Workshop Proceedings of the 19th International Conference on Intelligent Environments (IE2023)
G. Bekaroo et al. (Eds.)

doi:10.3233/AISE230011

A Knowledge Graph-Based Method for the Geolocation of Tweets

Fernando LOVERA [a,1], Yudith CARDINALE [a,b], Davide BUSCALDI [c], and Thierry CHARNOIS [c]

[a] *Universidad Simón Bolívar, Caracas, Venezuela*
[b] *Grupo de Investigación en Ciencia de Datos (GRID), Universidad Internacional de Valencia, Spain*
[c] *Northern Paris Computer Science Lab, LIPN, Paris France*
ORCiD ID: Fernando Lovera https://orcid.org/0000-0002-3042-5953, Yudith Cardinale https://orcid.org/0000-0002-5966-0113, Davide Buscaldi https://orcid.org/0000-0003-1112-3789, Thierry Charnois https://orcid.org/0000-0001-9700-5075

Abstract. Twitter geolocation is useful for various purposes, including tracking COVID-19 perceptions, analyzing political trends, and managing natural disasters. However, accurately predicting geolocations based on tweet content remains a challenge. In the past, machine learning approaches have tried to solve this problem by training prediction models on previously seen data, but these models often struggle to generalize to unseen places. To overcome these limitations, in this work we present a framework based on Natural Language Processing (NLP), Knowledge Graphs (KG), and Semantic Web to find geographical entities on tweets' content. KG facilitate the extraction of structured knowledge of texts in order to study their semantic analysis based on NLP techniques to search associated geographical coordinates to the entities of that KG; if there is explicit mention of places in the tweet, the Semantic Web is used to find geographical information associated with the entities present in the tweets' content. To evaluate the precision of the prediction algorithm, we compare our predicted latitude and longitude coordinates with AlbertaT6 floods dataset. Our results show an $F1$ score up to 0.851 within a 10 kilometer radius.

Keywords. Knowledge Graph, Geolocation, Ontologies, Natural Disasters, Twitter Analysis

1. Introduction

Twitter has been used in a variety of studies, such as Sentiment Analysis [1,2], examining political potential and trends [3], location based recommendation systems [4], advertising [5], demographic analysis [6], monitoring natural disasters [7]. Geolocation is particularly useful for applications related to natural disasters and crisis detection, as it allows for the identification of regional behavior and the provision of targeted assistance [8,9,10].

[1] Corresponding Author: Lovera Fernando, flovera@usb.ve

Geolocation in this context refers to the geographical identification of places, for example, the location of users or entities on texts. Geolocation is a growing topic in the development of smart cities [11], and it is essential for many new applications in Twitter, such as recommendation systems and disaster management [9]. In particular, in the case of crisis scenarios, such as identifying a request for help, it is important not only to know that someone is in danger, but also where these people are located. Nevertheless, geolocation still remains as an unfinished problem, since it is difficult to make precise geolocation predictions based on tweets' content [12]. Identifying the location of tweets is a difficult task, since coordinates are rarely available: according to Twitter[2], only 1-2% of the tweets are geographically tagged. This is due to the fact that people deactivate the localization functions in their devices both to enhance battery duration and to preserve their privacy.

Developing systems that are able to identify the origin of a tweet from text has been the focus of recent works. In particular, the shared task[3] at the 2nd workshop on noisy user-generated text (WNUT) was focused on twitter geolocation prediction [13]. It saw the participation of five teams with methods ranging from multinomial naive Bayes to neural networks, using training data collected from 1 million users; the best method obtained 0.409 in classification accuracy, and median and mean distances of 69.5 and 1,792.5 Kms, respectively [14]. Thomas and Hennig use a neural network based on Long Short-Term Memory (LSTM), which on the same dataset obtained 0.423 accuracy, 51.1 and 1,400.4 Kms in mean and average distances [15].

Identifying the geolocation of tweets is a problem that can leverage Linked Data and the Semantic Web [16]. The challenges of geolocation under this perspective are quite different. In general, there are ontologies that contain geographical data of almost every place on earth; this information is created as a complement to other types of information. For example, in DBpedia ontology is easy to find names of places and also their locations [17], in OpenStreetMap (a user-generated dataset of the streets), there is information about specific streets [18]. This information is also accessed through SPARQL[4] or through libraries in different programming languages[5]. Under this perspective, the main challenge is to understand what to ask to these ontologies in order to retrieve the appropriate information (geographical coordinates). It is important to mention that not being able to retrieve the information from the ontology is perfectly a valid case; this does not happen under the Machine Learning perspective, where the intelligent model always returns a value.

