of candidate n-grams. The 9th heuristic leverages the dictionary of other entity types (e.g., *Person*), such as to judge if a candidate n-gram ('*Houston*') with its preceding or succeeding word ('*Alexander*') in texts appears in the dictionary of person names. The 10th heuristic leverages other related place names (e.g., nearby or with the same name) to judge if an n-gram is valid or not. For example, '*IN*' is ambiguous. However, when it precedes '*Chennai*' in texts (e.g., '*Chennai, IN*'), which is likely to be a location and related to '*IN*', then '*IN*' is treated as a valid location. If an n-gram can be determined as a valid location by some heuristics, the other n-grams with the same name in the text are also treated as a valid location, such as '*stay safe Houston, flood in Houston is serious*', where both '*Houston*' are judged as valid locations since the preceding word ('*in*') of the second '*Houston*' is a spatial indicator.

Table 2. Common heuristics used for disambiguation in gazetteer matching-based approaches

ID	Heuristics	Examples
1	Limit the length of places in gazetteers	Keep only 1- and 2-grams
2	Limit the type of places in gazetteers	Keep only continent, country, state, and city
3	Limit the scale of places in gazetteers	Keep only places with a population over 1000
4	Limit the spatial range of gazetteers	Use the gazetteers in the area of Florence
5	Filter place names of common (stop) words	'today', 'long', 'that building' and 'the street'
6	Use spatial indicators in texts	'in', 'near', and 'at' that appear before a place
7	Use orthographic cues	Capitalization of words, such as 'Houston' and 'Germany'
8	Filter candidates by POS tags	Keep only noun phrases in texts
9	Use a dictionary of other entity types	Person names, such as 'Washington Irving' and 'Houston Alexander'
10	Use other related place names	'Chennai, IN' and 'stay safe Houston, flood in Houston is serious'

Many studies used gazetteer matching-based approaches to recognize location references from texts [3, 6, 12, 15, 33, 41, 52, 126, 128, 140, 140, 142, 162, 168, 169, 189]. One of the earliest geoparsing approaches was proposed by Woodruff and Plaunt [189] to support georeferenced document indexing and retrieval. A gazetteer containing around 120,000 places in California is first built on the US Geological Survey's Geographic Names Information System (USGS 1985) and the land use data from the US Geological Survey's Geographic Information Retrieval and Analysis System (GIRAS). The place names in a document were identified by matching texts' n-grams containing non-stop words against the gazetteer. If a token had no matches in the gazetteer, it was deplurIALIZED (e.g., 'valleys' to 'valley') and rematched with the gazetteer. Identified place names were then geocoded to determine the geographic scope of the document. Amitay et al. [15] developed a system, named Web-a-Where, for recognizing and geocoding continents, countries, states, and cities as well as their abbreviations in web pages. A gazetteer was created by collecting about 75,000 place names across the world from different data sources: USGS, World-gazetteer.com <sup>10</sup>, UNSD <sup>11</sup>, and ISO 3166-1 <sup>12</sup>. The system first extracted candidate place names in a given page by matching against the gazetteer. Then, four heuristics

<sup>10</sup><http://www.world-gazetteer.com>

<sup>11</sup><http://unstats.un.org/unsd>

<sup>12</sup><https://www.iso.org/iso-3166-country-codes.html>

were sequentially used to disambiguate and geocode the candidate place name, such as the vicinity of two candidate places (e.g., Chicago, IL) and the population of places. The approach was evaluated on 600 web pages, containing over 7000 place names. Clough [33] proposed geoparsing web pages. In the recognizing step, candidate place names are first identified by matching against gazetteers, which are then filtered by using stop words and by using context cues, such as to filter person names with simple heuristic  $< title >< loc >$  (e.g., Mr. Sheffield), where  $< loc >$  is a candidate place name and also in the dictionary of person names. Used gazetteers include the Ordnance Survey 1:50,000 Scale Gazetteer for the UK (OS <sup>13</sup>), Seamless Administrative Boundaries of Europe dataset (SABE <sup>14</sup>), and Getty Thesaurus of Geographic Names (TGN <sup>15</sup>). Pouliquen et al. [142] proposed geoparsing multilingual texts. Candidate place names are first identified by matching with a multilingual gazetteer, which are then disambiguated through a dictionary of person names (e.g., ‘George Bush’ and ‘Tony Blair’) and stop words (e.g., ‘And’, ‘Du’, ‘Auch’) in a multilingual context. The multilingual gazetteer is created from three sources: Global Discovery database of place names (Global Discovery 2006), the multilingual KNAB database (KNAB 2006), and a European commission internal document. The approach was tested on 162 newspaper stories in five languages (i.e., German, English, Spanish, French, and Italian) from Europe Media Monitor.

Gazetteer matching-based approaches were also used to extract locations from tweets. For instance, Paradesi [140] proposed TwitterTagger for geoparsing tweets. It matched the noun phrases of a tweet text with the entries in gazetteers (i.e., USGS database), which was followed by disambiguating the matched entry with two heuristics. The first was to check if spatial indicators (e.g., ‘in’ and ‘near’) were used before a noun phrase. The second was to check whether other users used a spatial indicator before the same noun phrase in their tweets. Geo/geo ambiguities were removed by calculating the distance between the location of a place name in a tweet and the location of the user who posted the tweet as well as the location of other tweet users who mentioned the same place name. The approach was evaluated on 2000 annotated tweets. Middleton et al. [126] proposed a multilingual geoparser for tweets, named *geoparsepy*. To overcome the place name variation issues, a set of heuristics were applied to expand OSM place names. To deal with abbreviations, a multilingual corpus of the street and building types from OSM was used to compute obvious variants for common location types (e.g., ‘Southampton Uni’ for ‘Southampton University’). To overcome the ambiguity issue, uni-gram location names that are non-nouns are filtered using a multilingual WordNet corpus lookup, such as ‘ok’ and ‘us’, which can refer to locations or other types depending on their POS tag. Location phrases are then filtered using a multilingual stop word corpus. To remove the geo/geo ambiguities, a confidence score is calculated for each matched entry in the gazetteer based on several evidential features, such as the other location references in the tweet, the admin level, and the geotag of the tweet if available. de Bruijn et al. [41] proposed a geoparsing algorithm (named TAGGS) by employing both metadata and the contextual spatial information of groups of tweets referencing the same location regarding a specific disaster type. It matches the uni- and bi-grams of a tweet text with the GeoNames gazetteer. The found candidates are then filtered with several heuristics, such as discarding the candidates with the 1000 most common words. Some studies [19, 138] leveraged gazetteer matching-based semantic annotators to realize their own geoparsing tool. For instance, Nizzoli et al. [138] employed TagMe [52] to identify entities of an input text and to link them to the corresponding entity in a knowledge graph (i.e., DBpedia). Then, they traversed the knowledge graph to expand the available information for the geoparsing task. Finally, they exploited all available information for learning a regression model that selects the best entity in the knowledge graph for annotated places in the text.

<sup>13</sup><http://www.ordnancesurvey.co.uk/oswebsite/products/50kgazetteer/>

<sup>14</sup>[http://www.eurogeographics.org/eng/03\\_projects\\_sabe.asp](http://www.eurogeographics.org/eng/03_projects_sabe.asp)

<sup>15</sup>[http://www.getty.edu/research/conducting\\_research/vocabularies/tgn/](http://www.getty.edu/research/conducting_research/vocabularies/tgn/)

Studies, such as [3, 6, 22, 128, 167, 191, 192], focused only on local events whose geographical scope is known, such as floods or traffic accidents happened in a certain city. Therefore, they would normally use a local gazetteer that contains only the places in a certain region, which can dramatically mitigate the issues of geo/non-geo ambiguities and geo/geo ambiguities. Although the proposed geoparsing approaches are not globally applicable, they are effective in dealing with local events. For instance, Al-Olimat et al. [6] proposed a Location Name Extraction tool (LNEx), which used n-gram statistics and location-related dictionaries to handle the abbreviations and automatically filter and augment the place names in the OpenStreetMap gazetteer (handling name contractions and auxiliary contents). Ahmed et al. [3] used tweets to monitor the traffic congestion in real-time. Specifically, tweets related to traffic congestion are first detected by using supervised and unsupervised machine learning techniques, and the road names in the tweets are then extracted by matching with a list of road names in the city of Chennai. Jaro-Winkler metric is used to calculate the similarity between the n-grams in tweets and the road names in the list to overcome the challenge of place name variants. 1551 congestion-related tweets were used to detect the congestion situations in Chennai in two short periods, lasting 7 months in total. Milusheva et al. [128] used traffic-related tweets to derive the locations of road traffic crash in Nairobi, Kenya, for the purpose of road safety improvements. Specifically, they applied a machine learning model to capture the occurrence of a crash and developed a gazetteer matching-based geoparsing algorithm to identify its location. A gazetteer of landmarks (e.g., roads, schools, and bus stops) for five counties that constitute the Nairobi metro area was created from OpenStreetMap, GeoNames, and Google Places. The location of crashes is then determined by matching the n-grams of the tweets with the entries in the gazetteer.

It is simple to implement gazetteer matching-based approaches and they can easily adapt to a multilingual context. Moreover, they are effective in certain applications, such as the ones whose geographic scope is limited to a small region (a city) or the ones require only coarse-grained locations, such as country names. However, it remains a challenge to propose a general and globally applicable approach for location reference recognition by using gazetteer matching and simple heuristics since the name variants and geo/non-geo ambiguity issues are ubiquitous in natural language texts. To overcome this challenge, many studies combined gazetteer matching with rules and/or statistical learning to compensating the shortcomings of each other, which will be introduced in the following subsections.

**3.1.3 Statistical learning-based approaches.** Statistical learning-based approaches are built on annotated training corpora containing texts associated with the expected location references. The annotated corpora are used to train a model via manually defined features, such as infrequent strings, length, capitalization, and contextual features, and/or features automatically learned by deep learning algorithms. The trained model is then applied to unlabeled texts, and the same features are computed to decide on the association of texts and location references. The basic architecture of statistical learning-based approaches is illustrated in Figure 5. Statistical learning-based approaches generally use either traditional machine learning models, such as Random Forest (RF) [56], or deep learning models, such as Long Short-Term Memory (LSTM) [180]. Statistical learning-based approaches can be further divided into two groups: *learning-based named entity recognition (NER)* and *learning-based place name extraction (PNE)*. In the following, we discuss these two groups of approaches respectively.

**Learning-based NER:** Location reference recognition can be considered as a subtask of NER, which has been extensively studied. Therefore, many studies [58, 70, 72, 93, 110, 120, 173] used existing statistical-based NER models or retrain them to extract location references from texts. For instance, Lingad et al. [110] used OpenNLP<sup>16</sup>, TwitterNLP [151], Yahoo!Placemaker, and Stanford NER to extract place names from 2878 disaster-related tweets. Stanford NER

<sup>16</sup><https://opennlp.apache.org/>

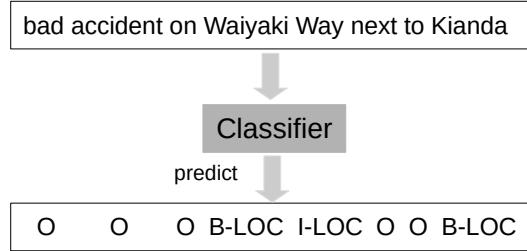


Fig. 5. Basic architecture of statistical learning-based approaches. *O* denotes non-type. *B-LOC* and *I-LOC* denote the beginning and inner part of a location reference, respectively.

and OpenNLP were also retrained and evaluated by using 10-fold cross-validation in their study. The results show retrained models achieved a much higher F1 score than pretrained models. Karimzadeh et al. [94] proposed a geoparsing system for tweets, named GeoTxt. It combined six publicly available NERs for location reference recognition, which are Stanford NER, Illinois CogComp [150], GATE ANNIE [23], MIT IE<sup>17</sup>, Apache OpenNLP, and LingPipe<sup>18</sup>. Belcastro et al. [22] utilized tweets to discover sub-events that occurred after a disaster, such as collapsed buildings, broken gas pipes, and flooded roads. Geoparsing tweets is one of the key tasks. Specifically, CoreNLP [119] is adopted to recognize street and district names, which are then geocoded by matching with a local gazetteer, covering the disaster area. Fan et al. [50] proposed uncovering the unfolding of disaster events based on tweets. Situational information (e.g., affected individuals, injured people, recuse, and damage) were extracted by using Stanford NER, which were then filtered and geocoded by keeping only the matched places in Google Geocoding API and by excluding the places outside affected areas. To understand the geographic origin and patterns of spread of the potato disease ‘late blight’ for the 19th-century US and European outbreaks and the means by which it spread, Tateosian et al. [171] used CLAVIN<sup>19</sup> to geoparse two historical collections: US Patent Office Reports 1841-1850 and Google Books Corpus. CLAVIN is an open-sourced geoparser that utilizes Apache OpenNLP for place name extraction. Mircea [129] implemented a prototype dashboard for real-time classification, geolocation, and interactive visualization of COVID-19 tweets. spaCy<sup>20</sup> was used to extract cities and countries from tweet content and user profiles. Mao et al. [120] proposed mapping near-real-time power outages from tweets using a retrained NeuroNER model [44]. Suat-Rojas et al. [167] utilized a retrained spaCy NER to detect and analyze traffic accidents from Spanish tweets in a city of Colombia.

Recently, many deep learning-based NERs have also been proposed. For example, Limsopatham and Collier [109] proposed recognizing name entities from tweets by enabling BiLSTM to automatically learn orthographic features using both the character embedding and word embedding. Akbik et al. [5] proposed Flair, an NLP tool that used contextual string embeddings for sequence labeling tasks, such as part-of-speech (POS) tagging and NER. Qi et al. [147] proposed a deep learning-based NLP toolkit, named Stanza, which adopted a contextualized string representation-based tagger. Recently, the fully-connected self-attention architecture (a.k.a. Transformer) attracts a lot of attention due to its parallelism and advantage in modeling long-range contexts. For instance, Ushio and Camacho-Collados [174] presented a python library for NER model fine-tuning, named T-NER. It facilities the training and testing of a Transformer-based

<sup>17</sup><https://github.com/mit-nlp/MITIE>

<sup>18</sup><http://alias-i.com/lingpipe/demos/tutorial/ne/read-me.html>

<sup>19</sup><https://github.com/Novetta/CLAVIN>

<sup>20</sup><https://spacy.io/>

NER model. Nine public NER datasets from different domains are compiled as part of the T-NER library, such as CoNLL 2003, OntoNotes 5.0, and WNUT 2017 datasets.

