A Comparative Study of Geocoder Performance on Unstructured Tweet Locations

GI_Forum 2023, Issue 1
Page: 110 - 117
Short Paper
Corresponding Author:
helenngonidzashe.serere@plus.ac.at
DOI: 10.1553/giscience2023_01_s110

Helen Ngonidzashe Serere[1], Umut Nefta Kanilmaz[1], Sruthi Ketineni[1], Bernd Resch[1,2]

[1]University of Salzburg, Austria
[2]Havard University, USA

Abstract

Geocoding is a process of converting human-readable addresses into latitude and longitude points. Whilst most geocoders tend to perform well on structured addresses, their performance drops significantly in the presence of unstructured addresses, such as locations written in informal language. In this paper, we make an extensive comparison of geocoder performance on unstructured location mentions within tweets. Using nine geocoders and a worldwide English-language Twitter dataset, we compare the geocoders' recall, precision, consensus and bias values. As in previous similar studies, Google Maps showed the highest overall performance. However, with the exception of Google Maps, we found that geocoders which use open data have higher performance than those which do not. The open-data geocoders showed the least per-continent bias and the highest consensus with Google Maps. These results suggest the possibility of improving geocoder performance on unstructured locations by extending or enhancing the quality of openly available datasets.

Keywords:

commercial geocoders, natural language, Twitter, open data, spaCy

1 Introduction

Geocoding is omnipresent in our day-to-day lives. Tourists searching for nearby restaurants, emergency responders wanting to locate victims (Singh et al. 2019), or planners wanting to understand traffic flows (Das & Purves, 2020), to give but a few examples, all involve geocoding. However, generating accurate results is not a simple task but depends on various factors, including, among other things, the quality of the underlying reference database and the geocoder's robustness in dealing with natural language (Karimi, Durcik & Rasdorf, 2004). Whilst most geocoders perform well on structured addresses, their performance decreases in the presence of unstructured or partial addresses that may also include spelling, syntax or formatting errors.

This paper seeks to evaluate geocoder performance on unstructured locations, specifically ones embedded within tweets. We chose to use a Twitter dataset because of the existing need to increase the percentage of tweets with a geographic location (Karami et al., 2021; Singh et al., 2019). Additionally, the presence of geotagged tweets – that is, tweets with an attached GNSS position – allow for validating the geocoders' performance.

For studies seeking to geocode unstructured locations using off-the-shelf geocoders, we show which geocoders perform best on unstructured data and underline how the choice of geocoder can have a significant impact on the analysis.

2 Related work

Several studies have compared geocoder performance. Whitsel et al. (2006) assessed the quality of four commercial geocoders on 3,615 USA addresses. The authors found differences in geocoder performance and concluded that there was a need for an informed selection of geocoder. Lovasi et al. (2007) compared two geocoders, one using a single-stage method and the other a multi-stage one, on 8,157 Washington State addresses. They found that the multi-stage geocoder performed better than the single-stage one. Roongpiboonsopit and Karimi (2010) arrived at the same conclusion after comparing Google, Geocoder.us, Microsoft Virtual Earth, MapQuest and Yahoo! Maps on a sample of USA addresses. More recently, Owusu et al. (2017) compared Google Maps and MapQuest to validate the geocoding results obtained from ArcGIS. The authors found that Google Maps performed better than MapQuest.

While earlier studies carried out their comparisons on a small number of geocoders, on structured addresses, and on a limited geographical area, in this study we increase the number of geocoders for comparison and use a worldwide dataset with unstructured locations. To the best of our knowledge, no study has been done that matches the number of geocoders or the scale of our analysis on unstructured locations.

3 Geocoder Selection

Our first step was to select geocoders that could geocode unstructured locations on a global scale. For this, we restricted our selection to geocoders that had:

- A well-documented API with free-of-charge trial access
- Worldwide location coverage
- An API limit of at least 1,000 requests per day
- A client-side implementation in GeoPy[1].