In this work, we propose a singular framework based on different technologies, such as Knowledge Graphs (KG), Natural Language Processing (NLP), parsing trees to generate predictions on geographical coordinates (latitude, longitude) based on tweets' contents, and Semantic Web. The starting point of the analysis is whether the text contains geographical entities or not, if it contains geographical entities, they are retrieved from Nominatim ontology; if not, then a KG is created that helps to associate the text with a known place in the DBpedia ontology.

The proposed framework is able to handle cases where no coordinates can be predicted, which is a significant difference from machine learning approaches that always

[2] https://developer.twitter.com/en/docs/tutorials/tweet-geo-metadata
[3] https://noisy-text.github.io/2016/geo-shared-task.html
[4] https://www.w3.org/TR/rdf-sparql-query/
[5] https://pypi.org/project/OSMPythonTools/

return a value. Instead of predicting the origin of a tweet, the focus of the framework is to identify the relevance of a tweet to a specific area. This means that the prediction may be affected if people in one place are discussing another place. We compare our framework to two geotagging tools as baselines for our metrics and find that our approach performs significantly better, as demonstrated by our $F1$, precision, and recall scores. We evaluate our method using the AlbertaT6 floods dataset, which achieves a precision of up to 0.928 within a 10 kilometer radius.

2. Related work

Literature shows many studies that approach the geolocation problem under the natural disasters emergencies perspective. Due to the features of the problem, geolocation in this case could have the impact of saving a life. Reynard and Shirgaokar use Machine Learning and geospatial techniques to geolocate tweets, which were both about Hurricane Irma and located within Florida [19]. Authors perform a comprehensive analysis over their tweets, including sentiment analysis, they relied on a multi-nominal logistic regression specification to study which features of the tweet, user, or location were likely to be associated with negative or positive sentiments. Between the contributions of the study, we found a suggestion to develop a probability-based understanding of needs to guide disaster managers and a proposition with a clean methodology for extracting embedded information within social media data. Ebrahimi et al. create an exhaustive hybrid neural network, by using the tweets' text, the network of the user (its connection with other users), and tweets' metadata (user name, user description, user-declared location, timezone, user language, tweet creation time, user UTC offset) as sources of information [20]. These fields are then processed by a different sub-network to generate a feature vector representation R_j, then these feature vectors are joined to build a final user representation $\hat{R}$, which is fed into a linear classification layer, achieving an accuracy of 70.8% in their predictions. Auclair et al. classify tweets associated to a natural disaster [21], they perform unsupervised classification to extract the thematic information and then introduce geolocation through the use of Named-Entity Recognition (NER). On Table 1, we show a comparison of these recent studies regarding natural disasters.

We have observed that supervised approaches have the tendency of using Machine Learning algorithms combined with bag-of-words. The study described in [22] presents a hierarchy of logistic regression classifiers model trained on Twitter, Wikipedia, and Flickr data. Han et al. suggest a similar approach [23], using a neural network with a denoising autoencoder, while Rahimi et al. use a multi-level regularization and a multi-level perceptron architecture [24]. In the case of unsupervised learning approaches, we found interesting studies that address the geolocation problem where the algorithm must first self-discover what patterns are in the text and then based on those patterns form groups, that would help to make comparisons and generate predictions. Eisenstein et al. present a multi-level generative model that is able to reason about geographical regions and latent topics [25]. The model uses the K-Nearest Neighbors (KNN) approach to generate clusters of information in the United States of America based on topics such as: popular music, emoticons on the text, and chit chat. Cha et al. geolocate users by identifying features extracted from their social media texts, by a two-step procedure that consists of an upconversion and iterative refinement by joint sparse coding [26]. Sanjar et al.