**Learning-based PNE:** Apart from using or retraining existing NER models, many studies also trained their own models for location reference recognition by using machine learning [137, 154, 164] and deep learning models [9, 27, 30, 99, 116, 122, 174, 191]. For instance, Nissim et al. [137] trained the Curran and Clark (C&C) maximum entropy tagger [37] for recognizing location references from Scottish historical documents, using the built-in C&C features, including morphological and orthographical features, information about the word itself, POS tags, named entity tag history (with a window size of 2), and contextual features. The model was evaluated on 648 Scottish historical documents, containing 10,868 sentences and 5682 places. Kumar and Singh [99] implemented a multi-channel convolutional neural network (CNN) architecture to extract location references from tweets. The model was evaluated on 5107 earthquake-related tweets with 6690 place names by using 10-fold cross-validation. Xu et al. [191] proposed DLocRL, a deep learning pipeline for fine-grained location recognition and linking in tweets. Specifically, they first used BiLSTM-CRF to train a Point of Interest (POI) recognizer. Then, given an input pair (POI, Profile), a linking module was trained to judge whether the location profile corresponds to the POI. The profile is an entry in a POI dictionary. The approach was evaluated on the Singaporean national Twitter dataset that was first used in [105], containing 3611 tweets and 1542 POIs. Cadorel et al. [27] proposed to extract a property’s location and neighborhood from French housing advertisements by recognizing place names and retrieving relationships between them. Specifically, a BiLSTM-CRF network with a concatenation of several text representations (CamenBERT [121], Flair, and Word2Vec [127]) was used to extract place names.

To mitigate the effort for manually annotating a large training dataset, semi-supervised approaches have been developed. For instance, Wang et al. [180] proposed to generate labeled training data from Wikipedia articles to train a BiLSTM model called NeuroTPR. Their model contains several layers to account for the linguistic irregularities in Twitter texts, such as the use of character embeddings to capture the morphological features of words, and contextual embeddings to capture the semantics of tokens in tweets. The approach was evaluated on 1000 tweets related to the 2017 Hurricane in Texas and Louisiana. Qiu et al. [148] proposed ChineseTR, a weakly supervised Chinese toponym recognizer. It first generated training examples based on word collections and associated word frequencies from various texts. Based on the training examples, a BiLSTM-CRF network built on BERT word embeddings was explored to train a toponym recognizer. The approach was evaluated on three Chinese NLP datasets (i.e., WeiboNER, Boson, and MSRA)<sup>21</sup>. Khanal and Caragea [96] used a multi-task learning setting to augment the learning of fine-grained location identification. The three tasks related to crisis events are key-phrase identification, eyewitness-account classification, and humanitarian category classification. The learning is conducted on one of the three popular transformer-based models: BERT [45], Albert [103], and RoBERTa [115]. Several public datasets for the training of the three tasks were utilized in the multi-task learning. The proposed approach was evaluated on two disaster-related twitter datasets that were used in Middleton et al. [126], which contain 1907 and 1762 tweets, respectively.

Given abundant annotated data, statistical learning-based approaches can automatically recognize location references according to the contextual cues and the intrinsic features of location references without requiring additional expert knowledge and gazetteers. However, a large number of labeled training sentences are often not available, making it difficult to use these approaches in many situations [69]. Furthermore, deep learning based models normally take much more time to recognize place names from texts than rule and gazetteer matching-based approaches.

---

<sup>21</sup><https://github.com/InsaneLife/ChineseNLPCorpus>

**3.1.4 Hybrid approaches.** Every single technique has its own drawbacks. Thus, researchers have proposed fusing different techniques to achieve the best of all [23, 49, 78, 105, 118, 186, 193]. Hybrid approaches can be further divided into four types based on the way they combine the previous three approaches: fusing rule and gazetteer, fusing rule and statistical learning, fusing gazetteer and statistical learning, and fusing rule, gazetteer, and statistical learning.

**Fusing rule and gazetteer:** Many studies [118, 122, 131, 132, 143, 182, 186] fused rules and gazetteers to overcome the shortcomings of each other. Manually defined rules are fragile and the detected location references can be thus further verified by gazetteers. Inversely, rules can be used to mitigate the two challenges faced by gazetteer matching, and can help remove the ambiguities of the location references detected by gazetteer matching and by recognizing those references that are not included in gazetteers. For instance, Pouliquen et al. [143] proposed identifying cities and countries from newspapers in multiple languages. Location references are recognized by matching texts' n-grams written in upper case with a multi-language gazetteer, named Global Discovery gazetteer. The matches are then filtered by stop words and person names. To recognize the morphological variants of places, regular expressions are used to list all possible suffixes and suffix combinations of location references. By doing so, some unseen places in gazetteers can be recognized, such as '*Lontoolaisen*', because it consists of '*Lontoo*' that is in the gazetteer and the suffix '*laisen*'. To remove geo/geo ambiguities, several heuristics are utilized, such as discarding small places in the gazetteer, leveraging the importance of places, and determining the country of an article. The approach was evaluated on 28 texts with 1650 places in 8 languages (e.g., English, Spanish, and Russian). To help understand the origins, mutations, and the geospatial transmission patterns of viruses, such as influenza, rabies, and Ebola, Weissenbacher et al. [186] presented a geoparsing system for research articles related to phylogeography. GeoNames was first searched to detect location references in articles, and then a black-list (e.g., '*How*', '*Although*', '*Gene*', and '*Body*') and a set of rules were created to remove noisy entities found in GeoNames. Malmasi and Dras [118] proposed detecting location references in tweets. A POS rule-based tree splitting method is first used to extract noun phrases, and the n-grams of the noun phrases are then matched with the entries of GeoNames. Dutt et al. [49] presented SAVITR, a system that geoparses and visualizes tweets during emergencies. They used a POS tagger to find proper nouns, and then used REs to mitigate the ambiguity of proper nouns with the prefix and suffix words (e.g., '*road*', '*south*', and '*city*') of place names. Last, the phrases extracted by the above methods are verified and geocoded by using a gazetteer (i.e., GeoNames or OSM) in India. Martínez and Pascual [122] presented LORE, a knowledge-based model that captures location references from English and Spanish tweets. First, bi-grams and uni-grams in the tweets are matched with entries in the GeoNames gazetteer and then filtered by heuristics. Second, linguistic patterns involving location-indicative words (e.g., '*city*' and '*street*'), location markers (e.g., '*north*' and '*10km*'), and POS tags are derived to recognize location expressions, such as '*25 miles NW of London City*'. They derived the linguistic patterns from 500 English tweets and 100 Spanish tweets, and then used 900 English tweets and 500 Spanish tweets to test LORE.

**Fusing rule and statistical learning:** Statistical learning models might not generalize well due to limited training samples, and manually defined rules can be added to boost the performance of the trained models, e.g., by correcting evident errors. For instance, Acheson and Purves [2] proposed geoparsing scientific articles in the form of PDF by first recognizing candidate location references using Stanford NER and then filter the candidates using rules, such as to include candidates with '*University*' or '*Institute*' and reject candidates with '*Inc*' and '*GmbH*'. These rules are derived based on observations from the training sets. Google Geocoding API was then used to obtain the spatial representation of the detected location references. The approach was evaluated on two article corpora in the domain of Orchards and Cancer, containing 150 and 200 articles, respectively. Das and Purves [38] proposed detecting traffic events (e.g., traffic accidents and congestion) in India by using tweets. First, tweets are classified as traffic relevant or not relevant using a

supervised model. A hybrid method is then used to recognize location references in traffic-relevant tweets. Finally, the location references are geocoded by using OSM Nomination API<sup>22</sup>. Specifically, the approach combines the detected location references by Stanford NER, retrained OpenNLP, and a rule-based system involving spatial indicators (e.g., ‘in’, ‘at’, and ‘near’), POS tags, and 85 words of place categories (e.g., ‘hospital’, ‘road’, and ‘clinic’).

**Fusing gazetteer and statistical learning:** In this type of hybrid approaches, gazetteers are used in two main ways: (1) to combine the detection result of statistical learning models with gazetteer matching; (2) to use gazetteer matching result (e.g., if an n-gram is in gazetteers or not) as input features for statistical learning models. Examples of the first way are [56, 71, 78, 105]. For instance, to improve users’ experience in searching their needed resources from digital libraries, Freire et al. [56] proposed geoparsing descriptive metadata records associated with digital resources. Initial location references are recognized by matching tokens of records with candidate entries in GeoNames. A Random Forest classifier was then trained to disambiguate and link the initial location references to the final entry. Li and Sun [105] proposed recognizing POIs in tweets. Candidate POIs in tweets were first extracted by matching with a POI inventory, which was constructed from check-in data in Foursquare. A trained time-aware POI tagger based on CRF was then utilized to remove the ambiguity of the candidates based on the context cues in the text. Hoang and Mothe [78] combined the detection results of multiple publicly available tools, such as Ritter’s tool [151], Gate NLP framework [23], and Stanford NER, and then filtered the results using DBpedia. Different configurations of the NER tools and DBpedia were tested on the Ritter’s dataset [151] and MSM2013 dataset [28]. Examples of the second way include [51, 88, 117, 141, 185]. For instance, Inkpen et al. [88] trained three CRF models for recognizing city, province/state, and country mentions based on manually defined features, including gazetteer features (if a phrase is in GeoNames or not). The models are intended to not only detect location references in tweets but also to categorize them into three types. The models were evaluated by using 10-fold cross-validation on 6000 tweets, containing 1270 countries, 772 states or provinces, and 2327 cities. To support viral phylogeographic studies, Weissenbacher et al. [185] proposed recognizing location references in research articles pertaining to virus-related GenBank records by using a CRF model. Lexical (i.e., POS tags), semantic, and gazetteer features. The proposed approach was evaluated on the same dataset as Weissenbacher et al. [186]. Fernández-Martínez and Periñán Pascual [51] proposed nLORE, a BiLSTM-CRF architecture for location reference recognition, exploiting both linguistic and gazetteer features from LORE [123]. The model was trained on 7000 tweets and then tested on 1063 tweets.

**Fusing rule, gazetteer, and statistical learning:** Some studies combined all the three techniques for location reference recognition [48, 59, 80, 81, 106, 117]. For instance, Gelernter and Zhang [59] proposed a cross-lingual location reference recognition for tweets, combining the results of a named location parser by using gazetteer matching, a rule-based building parser, a rule-based street parser, and a trained CRF-based named entity parser. The rules of the street and building parsers are created based on POS tags and indicator words, such as adjective plus noun and street and building indicators (e.g., ‘street’ and ‘highway’ in English and ‘calle’ and ‘carretera’ in Spanish). 4488 Spanish crisis-related tweets with 3182 tweets as training set and the rest as test set are used to evaluate the Spanish extractor. The Spanish dataset is then translated into English with the Google translator and used to evaluate the English extractor. Magge et al. [117] used a deep feedforward neural network to determine whether a given phrase in biomedical articles is a toponym or not. Rules are used to generate 8 million training samples from unannotated datasets. The generated samples are then used together with manually annotated training samples to train a model. The phrase’s context (represented by word embeddings), properties of the phrase (e.g., if it is in GeoNames), properties of the document (i.e.,

<sup>22</sup><https://github.com/osm-search/Nominatim>

abstract, introduction, body, or table) are concatenated into the input vector of the deep learning model. Dutt et al. [48] proposed understanding five important aspects of need-tweets and availability-tweets during disasters, including what resource (e.g., water, food, shelter, and medicines) and what quantity is needed/available, the geographical location of the need/availability, and who needs/is providing the resource. With regard to geoparsing, the authors improved their previously proposed system, Savitr [49], by combining the location references detected by spaCy and a rule system and then filtering the location references through a gazetteer. More recently, we proposed two place name extractors for tweets. The first extractor is named GazPNE [80], which is a neural classifier first trained based on place names in OpenStreetMap in the region of the US and India and non-place names that are synthesized by rules. Because GazPNE still suffers from ambiguity issues due to its limited use of context information, we developed a second and more robust approach, GazPNE2 [81]. It utilizes two pretrained transformer models, BERT and BERTweet [136] to disambiguate the detected location references and achieved an improved F1 score of 0.8 on 19 public twitter datasets.

### 3.2 Comparative studies

In addition to individual studies that focused on developing new methods, researchers also conducted experiments to compare existing methods based on the same datasets. Liu et al. [112] created a medium-scale corpus of locative expressions from multiple social media sources which include the TellUsWhere corpus [187], two sets of micro-blog posts from Twitter, comments from YouTube, forums, blog posts from tier one of the ICWSM-2011 Spinn3r dataset <sup>23</sup>, Wikipedia, and documents from the British National Corpus [26]. They then compared the performance of a couple of location reference recognition models over these seven corpora, which include Locative Expression Recogniser (LER) [111], retrained Stanford NER, pretrained Stanford NER, GeoLocator [57], UnLockText, and Twitter NLP. The results show the pretrained Stanford NER achieves the best overall performance. Gritta et al. [67] evaluated the performance of five geoparsers (GeoTxt, Edinburgh Geoparser [68], Yahoo! PlaceSpotter, CLAVIN, and Topocluster [42]) on two datasets, Local-Global Lexicon (LGL) [108], and WikToR that was programmatically created by the author. For location reference recognition, GeoTxt uses Stanford NER, Edinburgh Geoparser uses LT-TTT2, TopoCluster uses Stanford NER, and CLAVIN uses Apache OpenNLP. The evaluation results showed that Stanford NER performs the best in location reference recognition, and Edinburgh Geoparser and CLAVIN perform the best in geocoding. Wang and Hu [179] developed an extensible and unified platform for evaluating geoparsers, named EUPEG, which enables direct comparison of nine geoparsers on eight public corpora, which are LGL, GeoVirus [64], TR-News [93], GeoWebNews [65], WikToR [67], GeoCorpora [177], Hu2014 [83], and Ju2016 [92]. The compared geoparsers include GeoTxt, Edinburgh Geoparser, TopoCluster, CLAVIN , Yahoo! PlaceSpotter, CamCoder [64] that uses spaCy NER for location reference recognition, DBpedia Spotlight [125], and two systems that use Stanford NER and spaCy NER for location reference recognition, respectively, and population-based heuristics for disambiguation in geocoding step. Won et al. [188] evaluated the performance of five NERs and voting systems that combine the NERs in extracting place names from two collections of historical correspondence, named Mary Hamilton Papers and the Samuel Hartlib collection. The NERs include NER-Tagger [101], Stanford NER, spaCy, Edinburgh Geoparser, and Polyglot-NER [7]. The results showed that although the individual performance of each NER system was corpus dependent, the ensemble combination can achieve consistent measures of precision and recall, outperforming the individual NER systems. At the International Workshop on Semantic Evaluation 2019 <sup>24</sup>, a task for toponym resolution in scientific articles was launched. The evaluation results were presented in [184]. Several systems were evaluated on a corpus of 150 full PubMed articles as 105 articles for