Table 1. Geolocation in natural disasters

Work	Model	Dataset	Result
Eisenstein, J. et al.[20]	Neural network	TWITTERUS and WNUT	70.8% accuracy
Auclair, S. et al. [21]	SVM	Project SOS [30]	88% $F1$ score
Han, B. et al. [23]	Neural Network		43.7% accuracy
Rahimi, A. et al. [24]	Perceptron	GEOTEXT and TWITTER-US	50.2% accuracy
Cha, M. et al. [26]	Sparse coding	GEOTEXT	error of 568 km
Sanjar, K. et al. [27]	KNN	House Prices Kaggle	error of 630 km
Lieberman, M.D. et al. [28]	ML and Gazetteer	ACE 2005 English SpatialML	0.787 $F1$ score
Mourad, A. et al. [29]	Text-based	CSIRO Data61 at the WNUT	52.9% accuracy

suggest a KNN-based most correlated features (KNN–MCF) algorithm to use geolocation to predict house's pricing [27]. Although these studies show great results, they have the tendency not to scale to larger datasets. Lieberman et al. show another great popular learning technique used in geolocation [28]: the semi-supervised learning. In this study, a geographical dictionary is built due to references for information about places and place names used in conjunction with an atlas and a map. It contains information regarding the geographical region of a country or continent. It also includes physical features, like mountains, roads and finally social statistics. Mourad et al. provide a practical guide of Twitter user geolocation [29]; they demonstrate that the choice of effectiveness metric has a crucial consequences on the conclusions given by a geolocation system experiment, and conclude that the evaluation of geolocation models should be performed on datasets with different characteristics or domains to warranty their consistent performance.

In this paper we leverage Semantic Web technologies, such as KGs and ontologies to analyse the tweet's text to geolocalise its content. The study described in [28] presents quite interesting results regarding to the geolocation and achieving quite high accuracy. We investigate the semantics of the text before the actually querying the onthologies.

3. Our proposal

In this section, we describe a framework for Twitter to geolocate tweets based on KG. The proposed framework is divided in three main modules, as shown in Figure 1: (i) *Natural Language Processing* dedicated to perform the cleaning of the data; (ii) *Identification of geographical entities* mainly involves Named-Entity Recognition (NER), which is a task that aims to classify or locate named entities mentioned in the text into categories, such as organizations, locations, person names [31]; with these entities the KG is created; and (iii) *Knowledge graph & ontologies* more oriented to find the values of geographical coordinates and the creation of queries that can accomplish this.

3.1. First stage: Natural Language Processing

Normally, pre-processing is a step required since raw data tends to need the process of recognition and deletion of inaccurate or corrupt data, in this case text data. The ex-

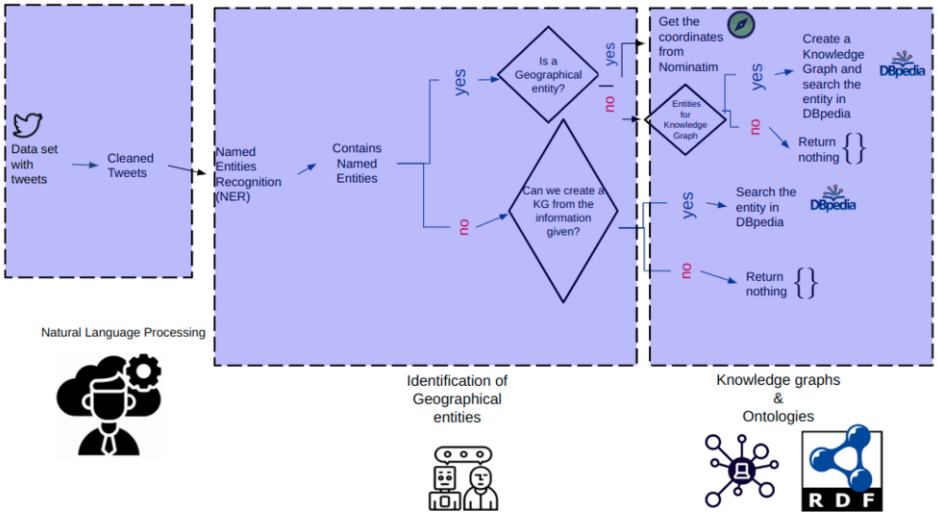

Figure 1. Stages of proposal: Natural Language Processing; Identification of Geographical entities; Knowledge graphs & ontologies.

traction and discovery of the knowledge hidden in unstructured text is of great importance [32]. Early identification of noise and proper data cleaning (special treatment) is necessary, so that further analysis can be performed.

Each step within the pre-processing of data is as follows: (i) *Elimination of characters* that do not help in the prediction process; (ii) *Transform* the full text to lowercase [33]; (iii) *Expansion* of contractions in English; (iv) *Compression* of repeated contiguous letters; (v) *Elimination* of long words; (vi) *Correction* of spelling errors; (vii) *Elimination* of stopwords; (viii) *Elimination* of repeated words; (ix) *Elimination* of null entries (tweets); (x) Lemmatization.

3.2. Second stage: Identification of geographical entities

Once the cleaned tweets are obtained, they are ready for the analysis to investigate the tweets' geolocation. As a pre-condition for this stage, we need to deal with ready-to-use data. We assume that the tweets are already cleaned and ready to be studied further.

Basically, for identifying the geographical entities, we apply NER techniques, based on NLP, dealing with classifying entities from raw text. This classification could be into organizations, entities, places, time, or even money.

We use the library *Spacy* in the implementation of the NER of our proposal. It has built-in methods for NER, which allows identifying easily if a tweet contains geographical entities or not. *Spacy* is used to detect geographical entities in the following way: first, all the entities are recognized; then, it is identified if these entities are geographical or not (non-geographical entities are filtered out) depending on how they were catalog. Once the entity is identified, a query to the Nominatim is performed, in order to return the geographical coordinates associated to the entities. The Nominatim Geocoder is publicly available[6]; therefore, it can be queried through libraries; nonetheless, it is also possible to create an own geocoder.

[6]https://github.com/mocnik-science/osm-python-tools/blob/master/docs/nominatim.md

3.3. Third stage: Knowledge graphs & Ontologies.

Geographical entities are not always present in the text; when this happens, we need to find another clue that could give us an estimate in the search of geographical coordinates. In our proposal, we build KG to try to build an intuition of the places present in the text (if any) and later ask for coordinates to the DBpedia knowledge base. The purpose of the KG is to link the tweet's content with some other entities, which we can locate.

Once the KG is created, we only need to query DBpedia with the information from the entities of such KG. This can be easily done through the use of SPARQL endpoints. Hence, with very little effort we can retrieve geographical information from the semantic database DBpedia, if we make the right query to it. If we query DBpedia, we can retrieve geographical information from the semantic database.

Sometimes we cannot form any kind of semantic representation of the tweets' content, these are rare cases, but they can still exists. When this happens, we try to study the metadata associated with the tweet id (also present in the dataset). The location mentioned in the tweet's text is not necessarily the same as the one mentioned in the metadata. This is a guess based on common sense, but it could be flawed (metadata could not reflect the reality of geolocation of the tweet); therefore, checking the tweet's metadata is the last resource of information for the proposed framework to try to find the coordinates. When we reach this stage, we associate the geolocation with the coordinates present in the metadata. To extract the information from the metadata, we use Twitter API by taking the tweet's ids as inputs.

At these stages, the metadata could not be available, thus the framework is unable to find an answer and make a precise guess of what the geolocations could be. In these cases, the proposed framework does not return any value, offering a more honest predictor; it does not always return a value regardless of what the input is, contrary to Machine Learning models that always return a value.

4. Evaluation and results

To evaluate the performance of the proposed framework, we use AlbetaT6 floods dataset and compare the results with two baselines in terms of $F1$ score, *precision*, and *recall*. In the following, we first describe the dataset used, afterwards we show the results after applying the three stages of the model proposed and the two baselines, to finally discuss the results.

4.1. Data Exploration and Pre-Processing

To evaluate the performance of the framework, we collect a dataset, called AlbertaT6, available in CrisisLex[7]. CrisisLex is a public repository of crisis-related social media data and tools. AlbertaT6 dataset is related to the Alberta flooding. The government of Canada lists the Alberta flood as the worst misfortune to have occurred in this country [34]. This dataset is a subset of the dataset CrisisLexT6, restricted to tweets that have geolocation. To create the dataset, we first took all Alberta's tweets from CrisisLexT6, and then with the Twitter API, their geolocation with GPS coordinates was verified. This