<sup>23</sup><https://www.icwsm.org/2011/data.php>

<sup>24</sup><https://alt.qcri.org/semeval2019/>

Table 3. Main features of tools evaluated in this study

Tool and Version	Category	Recognized Entity Type	Target Data	Development Language	Publishing Time
Stanford NER 4.3.1	statistical learning	4 classes	formal texts	Java	2021-11
spaCy 3.2.1	statistical learning	18 classes	formal texts	Python	2021-12
Stanza 1.2	statistical learning	18 classes	formal texts	Python	2021-01
OpenNLP 1.9.4	statistical learning	4 classes	formal texts	Java	2021-11
DBpedia Spotlight	statistical learning	*	formal texts	Python	2021-02
NER-Tagger	statistical learning	4 classes	tweets	Python	2016-04
Polyglot 16.07.04	statistical learning	3 classes	formal texts	Python	2016-07
NeuroNER	statistical learning	4 classes	formal texts	Python	2017-05
CogComp 4.0	statistical learning	4 classes	formal texts	Java	2018-08
OSU TwitterNLP	hybrid	10 classes	tweets	Java	2011-07
TwitIE-Gate 9.0.1	hybrid	4 classes	tweets	Java	2013-09
TNER	statistical learning	28 classes	formal texts	Python	2021-04
Flair NER	statistical learning	4 classes	formal texts	Python	2021-02
Flair NER (Ont)	statistical learning	18 classes	formal texts	Python	2021-03
BERT-base-NER	statistical learning	4 classes	formal texts	Python	2020-04
CLIFF 2.6.1	statistical learning	LOC	formal texts	Python	2020-04
Edinburgh 1.2	hybrid	LOC	formal texts	c	2021-07
GazPNE2	hybrid	LOC	tweets	Python	2022-02
LORE	hybrid	LOC	tweets	c++	2020-12
nLORE	hybrid	LOC	tweets	c++	2021-06
SPENS	hybrid	LOC	formal texts	N/A	2018-03
RSD	hybrid	LOC	tweets	N/A	2018-03
RGD	hybrid	LOC	tweets	N/A	2018-03
RS	hybrid	LOC	tweets	N/A	2018-03
BaseSemEval12	hybrid	LOC	formal texts	Python	2018-09
NeuroTPR	statistical learning	LOC	tweets	Python	2020-10
Geoparserpy 2.1.4	gazetteer matching	LOC	tweets	Python	2020-06

training and 45 articles for testing, containing in total 8360 toponyms. In the subtask of toponym recognition, all systems except one adopted Deep Recurrent Neural Networks. The highest F1 score was achieved by the system proposed by a team from Alibaba Group, which adopted an architecture of BiLSTM-CRF that was trained on OntoNote5.0, CoNLL13, and weakly labeled training corpora. All systems relied on handcrafted features for toponym resolution, including the lexical context of the toponyms and their importance (e.g., population), and a gradient boosting algorithm performs the best.

There are two major differences between this study and the aforementioned comparative studies. First, these existing comparative studies focused on the entire workflow of geoparsing, while we focus on a narrower topic, i.e., location reference recognition, and provide a deeper review and comparison of methods on this topic. Second, our comparative experiments (presented in the following section) are more comprehensive than existing studies. We used more datasets (26 datasets, containing 39,736 places across the world) and compared 27 different approaches. In the following, we present the results from the comparative experiments.

## 4 COMPARISON OF EXISTING APPROACHES

### 4.1 Methods

To inform future methodological developments for location reference recognition and help guide the selection of proper approaches based on application needs, we examine numerous characteristics of existing approaches for location reference recognition. We use or implement the 27 most widely used approaches including both general NERs and location-specific approaches. Note that, we do not include several approaches, such as LNEx [6], GazPNE [80], and SAVITR [49], since they can only be applied to a local region rather than the entire globe, while the place names

of our test datasets are across the globe. Table 3 summarizes the features of the compared approaches. The number after the name of a tool represents its version. The second column denotes the category of the approach with regard to its underlying principle. NERs can not only recognize *Location* but also other entity types, as shown in the third column. 3 classes, 4 classes, 10 classes, and 18 classes denote {LOC, PER, ORG}, {LOC, PER, ORG, MISC}, {PERSON, GEO-LOCATION, COMPANY, PRODUCT, FACILITY, TV-SHOW, MOVIE, SPORTSTEAM, BAND, OTHER}, and {LOC, PERSON, ORG, FAC, GPE, CARDINAL, DATE, EVENT, LANGUAGE, LAW, MONEY, NORP, ORDINAL, PERCENT, PRODUCT, QUANTITY, TIME, WORK\_OF\_ART}, respectively. 28 classes includes the entities of the 18 classes and the entities of {CELL TYPE, CELL LINE, CHEMICAL, CORPORATION, DISEASE, DNA, GROUP, PROTEIN, RNA, OTHER}. Note that, DBpedia Spotlight can recognize quite a few detailed classes, such as Place, People, Event, and Color. The fourth column refers to the type of datasets on which it was developed. The fifth column refers to the development language of the approach. *N/A* means the code of the approach is unavailable and we reimplement the approach by ourselves. The last column refers to the time that a tool of a certain version was proposed or updated.

- **Stanford NER (4.3.1)** [53] : It is a Java implementation <sup>25</sup> of a NER based on CRF, which was developed and maintained by the Stanford Natural Language Processing Group. We kept the entities of LOC (location) detected by Stanford NER as locations.
- **spaCy (3.2.1)**: It is a general NLP tool. We used its retrained model (*en\_core\_web\_lg*) and kept the entities of LOC, FAC (facility), and GPE (geopolitical entity) detected by spaCy as locations.
- **Stanza (1.2)** [147]: It is a general NLP toolkit <sup>26</sup> and includes an NER tool, which was built on BiLSTM and CRF. We kept the entities of LOC, FAC, and GPE as locations.
- **OpenNLP (1.9.4)** [125]: The Apache OpenNLP library is an open sourced and machine learning based toolkit for the processing of natural language text. We kept the entities of Location detected by OpenNLP as locations.
- **DBpedia Spotlight** [125]: It is for recognizing and linking entities based on a knowledge base—DBpedia. We treated the place mentions detected by this tool <sup>27</sup> as locations in the evaluation.
- **NER-Tagger** [101]: It is a NER tool targeted for tweets. It was built on BiLSTM and CRF. We used the pretrained model and implementation <sup>28</sup> to detect locations that were tagged with B-LOC and I-LOC.
- **Polyglot (16.07.04)** [7]: It is a natural language pipeline <sup>29</sup> and includes a multi-language NER tool. The entities tagged with I-LOC by this tool were regarded as locations.
- **NeuroNER** [44]: It is a BiLSTM-CRF based NER system developed by MIT. We used the pretrained model and implementation <sup>30</sup> to tag entities, and the locations were those with the tags of LOC, FAC, and GPE.
- **CogComp (4.0)** [150]: It is a NER tagger <sup>31</sup>, which was developed by the University of Illinois. The entities tagged with LOC were taken as the locations identified by this tool.
- **OSU Twitter NLP** [151]: It is a twitter-specific NER tool <sup>32</sup> that particularly targets twitter texts. The entities tagged with GEO-LOCATION and FACILITY by this tool were treated as locations.

<sup>25</sup><https://nlp.stanford.edu/software/CRF-NER.html>

<sup>26</sup><https://stanfordnlp.github.io/stanza/>

<sup>27</sup><https://www.dbpedia-spotlight.org/>

<sup>28</sup><https://github.com/glample/tagger>

<sup>29</sup><https://polyglot.readthedocs.io/en/latest/index.html>

<sup>30</sup><https://github.com/Franck-Dernoncourt/NeuroNER>

<sup>31</sup><https://github.com/CogComp/cogcomp-nlp/tree/master/ner>

<sup>32</sup>[https://github.com/aritter/twitter\\_nlp](https://github.com/aritter/twitter_nlp)

- **TwitIE-Gate (9.0.1)** [23]: It is a twitter-specific NER tool <sup>33</sup>, providing an executable pipeline on an open-source software toolkit GATE <sup>34</sup> (General Architecture for Text Engineering). The entities tagged with Location by this tool were treated as locations.
- **TNER** [174]: It is an All-Round Python Library <sup>35</sup> for Transformer-based Named Entity Recognition. Its recognized locations included those entities tagged with LOC, GPE, and FAC.
- **Flair NER** [4]: Flair is an NLP framework designed to facilitate training and distribution of sequence labeling and text classification. Flair-NER is the standard 4-class NER model trained on CoNLL-03. We used their trained model <sup>36</sup> directly and identified locations through the LOC tag.
- **Flair NER (Ontonotes)** [157]: This is the large 18-class NER model trained on Ontonotes that ships with Flair. It is named *Flair NER (Ont)* for short in this review. We used their trained model <sup>37</sup> directly and included entities tagged with LOC, GPE, and FAC as locations.
- **BERT-base-NER**: It is a fine-tuned BERT model that is ready to use for Named Entity Recognition. We used their trained NER model <sup>38</sup> directly. The locations were derived from the entities tagged with B-LOC and I-LOC.
- **GazPNE2** [81]: It fuses global gazetteers and two pretrained transformer models. The latest version <sup>39</sup> utilized Stanza to accelerate GazPNE2 and detect hard examples for GazPNE2.
- **CLIFF (2.6.1)** [46]: It integrates the results of Stanford NER and a modified CLAVIN (Cartographic Location and Vicinity Indexer) geoparser. In the evaluation, we used its implementation <sup>40</sup> and kept the detected place mentions as locations.
- **LORE** [123]: It is a rule-based location extractor for tweets. We used its implementation to extract locations.
- **nLORE** [122]: It is a deep learning model, an advanced version of **LORE**. We used the trained model provided by the author to extract locations.
- **Edinburgh Geoparser (1.2)** [68]: It is a geoparsing tool <sup>41</sup> developed by Edinburgh University, which combines rules and gazetteers to extract place names directly from text.
- **BaseSemEval12** [117]: It is a baseline system for SemEval-2019 Task 12 (i.e. Toponym Resolution in Scientific Papers) that uses a 2-layer feedforward neural network <sup>42</sup>. Its output place mentions were taken as the detected locations.
- **NeuroTPR** [180]: It is a neuro-net toponym recognition tool trained on recurrent neural networks. We used their trained model and implementation <sup>43</sup> to detect location mentions in texts.
- **Geoparserpy (2.1.4)** [126]: It is a representative gazetteer-based geoparser. We used the implementation of Geoparserpy and deployed the required OpenStreetMap gazetteer to extract place names.
- **SPENS** [188]: This approach combines the result of five different systems in a voting mechanism, including Stanford NER, Polyglot NER, Edinburgh Geoparser, NER-Tagger, and spaCy. It is thus named SPENS for short. We reimplement the approach by using the code and API of the five modules.

<sup>33</sup><https://gate.ac.uk/wiki/twitie.html>

<sup>34</sup><https://gate.ac.uk/>

<sup>35</sup><https://github.com/asahi417/tner>

<sup>36</sup><https://huggingface.co/flair/ner-english>

<sup>37</sup><https://huggingface.co/flair/ner-english-ontonotes-large>

<sup>38</sup><https://huggingface.co/dslim/bert-base-NER>

<sup>39</sup><https://github.com/uuhuhuy/GazPNE2>

<sup>40</sup><https://cliff.mediacloud.org/>

<sup>41</sup><http://www.ltg.ed.ac.uk/software/geoparser/>

<sup>42</sup><https://github.com/amagge/semeval-fnn-baseline>

<sup>43</sup><https://github.com/geoai-lab/NeuroTPR>

- **Ritter+Stanford NER+DBpedia** [78]: It uses DBpedia to filter the merged detection by Ritter’s tool (also named OSU Twitter NLP) and Stanford NER. We name this approach RSD for short and reimplement the approach by using the code and API of the three modules.
- **Ritter+GATE+DBpedia** [78]: It uses DBpedia to filter the merged detection by Ritter’s tool and GATE. We name this approach RGD for short and reimplement the approach by using the code and API of the three modules.
- **Ritter+Stanford NER** [78]: It merges detection by Ritter’s tool and Stanford NER. We name this approach RS for short and reimplement the approach by using the code and API of the two modules.

All methods were configured taking into account the corresponding research results to ensure to choose the optimal parameter settings. For example, we consider not only *Location* and *GPE* but also *Facility* detected by Stanza as a location since this can achieve the best F1 score on the total datasets.

## 4.2 Test data

We collect in total 26 commonly used datasets and use them as test data. The datasets include 3 formal datasets (i.e., news) and 23 informal datasets (i.e., tweets), containing 39,736 place names in total, as shown in Table 4. They can be categorized into two groups based on the purpose of the datasets: Location Extraction (LE) and NER. The former only annotates *Location* while the latter annotates not only *Location*, but also the other types, such as *Person*, *Organization*, and *Facility*. Note that, we do not use some available geoparsing datasets that were used to evaluate the geoparsing approaches in [179], such as WikToR [67] since we found they miss quite a few toponyms. For example, in WikToR, a text or article corresponds to a WikiPedia page entitled with a toponym whose coordinates are specified. The text explains the toponym. Only this toponym is automatically annotated, ignoring the other toponyms in the text. The dataset can be used to evaluate toponym resolution approaches but not toponym recognition approaches. The description of the used datasets is as follows:

- **LaFlood2016, HouFlood2015, CheFlood2015**<sup>44</sup>: They are three flood related datasets, which were created by Al-Olimat et al. [6]. The locations in the three datasets were annotated as one of the three types: *inLOC*, *outLOC*, and *ambLOC*, denoting the locations inside the area (e.g., ‘Houston’) of interest, outside the area, and ambiguous locations (e.g., ‘my house’), respectively. We only evaluate the tools on the *inLOC* and *outLOC* locations, ignoring the *ambLOC* locations. *Louisiana*, *Houston*, *Texas*, and *Chennai*, as well as their abbreviations, such as ‘La’, ‘Hou’, and ‘Tx’ appear frequently in the datasets. Moreover, many location mentions are in hashtags, such as '#laflood', '#txwx', and '#ChennaiRain'.
- **Harvey2017**<sup>45</sup>: The dataset is related to 2017 Hurricane Harvey and was created by Wang et al. [180]. The dataset contains many fine-grained locations, such as ‘398 Garden Oaks Blvd’ and ‘26206 longenbaugh rd’. No places appear in hashtags since they have been removed from the dataset.
- **NzEq2013, NyHuren2012**<sup>46</sup>: The two twitter datasets correspond to the New Zealand earthquake in 2013 and New York Hurricane in 2012, respectively. They were created by Middleton et al. [126]. We found several missing place names (e.g., ‘Christchurch’) which, however, appear frequently in the two datasets. To mitigate this issue, we manually create two missing place name lists (i.e. [(‘new’, ‘zealand’), (‘nz’), (‘uk’), (‘christchurch’), (‘chch’), (‘lyttleton’), (‘southland’), (‘wellington’), (‘south’, ‘island’)] and [(‘new’, ‘york’), (‘nyc’), (‘new’, ‘york’, ‘city’),

<sup>44</sup>The datasets can be obtained by filling out the Dataset Registration form [https://docs.google.com/forms/d/e/1FAIpQLScf6-DNwkgJXPS5e28Mj18hIW3Ap\\_Ym7Kna-SO7oSmiC72qGw/viewform](https://docs.google.com/forms/d/e/1FAIpQLScf6-DNwkgJXPS5e28Mj18hIW3Ap_Ym7Kna-SO7oSmiC72qGw/viewform)

<sup>45</sup><https://github.com/geai-lab/NeuroTPR/tree/master/Data/TestData/HarveyTweet2017>

<sup>46</sup><https://revealproject.eu/geoparse-benchmark-open-dataset/>

(‘ny’]) for the two datasets, respectively. We define that the detection of an entity which is not annotated in the dataset but in the corresponding missing list is a true positive. Moreover, sub place names exist in dataset NzEq2013. For example, in the text ‘*Christchurch hospital is now back in operation*’, both ‘*Christchurch hospital*’ and ‘*Christchurch*’ were annotated as *Location*. To tackle this issue, we removed sub place names from the dataset.

- **Martinez\_I, Martinez\_II, Martinez\_III:** The three twitter datasets correspond to multiple crises and emergency events (e.g., earthquake, flood, car accident, bombing, shooting, terrorist, and incident) that happened across the world. They were initially utilized in [51, 123]. One of the features of the datasets is that many fine-grained locations, such as ‘13219 S penrose Ave’ and ‘Exit 34’ as well as complex location expressions, such as ‘50 miles SW of Liverpool’ and ‘25mins away from Northumbria Street’ were annotated.
- **GeoCorpora<sup>47</sup>:** It was created by Wallgrün et al. [177]. In the dataset, location references in tweets were not only annotated but also linked to the toponyms of GeoNames. Therefore, it can be also used to evaluate the geocoding approaches. The dataset corresponds to multiple noteworthy events (e.g., earthquake, ebola, fire, flood, protest, and rebel) that happened across the world in 2014 and 2015. The majority of the annotated places are admin units, such as continent, country, state, and city.
- **CrisisBench-1000, HumAID-1000, COVID19-1000<sup>48</sup>:** The three datasets were created by ourselves. Specifically, we randomly selected 1000 tweets from CrisisBench [8], HumAID [54], and a COVID19 dataset [102], respectively, and then manually annotated place names, including admin units (e.g., country and village), traffic ways (e.g., street and highway), natural features (e.g., hill and river), and POIs (e.g., park and school).
- **BTC-A, BTC-B, BTC-E, BTC-F, BTC-G, BTC-H<sup>49</sup>:** The Board Twitter Corpus (BTC) was created by Derczynski et al. [43]. The datasets were sampled across different regions, temporal periods, and types of Twitter users. Apart from *Location*, *Organization* and *Person* were also annotated. Several annotated place names are in mentions (e.g., ‘@HoustonFlood’). However, they are normally ignored by existing location extractors. Thus, we remove the place name in the mentions from the six datasets.
- **NEEL2016<sup>50</sup>:** It is the gold dataset of 2016 Named Entity rEcognition and Linking (NEEL) Challenge. The dataset includes event-annotated tweets covering multiple noteworthy events from 2011 to 2013, such as the death of Amy Winehouse, the London Riots, the Oslo bombing, and the Westgate Shopping Mall shootout. Entities of different types, such as *Location*, *Person*, *Organization*, *Event*, and *Product* were not only annotated but also linked to entries in DBpedia. We used its training set, which contains 2135 tweets and 602 places.
- **Ritte’s dataset:**<sup>51</sup> It was initially used by Ritter et al. [151]. *Location*, *Facility*, *Person*, and *Organization* were annotated in the dataset. We used its training set, which contains 2394 tweets and 276 places.
- **MSM2013:**<sup>52</sup> It is the gold dataset of Concept Extraction Challenge held at the Making Sense of Microposts Workshop in 2013 (#MSM2013). Entities of *Person*, *Organization*, *Location*, and *MISC* were annotated. We used its training set, which contains 2815 tweets and 619 places.
- **WNUT2016<sup>53</sup>:** It is the gold data of the shared task on named entity recognition in Twitter. The task is part of the 2nd Workshop on Noisy User-generated Text (W-NUT 2016). Ten types of entities were annotated, such as *Location*, *Facility*, *Person*, and *Movie*. We used its training set, which contains 3850 tweets and 602 places.

<sup>47</sup> <https://github.com/geovista/GeoCorpora>

<sup>48</sup> [https://github.com/uhuohuy/GazPNE2/tree/main/data/test\\_data](https://github.com/uhuohuy/GazPNE2/tree/main/data/test_data)

<sup>49</sup> [https://github.com/GateNLP/broad\\_twitter\\_corpus](https://github.com/GateNLP/broad_twitter_corpus)

<sup>50</sup> <http://microposts2016.seas.upenn.edu/challenge.html>

<sup>51</sup> [https://github.com/ritter/twitter\\_nlp/blob/master/data/annotated/ner.txt](https://github.com/ritter/twitter_nlp/blob/master/data/annotated/ner.txt)

<sup>52</sup> [https://www.researchgate.net/profile/Andrea-Varga-4/publication/256682215\\_MSM2013\\_Concept\\_Extraction\\_Challenge\\_dataset](https://www.researchgate.net/profile/Andrea-Varga-4/publication/256682215_MSM2013_Concept_Extraction_Challenge_dataset)

<sup>53</sup> <https://metatext.io/datasets/wnut-2016>

Table 4. Summary of 26 datasets. There are in total 39,736 places.

Name	Source	Type	Tweet (Article) Count	Place Count	Resolved	Description
LaFlood2016 [6]	tweet	LE	1500	2295	No	Louisiana flood in 2016
HouFlood2015 [6]	tweet	LE	1500	3060	No	Houston flood in 2015
CheFlood2015 [6]	tweet	LE	1500	3671	No	Chennai flood in 2015
Harvey2017 [180]	tweet	LE	1000	2107	No	2017 Hurricane Harvey in Texas and Louisiana
NzEq2013 [126]	tweet	LE	1994	1252	No	New York hurricane in 2012
NyHurcn2012 [126]	tweet	LE	1997	764	No	New Zealand earthquake in 2013
Martinez_I [123]	tweet	LE	800	539	No	Multiple emergency events across the world
Martinez_II [123]	tweet	LE	1371	642	No	Multiple emergency events across the world
Martinez_III [51]	tweet	LE	8063	5122	No	Multiple emergency events across the world
CrisisBench-1000 [81]	tweet	LE	1000	1600	No	1000 tweets from CrisisBench [8]
HumAID-1000 [81]	tweet	LE	1000	1500	No	1000 tweets from HumAid [54]
COVID19-1000 [81]	tweet	LE	1000	800	No	1000 tweets from COVID-19 [102]
GeoCorpora [177]	tweet	LE	6634	3083	Yes	Multiple events across the world
BTC-A [43]	tweet	NER	2000	229	No	Section A of Broad Twitter Corpus
BTC-B [43]	tweet	NER	200	148	No	Section B of Broad Twitter Corpus
BTC-E [43]	tweet	NER	2000	572	No	Section E of Broad Twitter Corpus
BTC-F [43]	tweet	NER	2113	1330	No	Section F of Broad Twitter Corpus
BTC-G [43]	tweet	NER	1999	287	No	Section G of Broad Twitter Corpus
BTC-H [43]	tweet	NER	1000	119	No	Section H of Broad Twitter Corpus
NEEL2016 [152]	tweet	NER	2135	602	Yes	Dataset of Named Entity rEcognition and Linking Challenge in 2016
Ritte's dataset [151]	tweet	NER	2394	276	No	A general-purpose NER dataset initially used in [151]
MSM2013 [28]	tweet	NER	2815	619	No	Dataset of Concept Extraction Challenge at the Making Sense of Microposts Workshop in 2013
WNUT2016 [166]	tweet	NER	3850	602	No	Dataset of shared task on NER in Twitter at the Workshop on Noisy User-generated Text in 2016
LGL [108]	news	LE	588	5050	Yes	Local-Global Lexicon corpus
GeoVirus [64]	news	LE	229	2167	Yes	WikiNews related to global disease and epidemics
TR-News [93]	news	LE	118	1300	Yes	Annotated news articles from various news sources

- **LGL**<sup>54</sup>: Local-Global Lexicon (LGL) corpus was created by Lieberman et al. [108]. Toponyms were manually annotated and geocoded from 588 human-annotated news articles published by 78 local newspapers.
- **GeoVirus**<sup>55</sup>: GeoVirus was created by Gritta et al. [64] for the evaluation of geoparsing approaches in news related to disease outbreaks and epidemics, such as Ebola, Bird Flu, and Swine Flu. Toponyms were manually annotated and geocoded. Only admit units were annotated in the dataset. Buildings, POIs, streets, and rivers were ignored.
- **TR-News**<sup>56</sup>: TR-News was created by Kamalloo and Rafiei [93]. Toponyms were manually annotated and geocoded from 118 news articles from various news sources.

We adopted the standard comparison metrics: precision, recall, and F1-score. In the case of overlapping or partial matches, we penalize a tool by adding 1/2 FP (False Positive) and 1/2 FN (False Negative) (e.g., if the tool marks ‘*The Houston*’ instead of ‘*Houston*’), following Al-Olimat et al. [6].

$$Precision = \frac{TP}{TP + FP} \quad (1)$$

$$Recall = \frac{TP}{TP + FN} \quad (2)$$

<sup>54</sup><https://github.com/milangritta/Pragmatic-Guide-to-Geoparsing-Evaluation/blob/master/data/Corpora/lgl.xml>

<sup>55</sup><https://github.com/milangritta/Pragmatic-Guide-to-Geoparsing-Evaluation/blob/master/data/Corpora/GeoVirus.xml>

<sup>56</sup><https://github.com/milangritta/Pragmatic-Guide-to-Geoparsing-Evaluation/blob/master/data/Corpora/TR-News.xml>

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (3)$$

### 4.3 Results of location reference recognition

We run the 27 tools on all the test datasets, and their precision, recall, and F1 score are reported in Figure 6. The three metrics are obtained by calculating the sum of FP, FN, and TP of all the datasets rather than by averaging the three metrics of the datasets due to the imbalanced place name count (from 119 to 5122) of different datasets. The raw result of the tools on each dataset can be seen from <sup>57</sup>. The majority of tools can achieve very high precision with 21 of 27 achieving a precision over 0.7. Conversely, most of them obtain very low recall with only 2 (GazPNE2 and LORE) of 27 achieving a recall over 0.7. This means that most of the tools missed many location references. Therefore, we further investigate the reason in Section 4.4 by analyzing the detection accuracy of the tools in different types of texts and different types of location references. We can also observe that the top five best performing tools, GazPNE2, Flair NER (Ont), nLORE, Flair NER, and Stanza, are all based on deep learning techniques and were proposed in the last four years, showing the superior performance and progress of deep learning on this task. Furthermore, two voting-based systems, SPENS and RS also achieve impressive results by simply combining the detection results of several classic tools, improving the performance of each single tool. The two voting systems are comparable with Stanza and Flair NER. This suggests a great potential of using voting mechanism in location reference recognition.

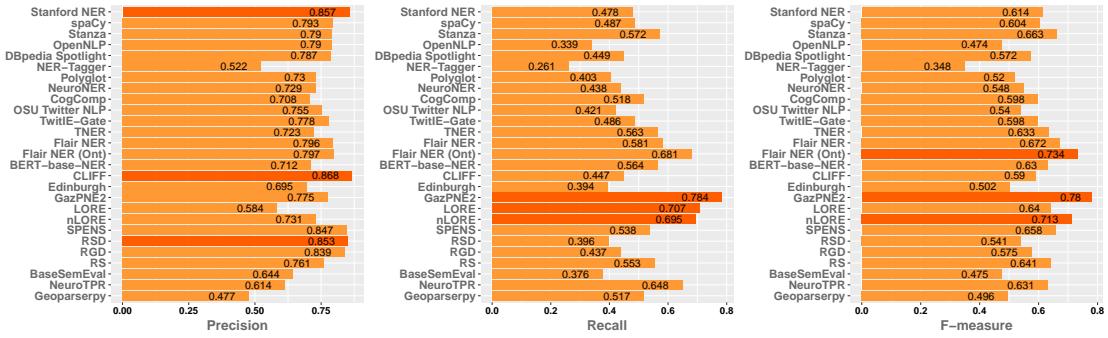


Fig. 6. Precision, recall, and F1 score of the tested tools on all the datasets with 39,736 location references.

### 4.4 Error analysis

To gain an insight into what mistakes these tools made, we carried out an error analysis. Specifically, we investigated the performance of the tools on formal and informal texts. We also examined their performance on location references in different categories and forms.

**4.4.1 Text type.** Among the test datasets, 3 datasets are formal texts and 23 datasets are informal texts, containing 8517 and 31,219 places, respectively. The performance of the tools on the datasets of formal and informal texts is shown in Figure 7 and Figure 8, respectively. We can observe that on formal texts, Flair NER (Ont), Flair NER, SPENS, Stanford NER, and Stanza perform the best, while on informal texts, GazPNE2 and nLORE perform the best. The main reason is

<sup>57</sup>[https://docs.google.com/spreadsheets/d/16cyuyDhty04hQE1gBfP4zq23Lr3OWJ5EY\\_bsVMoYxfQ/edit#gid=1536784545](https://docs.google.com/spreadsheets/d/16cyuyDhty04hQE1gBfP4zq23Lr3OWJ5EY_bsVMoYxfQ/edit#gid=1536784545)

that the former five tools are all trained on formal texts. Conversely, the latter two tools were proposed to deal with tweets and thus work better on twitter datasets.

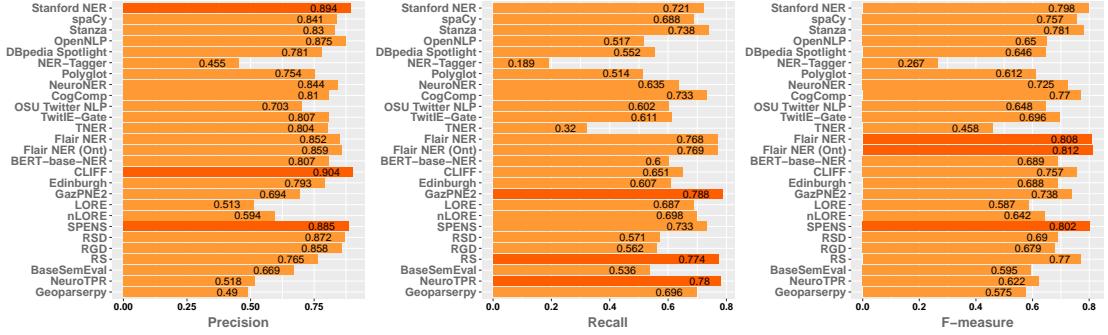


Fig. 7. Precision, recall, and F1 score of the tested tools on the datasets of formal texts, containing 8517 places.