[7] https://crisislex.org/

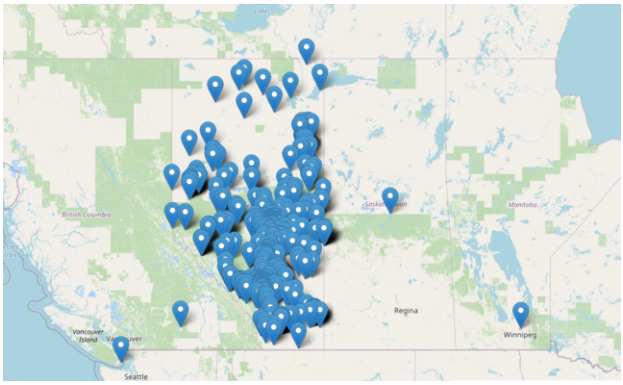

Figure 2. Location points contained in the Alberta dataset.

dataset contains 2152 tweets in total with information related to the tweet id (this is part of the tweet's metadata) and the tweet's content.

We plot the tweets collected in AlbertaT6 dataset in a map, shown in Figure 2, since we know the label of tweets' geolocation, so it is easy to recognize where this natural disaster occurred.

To clean the dataset, we went through all the steps required to pre-process the data as the first stage in the proposed framework. After the pre-processing of the data, we obtain 2151 tweets for the AlbertaT6 dataset.

4.2. Comparative Evaluation

After the cleaning process, we submitted the dataset to the second and third stages of the framework to predict their tweets' geolocations.

We choose two baselines since they represent the state-of-the-art in geolocation prediction and they have been used and tested for many researchers. As first baseline, we compared our results of predictions with the Pigeo project, which is dedicated to geolocation prediction [35]. Pigeo is a Python Geotagging tool that is used as a geolocation service based on pre-trained regression models. As the second baseline, we test the model Erechtheus, publicly available[8]. To study the behavior of our proposal, we computed the $F1$ score, the *precision*, and the *recall*, for the predictions, including a margin of error of 1, 5, and 10 KM, as shown in Table 2. The predictions of our framework improve as the radio of precision increases, obtaining up to 0.851 of $F1$ score. Table 3 contains the $F1$ score, the *precision*, and the *recall* metrics of the baselines Erectheus and Pigeo, which do not consider a radio of errors in KM, therefore we do not have the comparison with 1, 5, and 10 KM, respectively. Nonetheless, if we compare them with the worst results obtained from our framework with 1 KM of distance in radio (see Table 2), we observe that our proposal behaves with more accurate results. Our framework obtains an $F1$ score of 0.122 of precision that overcomes the $F1$ score of 0.003 obtained by Erectheus (see Table 3).

The tool Pigeo performed similarly to our framework obtaining an $F1$ score of 0.146 against the $F1$ score of the proposed framework of 0.122 (see Table 3 and Table 2,

[8]https://github.com/Erechtheus/geolocation

Table 2. *F*1 score, *precision*, and *recall* of our proposal within 1, 5, and 10 KM.

KM	*F*1 score	Precision	Recall
1KM	0.122	0.137	0.116
5 KM	0.712	0.776	0.658
10 KM	0.851	0.928	0.787

Table 3. Results of two baselines considered state-of-the-art

Model	*F*1 score	Precision	Recall
Erectheus	0.003	0.007	0.002
Pigeo	0.146	0.110	0.226

Table 4. Results of meta-information about our framework.

Model	Queries to DBpedia	Queries to Nominatim	Unknowns
Our framework	1649	175	327

respectively); nonetheless, our framework behaves better for the rest of the test cases. The tool Erectheus seems to fail to generalize towards unknown places since it behaved poorly for all the scenarios. We can also conclude that the comparison between these tools and the results of 1 KM of distance in radio from the framework reflects that it is flexible and it can try to make predictions even for unknown places; also, the framework recognizes when it is an unknown geolocation.

Table 4 shows the amount of queries that were performed by our framework towards DBpedia and Nominatim. We also count the number of unknown results that could came from these queries, this means that our framework could return any value. Our framework have 15.202% of tweets that could not locate (unknowns).

In many of these cases Erechtheus and Pigeo returned wrong answers, because they do not look beforehand if the text can be geolocalised or not, and this affect negatively their scores. Concerning the use of ontologies, our framework took most of the coordinates out of DBpedia, reaching more than 70% of the information retrieved from this ontology.