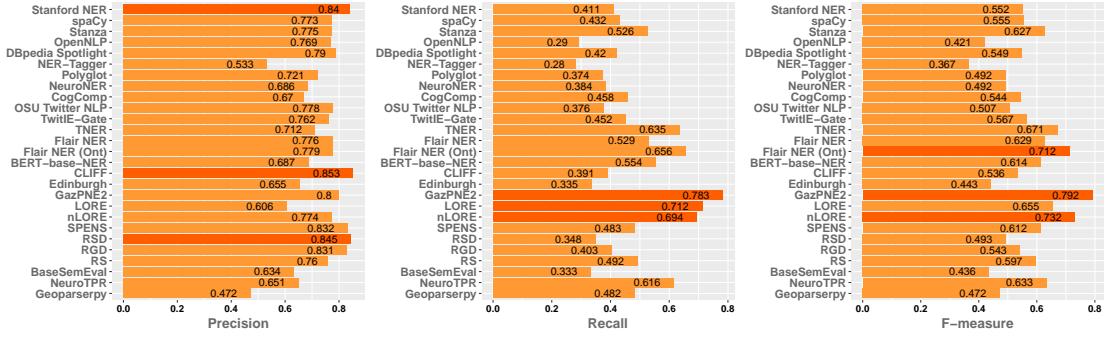


Fig. 8. Precision, recall, and F1 score of the tested tools on the datasets of informal texts, containing 31,219 places.

**4.4.2 Place category.** We divide the location references in the datasets into four categories: admin units (e.g., country, state, town, and suburb), traffic ways (e.g., street, road, highway, and bridge), natural features (e.g., river, creek, beach, and hill), and POIs (e.g., park, church, school, and library). We choose four datasets (i.e., Harvey2017, GeoCorpora, LGL, and TR-News) to conduct this experiment because the category of the places in the four datasets can be derived. The places in Harvey2017 were categorized into ten types [84]. We treat the types of house number addresses, street names, highway, exits of highways, and intersections of roads as traffic ways, the type of natural features as natural features, the types of other human-made features and local organizations as POIs, and the types of admin units and multiple areas as admin units. In the other three datasets, place names were linked to the entries of GeoNames. We treat the places whose feature codes<sup>58</sup> are *A* (e.g., country, state, and region) and *P* (e.g., city and village) in GeoNames as admin units, the places whose feature codes are *R* (e.g., road and railroad) as traffic ways, the places whose feature codes are *H* (e.g., stream and lake), *T* (e.g., mountain, hill, and rock), *U* (e.g., undersea and valley), *V* (e.g., forest and grove) as

<sup>58</sup><http://www.geonames.org/export/codes.html>

natural features, and the places whose feature codes are  $L$  (e.g., park and port) and  $S$  (e.g., sport, building, and farm) as POIs. There are in total 9790 admin units, 773 traffic ways, 263 natural features, and 754 POIs in the four datasets.

We define the detection rate as the proportion of correctly detected places among the total places of a certain category. Only an exact match is regarded as a correct detection. Figure 9 shows the detection rate of the tools on the four categories. We can observe that many tools show superior performance in recognizing coarse-grained places, with 13 of 27 recognizing over 60% of the admin units. However, most of them are incapable of recognizing fine-grained places. Only 2 of 27, 6 of 27, and 4 of 27 can recognize over 60% of the traffic ways, nature features, and POIs, respectively. The three categories refer to much more precise geographic scopes than admin units and are thus valuable in many key applications, such as emergency rescue and traffic event detection. It is worth mentioning that GazPNE2 can recognize over 70% of places in all four categories.

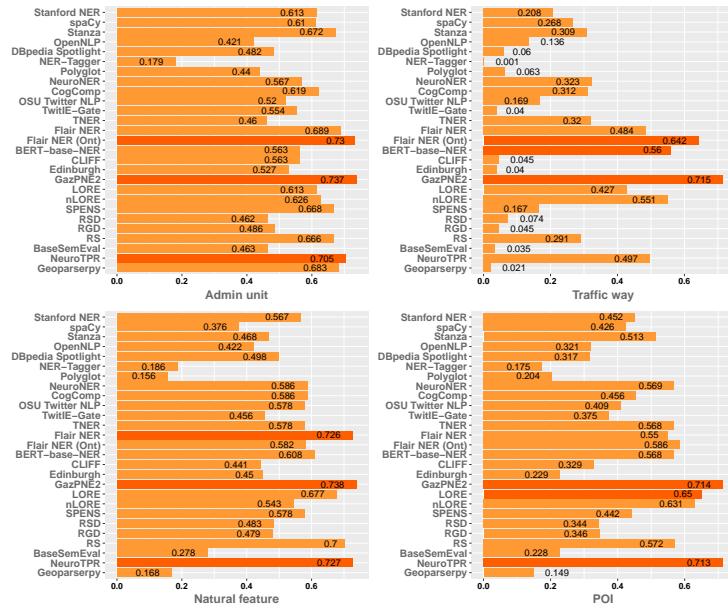


Fig. 9. Detection rate of tools on the four categories with 9790 admin units, 773 traffic ways, 263 natural features, and 754 POIs.

**4.4.3 Form of location references.** We consider three forms of location references: the place names with numbers (e.g., ‘500 Neches Ave’ and ‘Highway 25’), abbreviation of place names (e.g., ‘us’ and ‘tx’), and place names in hashtags (e.g., ‘#HoustonFlood’ and ‘#Chennai’). Place names with numbers mostly refer to fine-grained locations, such as highways, roads, and home addresses. We define that the abbreviation of place names consists of only one single word and its character length does not surpass 3. 1621, 3697, and 6560 place names are in the three forms (number, abbreviation, and hashtag), respectively. Figure 10 shows the detection rate of the tools on the place names of the three forms.

We can observe that it is a challenge to recognize place names with numbers, with only 4 of 27 achieving a detection rate of over 0.3. Conversely, recognizing abbreviations is easier, with over half (16 of 27) of the tools recognizing over 30% of the abbreviations. However, we can also observe that no tool can achieve a detection rate over 0.6 on the two

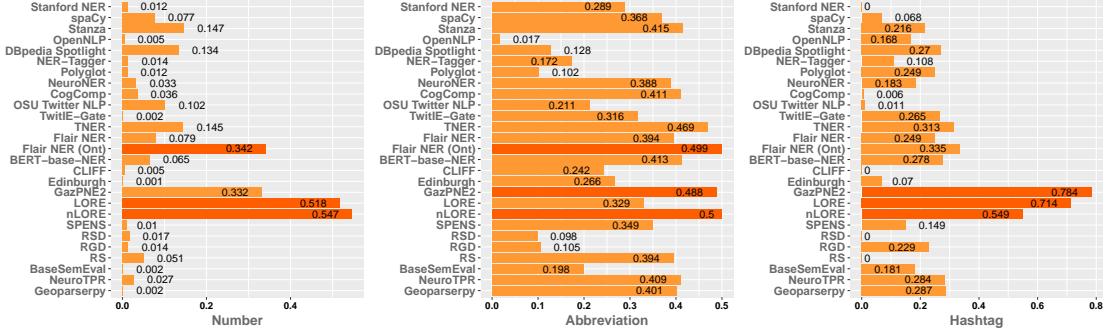


Fig. 10. Detection rate of tools on location references in different forms with 1621 place names with numbers, 3697 abbreviations, and 5650 place names in hashtags.

forms. It is also a challenge to recognize place names in hashtags, with only 5 of 27 achieving a detection rate of over 0.3. However, it is encouraging that GazPNE2 and LORE can recognize over 70% of the places in hashtags.

#### 4.5 Computational efficiency

In this section, we further investigate the computational efficiency (i.e., speed) of different approaches. In many applications, the texts that need to be geoparsed are of huge volumes, such as some major historical books and reports (e.g., the Old Bailey Online) that each comprise many millions or even billions of words [62] and over millions of tweets related to a crisis event [146]. This requires a rapid geoparsing procedure and the speed is thus a critical indicator.

We run each approach on the total datasets and record the consumed time of each approach. We do not count the consumed time during the training phase if an approach needs to be trained since it can be conducted offline and is a one-time process. Note that, we do not include Edinburgh Geoparser and DBpedia Spotlight in the comparison since they are online service and it is impossible to count their amount of time of processing done on the server. While most of approaches run on a MacBook Pro laptop with an Intel Core i7 (2.2 GHz 6-Core) and a RAM of 16 GB, three approaches (i.e., OSU TwitterNLP, LORE, and nLORE) run on a Lenovo laptop with an Intel Core i5 (2.5 GHZ 4-Core) and a RAM of 3.8 GB since they require a Linux or Windows environment. Figure 11 illustrates the speed of the approaches.

We can observe that the speed of different approaches varies drastically. It takes 6 minutes to 33 hours for these approaches to process the total datasets that contain 1,092,093 words. It is unexpected that OSU Twitter NLP takes nearly 9.6 hours. Therefore, RSD, RGD, and RS that use the OSU Twitter NLP take also nearly 10 hours. The other approaches that take over 5 hours are all deep learning-based. spaCy, Cliff, LORE, Ployglot, and OpenNLP are 20 times quicker than these approaches. However, the deep learning-based approaches achieve a much higher F1 score than the other approaches. Therefore, there exists a trade-off between correctness and computational efficiency.

## 5 CONCLUSIONS AND OUTLOOK

In this paper, we first summarized seven typical applications of geoparsing, and then presented a survey of existing approaches for location reference recognition. We grouped these existing approaches into four categories: rule-based, gazetteer matching-based, statistical learning-based, and hybrid approaches. In addition, we carried out a comparative study to systematically compare 27 existing approaches on 26 datasets. We evaluated these approaches from multiple perspectives including their overall location recognition accuracy, their performance on formal (e.g., news) and informal

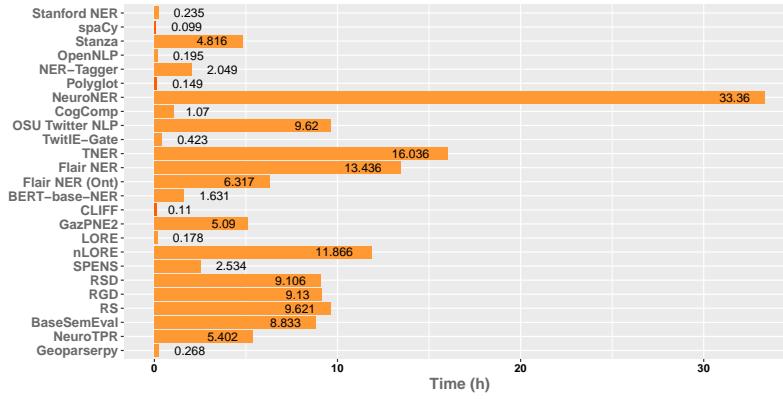


Fig. 11. Time consumption of the approaches running on the total test datasets.

(e.g., tweets) texts, and their capabilities to detect location references in different categories and forms. Finally, we also compared the computational efficiency of these existing approaches.

From the results, we can conclude that: (1) deep learning is so far the most promising technique in location reference recognition; (2) fusing existing approaches or tools in a voting mechanism can overcome each others' shortcomings and is more robust than each single approach; (3) the performance of different approaches varies on the type of texts and the attributes of location references and their computational efficiency also vary drastically. Users should choose the proper one according to their specific application demands. For example, for traffic management and disaster management that rely on informal texts and require fine-grained locations, GazPNE2 is a good option; for crime management and tourism management that rely on formal texts and require both coarse-grained and fine-grained locations, Flair NER (Ont) is a good option; for spatial humanities, disease surveillance, and GIR that need process a large number of formal texts and require just coarse-grained locations, Standford NER, CogComp, and spaCy are good options.

Several research directions can be further explored in the future.

- **Location reference recognition:** There is still space to improve the performance of the approaches for informal texts (e.g., tweets) since most of the existing approaches did not perform well on informal texts. Inspired by the success of two voting systems RS and SPENS, one of the promising ways might be further selecting several approaches from the 27 approaches and combining them in a voting system to achieve more satisfying results. We can also configure the combined approaches of the voting system to satisfy the requirements of different applications, such as a high recall on POIs of tweets.
- **Location reference geocoding:** While there exist many approaches for geocoding location references, they focused mainly on formal texts. Some studies proposed geoparsing tweets, but usually limited the application area to a known geographic region (e.g., a city where a flood occurs) [3, 6]. In those cases, simply searching a local gazetteer is often sufficient for geocoding. Only a few studies [94, 138] geoparsed tweets at a global scale. This task has two main challenges: geo/geo ambiguities caused by limited contexts in short texts of tweets and unseen place names caused by place name variants and the informal features of tweets that might contain abbreviations, slang, and misspellings. Three main ways might be explored to overcome the challenges: (1) to leverage deep learning [29, 55, 98] and clustering techniques [41] that can group tweets of the same topic to expand the context of tweets; (2) to combine multiple SOTA geocoders in a voting mechanism; (3) to generate

abundant and high-quality training examples by using geotagged tweets. The geotag of a tweet can drastically remove the geo/geo ambiguities of the place names in the tweet text, enabling the generation of high-quality training examples.

- **Datasets for geoparsing research:** Existing datasets for geoparsing research are often in the form of formal text such as news, and only a few twitter datasets (e.g., GeoCorpora and NEEL2016) are available for geoparsing research. Although there exist many other twitter datasets for general NER research, these datasets typically do not contain geographic coordinates for the labeled entities and therefore cannot support research involving the entire workflow of geoparsing. More datasets in informal texts with labeled location references and their geographic coordinates are needed. Furthermore, most of the location references in the existing datasets are admin units, such as countries and cities. Finer-grained location references, such as traffic ways and POIs, are much rare. However, they are important in many applications, such as determining the precise locations where rescue is needed during disasters. A large twitter dataset that contains many fine-grained locations across the world would be very helpful for advancing methods in recognizing and geocoding fine-grained location references in texts.

## REFERENCES

- [1] Ahmed Abdelkader, Emily Hand, and Hanan Samet. 2015. Brands in newsstand: Spatio-temporal browsing of business news. In *Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems*. 1–4.
- [2] Elise Acheson and Ross S Purves. 2021. Extracting and modeling geographic information from scientific articles. *PLoS one* 16, 1 (2021), e0244918.
- [3] Mohammed Faisal Ahmed, Lelitha Vanajakshi, and Ramasubramanian Suriyanarayanan. 2019. Real-Time Traffic Congestion Information from Tweets Using Supervised and Unsupervised Machine Learning Techniques. *Transportation in Developing Economies* 5, 2 (2019), 1–11.
- [4] Alan Akbik, Tanja Bergmann, Duncan Blythe, Kashif Rasul, Stefan Schweter, and Roland Vollgraf. 2019. FLAIR: An easy-to-use framework for state-of-the-art NLP. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (Demonstrations)*. 54–59.
- [5] Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. Contextual string embeddings for sequence labeling. In *Proceedings of the 27th international conference on computational linguistics*. 1638–1649.
- [6] Hussein Al-Olimat, Krishnaprasad Thirunarayan, Valerie Shalin, and Amit Sheth. 2018. Location Name Extraction from Targeted Text Streams using Gazetteer-based Statistical Language Models. *Proceedings of the 27th International Conference on Computational Linguistics* (Aug. 2018), 1986–1997. <https://www.aclweb.org/anthology/C18-1169>
- [7] Rami Al-Rfou, Vivek Kulkarni, Bryan Perozzi, and Steven Skiena. 2015. Polyglot-NER: Massive multilingual named entity recognition. In *Proceedings of the 2015 SIAM International Conference on Data Mining*. SIAM, 586–594.
- [8] Firoj Alam, Hassan Sajjad, Muhammad Imran, and Ferda Ofli. 2020. CrisisBench: Benchmarking Crisis-related Social Media Datasets for Humanitarian Information Processing. *arXiv preprint arXiv:2004.06774* (2020).
- [9] Edwin Aldana-Bobadilla, Alejandro Molina-Villegas, Ivan Lopez-Arevalo, Shanel Reyes-Palacios, Victor Muñiz-Sánchez, and Jean Arreola-Trapala. 2020. Adaptive Geoparsing Method for Toponym Recognition and Resolution in Unstructured Text. *Remote Sensing* 12, 18 (2020), 3041.
- [10] Farman Ali, Amjad Ali, Muhammad Imran, Rizwan Ali Naqvi, Muhammad Hameed Siddiqi, and Kyung-Sup Kwak. 2021. Traffic accident detection and condition analysis based on social networking data. *Accident Analysis & Prevention* 151 (2021), 105973.
- [11] Balsam Alkouz and Zaher Al Aghbari. 2020. SNSJam: Road traffic analysis and prediction by fusing data from multiple social networks. *Information Processing & Management* 57, 1 (2020), 102139.
- [12] Toph Allen, Kris A Murray, Carlos Zambrano-Torrello, Stephen S Morse, Carlo Rondinini, Moreno Di Marco, Nathan Breit, Kevin J Olival, and Peter Daszak. 2017. Global hotspots and correlates of emerging zoonotic diseases. *Nature communications* 8, 1 (2017), 1–10.
- [13] Ebtessam Alomari, Iyad Katib, Aiiad Albeshri, Tan Yigitcanlar, and Rashid Mahmood. 2021. Iktishaf+: A big data tool with automatic labeling for road traffic social sensing and event detection using distributed machine learning. *Sensors* 21, 9 (2021), 2993.
- [14] Meshrif Alruaij, Aladdin Ayesh, and Hussein Zedan. 2014. Crime profiling for the Arabic language using computational linguistic techniques. *Information processing & management* 50, 2 (2014), 315–341.
- [15] Einat Amitay, Nadav Har'El, Ron Sivan, and Aya Soffer. 2004. Web-a-where: geotagging web content. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*. 273–280.
- [16] Stelios Andreadis, Gerasimos Antzoulatos, Thanassis Mavropoulos, Panagiotis Giannakeris, Grigoris Tzionis, Nick Pantelidis, Konstantinos Ioannidis, Anastasios Karakostas, Ilias Gialampoukidis, Stefanos Vrochidis, et al. 2021. A social media analytics platform visualising the spread of COVID-19 in Italy via exploitation of automatically geotagged tweets. *Online Social Networks and Media* 23 (2021), 100134.

- [17] Raxy Arulanandam, Bastin Tony Roy Savarimuthu, and Maryam A Purvis. 2014. Extracting crime information from online newspaper articles. In *Proceedings of the second australasian web conference-volume 155*. 31–38.
- [18] Marco Avvenuti, Stefano Cresci, Fabio Del Vigna, Tiziano Fagni, and Maurizio Tesconi. 2018. CrisMap: a big data crisis mapping system based on damage detection and geoparsing. *Information Systems Frontiers* 20, 5 (2018), 993–1011.
- [19] Marco Avvenuti, Stefano Cresci, Leonardo Nizzoli, and Maurizio Tesconi. 2018. GSP (Geo-Semantic-Parsing): geoparsing and geotagging with machine learning on top of linked data. In *European Semantic Web Conference*. Springer, 17–32.
- [20] Dariusz B Baranowski, Maria K Flatau, Piotr J Flatau, Dwikorita Karnawati, Katarzyna Barabasz, Michal Labuz, Beata Latos, Jerome M Schmidt, Jaka AI Paski, et al. 2020. Social-media and newspaper reports reveal large-scale meteorological drivers of floods on Sumatra. *Nature communications* 11, 1 (2020), 1–10.
- [21] Moumita Basu, Sipra Das Bit, and Saptarshi Ghosh. 2022. Utilizing microblogs for optimized real-time resource allocation in post-disaster scenarios. *Social Network Analysis and Mining* 12, 1 (2022), 1–20.
- [22] Loris Belcastro, Fabrizio Marozzo, Domenico Talia, Paolo Trunfio, Francesco Branda, Themis Palpanas, and Muhammad Imran. 2021. Using social media for sub-event detection during disasters. *Journal of Big Data* 8, 1 (2021), 1–22.
- [23] Kalina Bontcheva, Leon Derczynski, Adam Funk, Mark A Greenwood, Diana Maynard, and Niraj Aswani. 2013. Twitie: An open-source information extraction pipeline for microblog text. In *Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013*. 83–90.
- [24] Karla AV Borges, Alberto HF Laender, Claudia B Medeiros, and Clodoveu A Davis Jr. 2007. Discovering geographic locations in web pages using urban addresses. In *Proceedings of the 4th ACM workshop on Geographical information retrieval*. 31–36.
- [25] Igo Ramalho Brilhante, Jose Antonio Macedo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. 2015. On planning sightseeing tours with TripBuilder. *Information Processing & Management* 51, 2 (2015), 1–15.
- [26] Lou Burnard. 1995. *Users reference guide for the British National Corpus*. Oxford University Computing Services.
- [27] Lucie Cadorel, Alicia Blanchi, and Andrea GB Tettamanzi. 2021. Geospatial Knowledge in Housing Advertisements: Capturing and Extracting Spatial Information from Text. In *Proceedings of the 11th on Knowledge Capture Conference*. 41–48.
- [28] Amparo Elizabeth Cano Basave, Andrea Varga, Matthew Rowe, Milan Stankovic, and Aba-Sah Dadzie. 2013. Making sense of microposts (# msm2013) concept extraction challenge. (2013).
- [29] Ana Bárbara Cardoso, Bruno Martins, and Jacinto Estima. 2021. A Novel Deep Learning Approach Using Contextual Embeddings for Toponym Resolution. *ISPRS International Journal of Geo-Information* 11, 1 (2021), 28.
- [30] Zi Chen, Badal Pokharel, Bingnan Li, and Samsung Lim. 2020. Location Extraction from Twitter Messages using Bidirectional Long Short-Term Memory Model. In *GISTAM*. 45–50.
- [31] Zhiyuan Cheng, 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.
- [32] Rumi Chunara, Jason R Andrews, and John S Brownstein. 2012. Social and news media enable estimation of epidemiological patterns early in the 2010 Haitian cholera outbreak. *The American journal of tropical medicine and hygiene* 86, 1 (2012), 39.
- [33] Paul Clough. 2005. Extracting metadata for spatially-aware information retrieval on the internet. In *Proceedings of the 2005 workshop on Geographic information retrieval*. 25–30.
- [34] Andrea Fronzetti Colladon, Barbara Guardabascio, and Rosy Innarella. 2019. Using social network and semantic analysis to analyze online travel forums and forecast tourism demand. *Decision Support Systems* 123 (2019), 113075.
- [35] Stefano Cresci, Andrea D'Errico, Davide Gazzé, Angelica Lo Duca, Andrea Marchetti, and Maurizio Tesconi. 2014. Towards a DBpedia of Tourism: the case of Tourpedia.. In *International Semantic Web Conference (Posters & Demos)*. 129–132.
- [36] Rafael Prieto Curiel, Stefano Cresci, Cristina Ioana Muntean, and Steven Richard Bishop. 2020. Crime and its fear in social media. *Palgrave Communications* 6, 1 (2020), 1–12.
- [37] James R Curran and Stephen Clark. 2003. Language independent NER using a maximum entropy tagger. In *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003*. 164–167.
- [38] Rahul Deb Das and Ross S Purves. 2019. Exploring the Potential of Twitter to Understand Traffic Events and Their Locations in Greater Mumbai, India. *IEEE Transactions on Intelligent Transportation Systems* 21, 12 (2019), 5213–5222.
- [39] Tirthankar Dasgupta, Lipika Dey, Rupsa Saha, and Abir Naskar. 2018. Automatic Curation and Visualization of Crime Related Information from Incrementally Crawled Multi-source News Reports. In *Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations*. 103–107.
- [40] Tirthankar Dasgupta, Abir Naskar, Rupsa Saha, and Lipika Dey. 2017. Crimeprofiler: Crime information extraction and visualization from news media. In *Proceedings of the International Conference on Web Intelligence*. 541–549.
- [41] Jens A de Brujin, Hans de Moel, Brenden Jongman, Jurjen Wagenaar, and Jeroen CJH Aerts. 2018. TAGGS: Grouping tweets to improve global geoparsing for disaster response. *Journal of Geovisualization and Spatial Analysis* 2, 1 (2018), 2.
- [42] Grant DeLozier, Jason Baldridge, and Loretta London. 2015. Gazetteer-independent toponym resolution using geographic word profiles. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- [43] Leon Derczynski, Kalina Bontcheva, and Ian Roberts. 2016. Broad twitter corpus: A diverse named entity recognition resource. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*. 1169–1179.

- [44] Franck Dernoncourt, Ji Young Lee, and Peter Szolovits. 2017. NeuroNER: an easy-to-use program for named-entity recognition based on neural networks. *arXiv preprint arXiv:1705.05487* (2017).
- [45] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [46] Catherine D'Ignazio, Rahul Bhargava, Ethan Zuckerman, and Luisa Beck. 2014. Cliff-clavin: Determining geographic focus for news articles. *NewsKDD: Data Science for News Publishing, at KDD 2014* (2014).
- [47] Christopher Donaldson, Ian N Gregory, and Joanna E Taylor. 2017. Locating the beautiful, picturesque, sublime and majestic: spatially analysing the application of aesthetic terminology in descriptions of the English Lake District. *Journal of Historical Geography* 56 (2017), 43–60.
- [48] Ritam Dutt, Moumita Basu, Kripabandhu Ghosh, and Saptarshi Ghosh. 2019. Utilizing microblogs for assisting post-disaster relief operations via matching resource needs and availabilities. *Information Processing & Management* 56, 5 (2019), 1680–1697.
- [49] Ritam Dutt, Kaustubh Hiware, Avijit Ghosh, and Rameshwar Bhaskaran. 2018. Savitr: A system for real-time location extraction from microblogs during emergencies. In *Companion Proceedings of the The Web Conference 2018*. 1643–1649.
- [50] Chao Fan, Fangsheng Wu, and Ali Mostafavi. 2020. A hybrid machine learning pipeline for automated mapping of events and locations from social media in disasters. *IEEE Access* 8 (2020), 10478–10490.
- [51] Nicolás José Fernández-Martínez and Carlos Perinián Pascual. 2021. *nLORE: A Linguistically Rich Deep-Learning System for Locative-Reference Extraction in Tweets*. <https://doi.org/10.3233/AISE210103>
- [52] Paolo Ferragina and Ugo Scaiella. 2010. Tagme: on-the-fly annotation of short text fragments (by wikipedia entities). In *Proceedings of the 19th ACM international conference on Information and knowledge management*. 1625–1628.
- [53] Jenny Rose Finkel, Trond Grenager, and Christopher Manning. 2005. Incorporating non-local information into information extraction systems by gibbs sampling. In *Proceedings of the 43rd annual meeting on association for computational linguistics*. Association for Computational Linguistics, 363–370.
- [54] Muhammad Imran Ferda Firoj Alam, Umair Qazi. 2021. HumAID: Human-Annotated Disaster Incidents Data from Twitter. In *15th International Conference on Web and Social Media (ICWSM)*.
- [55] Jacques Fize, Ludovic Moncla, and Bruno Martins. 2021. Deep Learning for Toponym Resolution: Geocoding Based on Pairs of Toponyms. *ISPRS International Journal of Geo-Information* 10, 12 (2021), 818.
- [56] Nuno Freire, José Borbinha, Pável Calado, and Bruno Martins. 2011. 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*. 339–348.
- [57] Judith Gelernter and Shilpa Balaji. 2013. An algorithm for local geoparsing of microtext. *GeoInformatica* 17, 4 (2013), 635–667.
- [58] Judith Gelernter and Nikolai Mushegian. 2011. Geo-parsing messages from microtext. *Transactions in GIS* 15, 6 (2011), 753–773.
- [59] Judith Gelernter and Wei Zhang. 2013. Cross-lingual geo-parsing for non-structured data. In *Proceedings of the 7th Workshop on Geographic Information Retrieval*. 64–71.
- [60] Prasanna Giridhar, Tarek Abdelzaher, Jemin George, and Lance Kaplan. 2015. On quality of event localization from social network feeds. In *2015 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*. IEEE, 75–80.
- [61] Ian Gregory, Paul Atkinson, Andrew Hardie, Amelia Joulain-Jay, Daniel Kershaw, Catherine Porter, Paul Rayson, and Christopher John Rupp. 2016. From digital resources to historical scholarship with the British library 19th century newspaper collection. (2016).
- [62] Ian Gregory, Christopher Donaldson, Patricia Murrieta-Flores, and Paul Rayson. 2015. Geoparsing, GIS, and textual analysis: current developments in spatial humanities research. *International Journal of Humanities and Arts Computing* 9, 1 (2015), 1–14.
- [63] Milan Gritta. 2019. *Where are you talking about? advances and challenges of geographic analysis of text with application to disease monitoring*. Ph.D. Dissertation. University of Cambridge.
- [64] Milan Gritta, Mohammad Pilehvar, and Nigel Collier. 2018. Which melbourne? augmenting geocoding with maps. (2018).
- [65] Milan Gritta, Mohammad Taher Pilehvar, and Nigel Collier. 2018. A pragmatic guide to geoparsing evaluation. *arXiv preprint arXiv:1810.12368* (2018).
- [66] Milan Gritta, Mohammad Taher Pilehvar, and Nigel Collier. 2020. A pragmatic guide to geoparsing evaluation. *Language resources and evaluation* 54, 3 (2020), 683–712.
- [67] Milan Gritta, Mohammad Taher Pilehvar, Nut Limsopatham, and Nigel Collier. 2018. What's missing in geographical parsing? *Language Resources and Evaluation* 52, 2 (2018), 603–623.
- [68] Claire Grover, Richard Tobin, Kate Byrne, Matthew Woppard, James Reid, Stuart Dunn, and Julian Ball. 2010. Use of the Edinburgh geoparser for georeferencing digitized historical collections. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 368, 1925 (2010), 3875–3889.
- [69] Marco Guerini, Simone Magnolini, Vevake Balaraman, and Bernardo Magnini. 2018. Toward zero-shot entity recognition in task-oriented conversational agents. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. 317–326.
- [70] Carlos Gutierrez, Paulo Figuerias, Pedro Oliveira, Ruben Costa, and Ricardo Jardim-Goncalves. 2015. Twitter mining for traffic events detection. In *2015 Science and information conference (SAI)*. IEEE, 371–378.
- [71] Mena B Habib and Maurice van Keulen. 2013. A hybrid approach for robust multilingual toponym extraction and disambiguation. In *Intelligent Information Systems Symposium*. Springer, 1–15.
- [72] Andrew Halterman. 2017. Mordecai: Full text geoparsing and event geocoding. *Journal of Open Source Software* 2, 9 (2017), 91.