These results were obtained thanks to the use of NER that could properly identify entities, and also the use of powerful data structures such as KG. The importance of KG as data structure is its potential to search information about a variety of topics, in this case geolocations. The Semantic Web has been constructed to represent information in a more structured manner and KG explode this. As we can see, we can make queries to the Semantic Web using engines, such as SPARQL or Virtuoso, among others, and immediately receive benefits from them.

5. Conclusions

In this work, we propose a framework to predict geographical coordinates from tweets' content. The framework consists of three separated stages with well-defined tasks that combine NER and NLP techniques with Knowledge Graphs (KG) and modern analysis over Semantic Web. The first stage is based on NLP techniques to pre-process the text and clean data. The second stage takes the outputs of the first stage and builds semantic

representations of the inputs given (tweets' text) to query ontologies in the following stage; this is achieved through the use of NER and KG. The last stage's task is to retrieve geographical information found in Semantic Ontologies, DBpedia, and Nominatim. Particularly, the NER and KG operate well with complex Semantic Web structures like DBpedia or with Nominatim. The proposed framework follows clues of the tweet passed as input, it exhaustes all the possible ways to deduce a geographical coordinate, in case of not finding a logical deduction, it simply does not return anything. We compare this mechanism with two proven geolocation tools and show the strengths and advantages of using this mechanism. We show great improvement in the precision of the model introducing KG and the Semantic Web to our solution.

We are working on testing the framework in real scenarios to demonstrate how these technologies and data structures can have a great impact on the problem of geolocation in the case of natural disasters. Future research will be carried out in the use of the KG as a fundamental data structure in a framework that works with short texts in order to make geolocation predictions. Furthermore, we plan to perform experiments with datasets with tweets written in other languages, such as France or Spanish) to measure the language impact on the measurements of $F1$ **score**, *precision*, and *recall*.

References

[1] Ghosh S, Hazra A, Raj A. A comparative study of different classification techniques for sentiment analysis. In: Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines. IGI Global; 2022. p. 174-83.

[2] Mohammad SM, Kiritchenko S, Zhu X. NRC-Canada: Building the state-of-the-art in sentiment analysis of tweets. arXiv preprint arXiv:13086242. 2013.

[3] Meeks L. Tweeted, deleted: theoretical, methodological, and ethical considerations for examining politicians' deleted tweets. Information, Communication & Society. 2018;21(1):1-13.

[4] Katarya R, Arora Y. A survey of recommendation systems in twitter. In: Internat. Conf. on Computational Intelligence & Communication Technology. IEEE; 2018. p. 1-5.

[5] Boerman SC, Kruikemeier S. Consumer responses to promoted tweets sent by brands and political parties. Computers in human behavior. 2016;65:285-94.

[6] Sloan L, Morgan J, Housley W, Williams M, Edwards A, Burnap P, et al. Knowing the Tweeters: Deriving sociologically relevant demographics from Twitter. Sociological research online. 2013;18(3):74-84.

[7] Pourebrahim N, Sultana S, Edwards J, Gochanour A, Mohanty S. Understanding communication dynamics on Twitter during natural disasters: A case study of Hurricane Sandy. Internat journal of disaster risk reduction. 2019;37:101176.

[8] Bruns A, Liang YE. Tools and methods for capturing Twitter data during natural disasters. First Monday. 2012.

[9] Scalia G, Francalanci C, Pernici B. CIME: Context-aware geolocation of emergency-related posts. GeoInformatica. 2022;26(1):125-57.

[10] Vieweg S, Hughes AL, Starbird K, Palen L. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In: SIGCHI conference on human factors in computing systems; 2010. p. 1079-88.

[11] Khatoon S, Asif A, Hasan MM, Alshamari M. In: Pardalos PM, Rassia ST, Tsokas A, editors. Social Media-Based Intelligence for Disaster Response and Management in Smart Cities; 2022. p. 211-35.

[12] Witono T, Yazid S. A Review of Internet Topology Research at the Autonomous System Level. In: Yang XS, Sherratt S, Dey N, Joshi A, editors. Internat. Congress on Information and Communication Technology; 2022. p. 581-98.

[13] Han B, Rahimi A, Derczynski L, Baldwin T. Twitter geolocation prediction shared task of the 2016 workshop on noisy user-generated text. In: Workshop on Noisy User-generated Text; 2016. p. 213-7.

[14] Miura Y, Taniguchi M, Taniguchi T, Ohkuma T. A simple scalable neural networks based model for geolocation prediction in Twitter. In: Workshop on Noisy User-generated Text; 2016. p. 235-9.

[15] Thomas P, Hennig L. Twitter geolocation prediction using neural networks. In: Internat. Conf. of the German Society for Computational Linguistics and Language Technology. Springer; 2017. p. 248-55.
[16] Cordeiro KdF, Marino T, Campos MLM, Borges MRS. Use of Linked Data in the design of information infrastructure for collaborative emergency management system. In: Internat. Conf. on Computer Supported Cooperative Work in Design; 2011. p. 764-71.
[17] Hausenblas M. Exploiting linked data to build web applications. IEEE Internet Computing. 2009;13(4):68-73.
[18] Bennett J. OpenStreetMap. Packt Publishing Ltd https://www.packtpub.com/product/openstreetmap/9781847197504; 2010.
[19] Reynard D, Shirgaokar M. Harnessing the power of machine learning: Can Twitter data be useful in guiding resource allocation decisions during a natural disaster? Transportation research part D: Transport and environment. 2019;77:449-63.
[20] Ebrahimi M, ShafieiBavani E, Wong R, Chen F. A Unified Neural Network Model for Geolocating Twitter Users. In: Conf. on Computational Natural Language Learning. Brussels, Belgium: Association for Computational Linguistics; 2018. p. 42-53.
[21] Auclair S, Boulahya F, Birregah B, Quique R, Ouaret R, Soulier E. SURICATE-Nat: Innovative citizen centered platform for Twitter based natural disaster monitoring. In: Internat. Conf. on Information and Communication Technologies for Disaster Management. IEEE; 2019. p. 1-8.
[22] Miyazaki T, Rahimi A, Cohn T, Baldwin T. Twitter geolocation using knowledge-based methods. In: EMNLP Workshop W-NUT: The 4th Workshop on Noisy User-generated Text; 2018. p. 7-16.
[23] Han B, Cook P, Baldwin T. Geolocation prediction in social media data by finding location indicative words. In: COLING; 2012. p. 1045-62.
[24] Rahimi A, Vu D, Cohn T, Baldwin T. Exploiting text and network context for geolocation of social media users. arXiv preprint arXiv:150604803. 2015.
[25] Eisenstein J, O'Connor B, Smith NA, Xing E. A latent variable model for geographic lexical variation. In: Conf. on empirical methods in natural language processing; 2010. p. 1277-87.
[26] Cha M, Gwon YL, Kung HT. Geolocation with Subsampled Microblog Social Media. In: ACM Internat. Conf. on Multimedia; 2015. p. 891–894.
[27] Sanjar K, Bekhzod O, Kim J, Paul A, Kim J. Missing Data Imputation for Geolocation-based Price Prediction Using KNN–MCF Method. ISPRS Internat Journal of Geo-Information. 2020;9(4).
[28] Lieberman MD, Samet H, Sankaranarayanan J. Geotagging with local lexicons to build indexes for textually-specified spatial data. In: Internat. Conf. on data engineering. IEEE; 2010. p. 201-12.
[29] Mourad A, Scholer F, Magdy W, Sanderson M. A Practical Guide for the Effective Evaluation of Twitter User Geolocation. Trans Soc Comput. 2019 dec;2(3).
[30] Project Social Sensing. 2022. Mapping natural hazards with people as sensors;. Accessed on 23 January 2022. https://socialsensing.com/.
[31] Al-Moslmi T, Ocaña MG, Opdahl AL, Veres C. Named entity extraction for knowledge graphs: A literature overview. IEEE Access. 2020;8:32862-81.
[32] Kao A, Poteet SR. Natural language processing and text mining. Springer Science & Business Media; 2007.
[33] Baby CJ, Khan FA, Swathi J. Home automation using IoT and a chatbot using natural language processing. In: Innovations in Power and Advanced Computing Technologies; 2017. p. 1-6.
[34] Macnab C, Boxall P, Parkins J. The Social Context of Flood Risk in Alberta: Perspectives from Municipal Planners, Insurance Agents, the General Public and Media Sources; 2021. Project Report #21-01.
[35] Rahimi A, Cohn T, Baldwin T. pigeo: A python geotagging tool. In: ACL System Demonstrations; 2016. p. 127-32.