- [73] Erum Haris and Keng Hoon Gan. 2017. Mining graphs from travel blogs: a review in the context of tour planning. *Information Technology & Tourism* 17, 4 (2017), 429–453.
- [74] Erum Haris, Keng Hoon Gan, and Tien-Ping Tan. 2020. Spatial information extraction from travel narratives: Analysing the notion of co-occurrence indicating closeness of tourist places. *Journal of Information Science* 46, 5 (2020), 581–599.
- [75] Jenine K Harris, Sarah Moreland-Russell, Rachel G Tabak, Lindsay R Ruhr, and Ryan C Maier. 2014. Communication about childhood obesity on Twitter. *American journal of public health* 104, 7 (2014), e62–e69.
- [76] Jingrui He, Wei Shen, Phani Divakaruni, Laura Wynter, and Rick Lawrence. 2013. Improving traffic prediction with tweet semantics. In *Twenty-Third International Joint Conference on Artificial Intelligence*.
- [77] Uta Hinrichs, Beatrice Alex, Jim Clifford, Andrew Watson, Aaron Quigley, Ewan Klein, and Colin M Coates. 2015. Trading consequences: A case study of combining text mining and visualization to facilitate document exploration. *Digital Scholarship in the Humanities* 30, suppl\_1 (2015), i50–i75.
- [78] Thi Bich Ngoc Hoang and Josiane Mothe. 2018. Location extraction from tweets. *Information Processing & Management* 54, 2 (2018), 129–144.
- [79] Alexander Hohl, Moongi Choi, Richard Medina, Neng Wan, and Ming Wen. 2021. Understanding Adverse Population Sentiment Towards the Spread of COVID-19 in the United States. *medRxiv* (2021).
- [80] Xuke Hu, Hussein Al-Olimat, Jens Kersten, Matti Wiegmann, Friederike Klan, Yeran Sun, and Hongchao Fan. 2021. GazPNE: Annotation-free Deep Learning for Place Name Extraction from Microblogs Leveraging Gazetteer and Synthetic Data by Rules. *International Journal of Geographical Information Science* (2021), 1–28. <https://doi.org/10.1080/13658816.2021.1947507>
- [81] Xuke Hu, Zhiyong Zhou, Yeran Sun, Jens Kersten, Friederike Klan, Hongchao Fan, and Matti Wiegmann. 2022. GazPNE2: A general place name extractor for microblogs fusing gazetteers and pretrained transformer models. *IEEE Internet of Things Journal* (2022), 1–1. <https://doi.org/10.1109/JIOT.2022.3150967>
- [82] Yingjie Hu and Benjamin Adams. 2021. Harvesting big geospatial data from natural language texts. In *Handbook of Big Geospatial Data*. Springer, 487–507.
- [83] Yingjie Hu, Krzysztof Janowicz, and Sathya Prasad. 2014. Improving wikipedia-based place name disambiguation in short texts using structured data from dbpedia. In *Proceedings of the 8th workshop on geographic information retrieval*. 1–8.
- [84] Yingjie Hu and Jimin Wang. 2020. How do people describe locations during a natural disaster: an analysis of tweets from Hurricane Harvey. *arXiv preprint arXiv:2009.12914* (2020).
- [85] Yingjie Hu and Ruo-Qian Wang. 2020. Understanding the removal of precise geotagging in tweets. *Nature Human Behaviour* 4, 12 (2020), 1219–1221.
- [86] Qunying Huang and Yu Xiao. 2015. Geographic situational awareness: mining tweets for disaster preparedness, emergency response, impact, and recovery. *ISPRS International Journal of Geo-Information* 4, 3 (2015), 1549–1568.
- [87] Xiao Huang, Zhenlong Li, Yuqin Jiang, Xiaoming Li, and Dwayne Porter. 2020. Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PloS one* 15, 11 (2020), e0241957.
- [88] Diana Inkpen, Ji Liu, Atefeh Farzindar, Farzaneh Kazemi, and Diman Ghazi. 2017. Location detection and disambiguation from twitter messages. *Journal of Intelligent Information Systems* 49, 2 (2017), 237–253.
- [89] Aminreza Iranmanesh and Resmiye Alpar Atun. 2021. Reading the changing dynamic of urban social distances during the COVID-19 pandemic via Twitter. *European Societies* 23, sup1 (2021), S872–S886.
- [90] Christopher B Jones, Ross Purves, Anne Ruas, Mark Sanderson, Monika Sester, M Van Kreveld, and Robert Weibel. 2002. Spatial information retrieval and geographical ontologies an overview of the SPIRIT project. In *Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*. 387–388.
- [91] Kate E Jones, Nikkita G Patel, Marc A Levy, Adam Storeygard, Deborah Balk, John L Gittleman, and Peter Daszak. 2008. Global trends in emerging infectious diseases. *Nature* 451, 7181 (2008), 990–993.
- [92] Yiting Ju, Benjamin Adams, Krzysztof Janowicz, Yingjie Hu, Bo Yan, and Grant McKenzie. 2016. Things and strings: improving place name disambiguation from short texts by combining entity co-occurrence with topic modeling. In *European Knowledge Acquisition Workshop*. Springer, 353–367.
- [93] Ehsan Kamalloo and Davood Rafiei. 2018. A coherent unsupervised model for toponym resolution. In *Proceedings of the 2018 World Wide Web Conference*. 1287–1296.
- [94] Morteza Karimzadeh, Scott Pezanowski, Alan M MacEachren, and Jan O Wallgrün. 2019. GeoTxt: A scalable geoparsing system for unstructured text geolocation. *Transactions in GIS* 23, 1 (2019), 118–136.
- [95] Mikaela Keller, Clark C Freifeld, and John S Brownstein. 2009. Automated vocabulary discovery for geo-parsing online epidemic intelligence. *BMC bioinformatics* 10, 1 (2009), 1–9.
- [96] Sarthak Khanal and Doina Caragea. 2021. Multi-task Learning to Enable Location Mention Identification in the Early Hours of a Crisis Event. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. 4051–4056.
- [97] Hiroshi Kori, Shun Hattori, Taro Tezuka, and Katsumi Tanaka. 2007. Automatic generation of multimedia tour guide from local blogs. In *International conference on multimedia modeling*. Springer, 690–699.
- [98] Sayali Kulkarni, Shailee Jain, Mohammad Javad Hosseini, Jason Baldridge, Eugene Ie, and Li Zhang. 2021. Multi-Level Gazetteer-Free Geocoding. In *Proceedings of Second International Combined Workshop on Spatial Language Understanding and Grounded Communication for Robotics*. 79–88.

- [99] Abhinav Kumar and Jyoti Prakash Singh. 2019. Location reference identification from tweets during emergencies: A deep learning approach. *International journal of disaster risk reduction* 33 (2019), 365–375.
- [100] Abhinav Kumar, Jyoti Prakash Singh, and Nripendra P Rana. 2017. Authenticity of geo-location and place name in tweets. (2017).
- [101] Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. Neural architectures for named entity recognition. *arXiv preprint arXiv:1603.01360* (2016).
- [102] Rabindra Lamsal. 2021. Design and analysis of a large-scale COVID-19 tweets dataset. *Applied Intelligence* 51, 5 (2021), 2790–2804.
- [103] Zhenzhong Lan, Mingda Chen, Sebastian Goodman, Kevin Gimpel, Piyush Sharma, and Radu Soricut. 2019. Albert: A lite bert for self-supervised learning of language representations. *arXiv preprint arXiv:1909.11942* (2019).
- [104] Jochen L Leidner and Michael D Lieberman. 2011. Detecting geographical references in the form of place names and associated spatial natural language. *Sigspatial Special* 3, 2 (2011), 5–11.
- [105] Chenliang Li 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.
- [106] Michael D Lieberman and Hanan Samet. 2011. Multifaceted toponym recognition for streaming news. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*. 843–852.
- [107] Michael D Lieberman and Hanan Samet. 2012. Adaptive context features for toponym resolution in streaming news. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*. 731–740.
- [108] Michael D Lieberman, Hanan Samet, and Jagan Sankaranarayanan. 2010. Geotagging with local lexicons to build indexes for textually-specified spatial data. In *2010 IEEE 26th international conference on data engineering (ICDE 2010)*. IEEE, 201–212.
- [109] Nut Limsoopatham and Nigel Collier. 2016. Bidirectional LSTM for named entity recognition in Twitter messages. *COLING 2016* (2016).
- [110] John Lingad, 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.
- [111] Fei Liu. 2013. Automatic identification of locative expressions from informal text. (2013).
- [112] Fei Liu, Maria Vasardani, and Timothy Baldwin. 2014. Automatic identification of locative expressions from social media text: A comparative analysis. In *Proceedings of the 4th International Workshop on Location and the Web*. 9–16.
- [113] Xiao Liu, Haixiang Guo, Yu-ru Lin, Yijing Li, and Jundong Hou. 2018. Analyzing spatial-temporal distribution of natural hazards in China by mining news sources. *Natural Hazards Review* 19, 3 (2018), 04018006.
- [114] Xinyue Liu, Xiangnan Kong, and Yanhua Li. 2016. Collective traffic prediction with partially observed traffic history using location-based social media. In *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*. 2179–2184.
- [115] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692* (2019).
- [116] Kai Ma, Yongjian Tan, Zhong Xie, Qinjun Qiu, and Siqiong Chen. 2022. Chinese toponym recognition with variant neural structures from social media messages based on BERT methods. *Journal of Geographical Systems* (2022), 1–27.
- [117] Arjun Magge, Davy Weissenbacher, Abeed Sarker, Matthew Scotch, and Graciela Gonzalez-Hernandez. 2018. Deep neural networks and distant supervision for geographic location mention extraction. *Bioinformatics* 34, 13 (2018), i565–i573.
- [118] Shervin Malmasi and Mark Dras. 2015. Location mention detection in tweets and microblogs. In *Conference of the Pacific Association for Computational Linguistics*. Springer, 123–134.
- [119] Christopher D Manning, Mihai Surdeanu, John Bauer, Jenny Rose Finkel, Steven Bethard, and David McClosky. 2014. The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations*. 55–60.
- [120] Huina Mao, Gautam Thakur, Kevin Sparks, Jibonananda Sanyal, and Budhendra Bhaduri. 2018. Mapping near-real-time power outages from social media. *International Journal of Digital Earth* (2018).
- [121] Louis Martin, Benjamin Muller, Pedro Javier Ortiz Suárez, Yoann Dupont, Laurent Romary, Éric Villemonte de La Clergerie, Djamé Seddah, and Benoît Sagot. 2019. CamemBERT: a tasty French language model. *arXiv preprint arXiv:1911.03894* (2019).
- [122] Nicolás José Fernández Martínez and Carlos Perinián-Pascual. 2020. Reglas basadas en conocimiento para la extracción de referencias locativas complejas en tweets. *Rael Revista Electronica de Linguistica Aplicada* 19 (2020), 136–164.
- [123] Nicolás José Fernández Martínez and Carlos Perinián-Pascual. 2020. Knowledge-based rules for the extraction of complex, fine-grained locative references from tweets. *RAEL: revista electrónica de lingüística aplicada* 19, 1 (2020), 136–163.
- [124] Fernando Melo and Bruno Martins. 2017. Automated geocoding of textual documents: A survey of current approaches. *Transactions in GIS* 21, 1 (2017), 3–38.
- [125] Pablo N Mendes, Max Jakob, Andrés García-Silva, and Christian Bizer. 2011. DBpedia spotlight: shedding light on the web of documents. In *Proceedings of the 7th international conference on semantic systems*. 1–8.
- [126] Stuart E Middleton, Giorgos Kordopatis-Zilos, Symeon Papadopoulos, and Yiannis Kompatsiaris. 2018. Location extraction from social media: Geoparsing, location disambiguation, and geotagging. *ACM Transactions on Information Systems (TOIS)* 36, 4 (2018), 1–27.
- [127] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).

- [128] Sveta Milusheva, Robert Marty, Guadalupe Bedoya, Sarah Williams, Elizabeth Resor, and Arianna Legovini. 2021. Applying machine learning and geolocation techniques to social media data (Twitter) to develop a resource for urban planning. *PLoS one* 16, 2 (2021), e0244317.
- [129] Andrei Mircea. 2020. Real-time classification, geolocation and interactive visualization of COVID-19 information shared on social media to better understand global developments. In *Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020*.
- [130] Ludovic Moncla, Mauro Gaio, Thierry Joliveau, and Yves-François Le Lay. 2017. Automated geoparsing of paris street names in 19th century novels. In *Proceedings of the 1st ACM SIGSPATIAL Workshop on Geospatial Humanities*. 1–8.
- [131] Ludovic Moncla, Mauro Gaio, and Sébastien Mustiere. 2014. Automatic itinerary reconstruction from texts. In *International Conference on Geographic Information Science*. Springer, 253–267.
- [132] Ludovic Moncla, Walter Renteria-Agualimpia, Javier Nogueras-Iso, and Mauro Gaio. 2014. Geocoding for texts with fine-grain toponyms: an experiment on a geoparsed hiking descriptions corpus. In *Proceedings of the 22nd ACM SIGSPATIAL international conference on advances in geographic information systems*. 183–192.
- [133] Bruno R Monteiro, Clodoveu A Davis Jr, and Fred Fonseca. 2016. A survey on the geographic scope of textual documents. *Computers & Geosciences* 96 (2016), 23–34.
- [134] Fred Morstatter, Jürgen Pfeffer, Huan Liu, and Kathleen Carley. 2013. Is the sample good enough? comparing data from twitter's streaming api with twitter's firehose. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 7.
- [135] Patricia Murrieta-Flores, Alistair Baron, Ian Gregory, Andrew Hardie, and Paul Rayson. 2015. Automatically analyzing large texts in a GIS environment: The registrar general's reports and cholera in the 19th century. *Transactions in GIS* 19, 2 (2015), 296–320.
- [136] Dat Quoc Nguyen, Thanh Vu, and Anh Tuan Nguyen. 2020. BERTweet: A pre-trained language model for English Tweets. *arXiv preprint arXiv:2005.10200* (2020).
- [137] Malvina Nissim, Colin Matheson, and James Reid. 2004. Recognising geographical entities in Scottish historical documents. In *Proceedings of the Workshop on Geographic Information Retrieval at SIGIR 2004*, Vol. 35. Citeseer.
- [138] Leonardo Nizzoli, Marco Avvenuti, Maurizio Tesconi, and Stefano Cresci. 2020. Geo-semantic-parsing: AI-powered geoparsing by traversing semantic knowledge graphs. *Decision Support Systems* 136 (2020), 113346.
- [139] Jesse O'Shea. 2017. Digital disease detection: A systematic review of event-based internet biosurveillance systems. *International journal of medical informatics* 101 (2017), 15–22.
- [140] Sharon Myrtle Paradesi. 2011. Geotagging tweets using their content. In *Twenty-Fourth International FLAIRS Conference*.
- [141] Kelly S Peterson, Julia Lewis, Olga V Patterson, Alec B Chapman, Daniel W Denhalter, Patricia A Lye, Vanessa W Stevens, Shantini D Gamage, Gary A Roselle, Katherine S Wallace, et al. 2021. Automated Travel History Extraction From Clinical Notes for Informing the Detection of Emergent Infectious Disease Events: Algorithm Development and Validation. *JMIR public health and surveillance* 7, 3 (2021), e26719.
- [142] Bruno Pouliquen, Marco Kimler, Ralf Steinberger, Camelia Ignat, Tamara Oellinger, Ken Blackler, Flavio Fuart, Wajdi Zaghouani, Anna Widiger, Ann-Charlotte Forslund, et al. 2006. Geocoding multilingual texts: Recognition, disambiguation and visualisation. *arXiv preprint cs/0609065* (2006).
- [143] Bruno Pouliquen, Ralf Steinberger, Camelia Ignat, and Tom De Groeve. 2004. Geographical information recognition and visualization in texts written in various languages. In *Proceedings of the 2004 ACM symposium on Applied computing*. 1051–1058.
- [144] Ross S Purves, Paul Clough, Christopher B Jones, Avi Arampatzis, Benedicte Bucher, David Finch, Gaihua Fu, Hideo Joho, Awase Khirni Syed, Subodh Vaid, et al. 2007. The design and implementation of SPIRIT: a spatially aware search engine for information retrieval on the Internet. *International journal of geographical information science* 21, 7 (2007), 717–745.
- [145] Ross S Purves, 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 (2018), 164–318.
- [146] Umair Qazi, Muhammad Imran, and Ferda Ofli. 2020. GeoCoV19: a dataset of hundreds of millions of multilingual COVID-19 tweets with location information. *SIGSPATIAL Special* 12, 1 (2020), 6–15.
- [147] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D Manning. 2020. Stanza: A python natural language processing toolkit for many human languages. *arXiv preprint arXiv:2003.07082* (2020).
- [148] Qinjun Qiu, Zhong Xie, Shu Wang, Yunqiang Zhu, Hairong Lv, and Kai Sun. [n. d.]. ChineseTR: A weakly supervised toponym recognition architecture based on automatic training data generator and deep neural network. *Transactions in GIS* ([n. d.]).
- [149] F Rahma and A Romadhony. 2021. Rule-Based Crime Information Extraction on Indonesian Digital News. In *2021 International Conference on Data Science and Its Applications (ICoDSA)*. IEEE, 10–15.
- [150] Lev Ratnoff and Dan Roth. 2009. Design challenges and misconceptions in named entity recognition. In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*. 147–155.
- [151] Alan Ritter, Sam Clark, Oren Etzioni, et al. 2011. Named entity recognition in tweets: an experimental study. In *Proceedings of the 2011 conference on empirical methods in natural language processing*. 1524–1534.
- [152] Giuseppe Rizzo, Amparo Elizabeth Cano Basave, Bianca Pereira, Andrea Varga, Matthew Rowe, Milan Stankovic, and A Dadzie. 2015. Making Sense of Microposts (# Microposts2015) Named Entity rEcognition and Linking (NEEL) Challenge.. In # MSM. 44–53.
- [153] CJ Rupp, Paul Rayson, Alistair Baron, Christopher Donaldson, Ian Gregory, Andrew Hardie, and Patricia Murrieta-Flores. 2013. Customising geoparsing and georeferencing for historical texts. In *2013 IEEE International Conference on Big Data*. IEEE, 59–62.
- [154] Meryem Sagcan and Pinar Karagoz. 2015. Toponym recognition in social media for estimating the location of events. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*. IEEE, 33–39.

- [155] SPCW Sandagiri, BTGS Kumara, and Banujan Kuhaneswaran. 2020. Detecting Crimes Related Twitter Posts Using SVM based Two Stages Filtering. In *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*. IEEE, 506–510.
- [156] Christopher Scheele, Manzhu Yu, and Qunying Huang. 2021. Geographic context-aware text mining: enhance social media message classification for situational awareness by integrating spatial and temporal features. *International Journal of Digital Earth* 14, 11 (2021), 1721–1743.
- [157] Stefan Schweter and Alan Akbik. 2020. FLERT: Document-Level Features for Named Entity Recognition. arXiv:2011.06993 [cs.CL]
- [158] Peter Scott, Martin K-F Bader, Treena Burgess, Giles Hardy, and Nari Williams. 2019. Global biogeography and invasion risk of the plant pathogen genus Phytophthora. *Environmental Science & Policy* 101 (2019), 175–182.
- [159] Jiang Shang, Xuke Hu, Fuqiang Gu, Di Wang, and Shengsheng Yu. 2015. Improvement schemes for indoor mobile location estimation: A survey. *Mathematical Problems in Engineering* 2015 (2015).
- [160] Lanyu Shang, Yang Zhang, Christina Youn, and Dong Wang. 2022. SAT-Geo: A social sensing based content-only approach to geolocating abnormal traffic events using syntax-based probabilistic learning. *Information Processing & Management* 59, 2 (2022), 102807.
- [161] Eric Shook and Victoria K Turner. 2016. The socio-environmental data explorer (SEDE): a social media-enhanced decision support system to explore risk perception to hazard events. *Cartography and Geographic Information Science* 43, 5 (2016), 427–441.
- [162] Mário J Silva, Bruno Martins, Marcirio Chaves, Ana Paula Afonso, and Nuno Cardoso. 2006. Adding geographic scopes to web resources. *Computers, Environment and Urban Systems* 30, 4 (2006), 378–399.
- [163] Jyoti Prakash Singh, Yogesh K Dwivedi, Nripendra P Rana, Abhinav Kumar, and Kawaljeet Kaur Kapoor. 2019. Event classification and location prediction from tweets during disasters. *Annals of Operations Research* 283, 1 (2019), 737–757.
- [164] N Sobhana, Pabitra Mitra, and SK Ghosh. 2010. Conditional random field based named entity recognition in geological text. *International Journal of Computer Applications* 1, 3 (2010), 143–147.
- [165] K Srinivasa and P Santhi Thilagam. 2019. Crime base: Towards building a knowledge base for crime entities and their relationships from online news papers. *Information Processing & Management* 56, 6 (2019), 102059.
- [166] Benjamin Strauss, Bethany Toma, Alan Ritter, Marie-Catherine De Marneffe, and Wei Xu. 2016. Results of the wnut16 named entity recognition shared task. In *Proceedings of the 2nd Workshop on Noisy User-generated Text (WNUT)*. 138–144.
- [167] Nestor Suat-Rojas, Camilo Gutierrez-Osorio, and Cesar Pedraza. 2022. Extraction and Analysis of Social Networks Data to Detect Traffic Accidents. *Information* 13, 1 (2022), 26.
- [168] Evan A Sultanik and Clayton Fink. 2012. Rapid geotagging and disambiguation of social media text via an indexed gazetteer.. In *ISCRAM*.
- [169] Javier Tamames and Victor de Lorenzo. 2010. EnvMine: A text-mining system for the automatic extraction of contextual information. *BMC bioinformatics* 11, 1 (2010), 1–10.
- [170] Andrea H Tapia, Kathleen A Moore, and Nichloas J Johnson. 2013. Beyond the trustworthy tweet: A deeper understanding of microblogged data use by disaster response and humanitarian relief organizations.. In *ISCRAM*.
- [171] Laura Tateosian, Rachael Guenter, Yi-Peng Yang, and Jean Ristaino. 2017. Tracking 19th century late blight from archival documents using text analytics and geoparsing. In *Free and open source software for geospatial (FOSS4G) conference proceedings*, Vol. 17. 17.
- [172] Benjamin E Teitler, Michael D Lieberman, Daniela Panozzo, Jagan Sankaranarayanan, Hanan Samet, and Jon Sperling. 2008. NewsStand: A new view on news. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*. 1–10.
- [173] Sayan Unarkard, Xue Li, and Mohamed A Sharaf. 2015. Emerging event detection in social networks with location sensitivity. *World Wide Web* 18, 5 (2015), 1393–1417.
- [174] Asahi Ushio and Jose Camacho-Collados. 2021. T-NER: An All-Round Python Library for Transformer-based Named Entity Recognition. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations*. 53–62.
- [175] Maria Vasardani, Stephan Winter, and Kai-Florian Richter. 2013. Locating place names from place descriptions. *International Journal of Geographical Information Science* 27, 12 (2013), 2509–2532.
- [176] Lara Vomfell, Wolfgang Karl Härdle, and Stefan Lessmann. 2018. Improving crime count forecasts using Twitter and taxi data. *Decision Support Systems* 113 (2018), 73–85.
- [177] Jan Oliver Wallgrün, Morteza Karimzadeh, Alan M MacEachren, and Scott Pezanowski. 2018. GeoCorpora: building a corpus to test and train microblog geoparsers. *International Journal of Geographical Information Science* 32, 1 (2018), 1–29.
- [178] Jimin Wang and Yingjie Hu. 2019. Are we there yet? Evaluating state-of-the-art neural network based geoparsers using EUPEG as a benchmarking platform. In *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Geospatial Humanities*. 1–6.
- [179] Jimin Wang and Yingjie Hu. 2019. Enhancing spatial and textual analysis with EUPEG: an extensible and unified platform for evaluating geoparsers. *Transactions in GIS* 23, 6 (2019), 1393–1419.
- [180] Jimin Wang, Yingjie Hu, and Kenneth Joseph. 2020. NeuroTPR: A neuro-net toponym recognition model for extracting locations from social media messages. *Transactions in GIS* (2020).
- [181] Ruo-Qian Wang, Yingjie Hu, Zikai Zhou, and Kevin Yang. 2020. Tracking flooding phase transitions and establishing a passive hotline with ai-enabled social media data. *IEEE Access* 8 (2020), 103395–103404.
- [182] Wei Wang and Kathleen Stewart. 2015. Spatiotemporal and semantic information extraction from Web news reports about natural hazards. *Computers, environment and urban systems* 50 (2015), 30–40.
- [183] Barney Warf and Santa Arias. 2008. *The spatial turn: Interdisciplinary perspectives*. Routledge.

- [184] Davy Weissenbacher, Arjun Magge, Karen O'Connor, Matthew Scotch, and Graciela Gonzalez. 2019. Semeval-2019 task 12: Toponym resolution in scientific papers. In *Proceedings of the 13th International Workshop on Semantic Evaluation*. 907–916.
- [185] Davy Weissenbacher, Abeed Sarker, Tasnia Tahsin, Matthew Scotch, and Graciela Gonzalez. 2017. Extracting geographic locations from the literature for virus phylogeography using supervised and distant supervision methods. *AMIA Summits on Translational Science Proceedings 2017* (2017), 114.
- [186] Davy Weissenbacher, Tasnia Tahsin, Rachel Beard, Mari Figaro, Robert Rivera, Matthew Scotch, and Graciela Gonzalez. 2015. Knowledge-driven geospatial location resolution for phylogeographic models of virus migration. *Bioinformatics* 31, 12 (2015), i348–i356.
- [187] Stephan Winter, Kai-Florian Richter, Tim Baldwin, Lawrence Cavedon, Lesley Stirling, Matt Duckham, Allison Kealy, and Abbas Rajabifard. 2011. Location-based mobile games for spatial knowledge acquisition. (2011).
- [188] Miguel Won, Patricia Murrieta-Flores, and Bruno Martins. 2018. ensemble named entity recognition (ner): evaluating ner Tools in the identification of Place names in historical corpora. *Frontiers in Digital Humanities* 5 (2018), 2.
- [189] Allison Gyle Woodruff and Christian Plaut. 1994. GIPSY: Automated geographic indexing of text documents. *Journal of the American Society for Information Science* 45, 9 (1994), 645–655.
- [190] Desheng Wu and Yiwen Cui. 2018. Disaster early warning and damage assessment analysis using social media data and geo-location information. *Decision support systems* 111 (2018), 48–59.
- [191] Canwen Xu, Jing Li, Xiangyang Luo, Jiaxin Pei, Chenliang Li, and Donghong Ji. 2019. DLocRL: A deep learning pipeline for fine-grained location recognition and linking in tweets. In *The World Wide Web Conference*. 3391–3397.
- [192] MM Yagoub, Aishah A Alseredi, Elfadil A Mohamed, Punitha Periyasamy, Reem Alameri, Salama Aldarmaki, and Yaqein Alhashmi. 2020. Newspapers as a validation proxy for GIS modeling in Fujairah, United Arab Emirates: identifying flood-prone areas. *Natural Hazards* 104, 1 (2020), 111–141.
- [193] Pranali Yenkar and SD Sawarkar. 2021. Gazetteer based unsupervised learning approach for location extraction from complaint tweets. In *IOP Conference Series: Materials Science and Engineering*. Vol. 1049. IOP Publishing, 012009.
- [194] Hua Yuan, Hualin Xu, Yu Qian, and Yan Li. 2016. Make your travel smarter: Summarizing urban tourism information from massive blog data. *International Journal of Information Management* 36, 6 (2016), 1306–1319.
- [195] Yongjian Zhu, Liqing Cao, Jingui Xie, Yugang Yu, Anfan Chen, and Fengming Huang. 2021. Using social media data to assess the impact of COVID-19 on mental health in China. *Psychological medicine* (2021), 1–8.
- [196] Lei Zou, Danqing Liao, Nina SN Lam, Michelle Meyer, Nasir G Gharaibeh, Heng Cai, Bing Zhou, and Dongying Li. 2021. Social Media for Emergency Rescue: An Analysis of Rescue Requests on Twitter during Hurricane Harvey. *arXiv preprint arXiv:2111.07187* (2021).