Based on these criteria, we selected nine geocoders. Table 1 provides an overview of the geocoders selected. The open-data category indicates whether, to the best of our knowledge, the geocoder uses data that is openly accessible.

[1] https://geopy.readthedocs.io/en/stable/#

Table 1: Overview of the selected geocoders. A dash (–) in the daily limit and monthly limit rows denotes an instance where the geocoder does not explicitly define a limit.

Geocoder	Bing Maps[2]	Geolake[3]	GeoNames[4]	Google Maps[5]	HERE[6]	MapTiler[7]	Nominatim[8]	OpenCage[9]	TomTom[10]
Daily limit	125,000	1,500	20,000	–	1,000	–	Unlimited	2,500	2,500
Monthly limit	–	–	–	40,000	30,000	100,000	Unlimited	75,000	–
Open data	No	No	Yes	No	No	No	Yes	Yes	No

4 Methodology

The aim was to compare geocoder performance on unstructured locations mentioned within tweets. Figure 1 shows the overall methodology along with the corresponding sub-steps. We provide more details for each step in the sub-sections below.

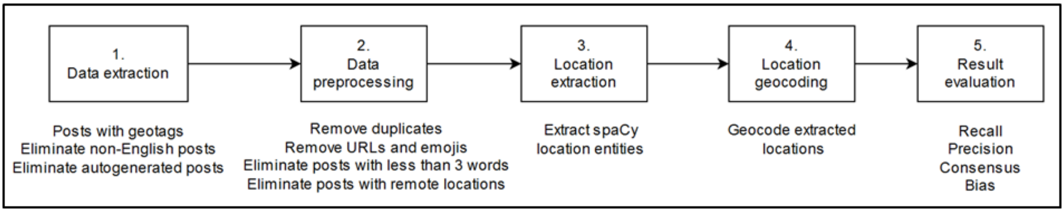

1. Data extraction
2. Data preprocessing
3. Location extraction
4. Location geocoding
5. Result evaluation
Posts with geotags
Eliminate non-English posts
Eliminate autogenerated posts
Remove duplicates
Remove URLs and emojis
Eliminate posts with less than 3 words
Eliminate posts with remote locations
Extract spaCy location entities
Geocode extracted locations
Recall
Precision
Consensus
Bias

Figure 1: Five-step overall workflow

Data extraction: We extracted a random sample of 125,000 tweets dated between August 2021 and September 2022. To allow for evaluation of the geocoder results, we selected only posts with a GNSS position attached. We restricted our extraction to posts generated in English to minimize errors caused by multi-language models and avoid possible translation errors.

[2] https://msdn.microsoft.com/en-us/library/ff701715.aspx
[3] https://geolake.com/
[4] https://www.geonames.org/
[5] https://developers.google.com/maps/documentation/geocoding/
[6] https://www.here.com/
[7] https://www.maptiler.com/
[8] https://nominatim.org/
[9] https://opencagedata.com/api
[10] https://www.tomtom.com/

Data preprocessing: Tweets are unstructured and noisy. Noisy elements in tweet text can degrade the performance of the models applied. We therefore pre-processed our extracted tweets by removing all emojis and URLs, posts that mentioned locations at some distance from the tweeter, posts with less than three words, and any duplicated posts. The pre-processing step returned 113,363 tweets out of the 125,000 initially extracted.

Location extraction: We extracted tweet locations using spaCy[11], a pre-trained, syntax-based Named Entity Recognition Model (NER). It is an open-source library for Natural Language Processing and provides NER models in multiple languages and multiple sizes. For this study, we used spaCy version 3.1.1 (known as en_core_web_trf), a high-accuracy English-language model. Using spaCy, we extracted location mentions from 44,787 of the 113,363 pre-processed tweets.

Location geocoding: We passed all extracted locations into our chosen geocoders (Table 1). Each geocoder returned a list of latitude and longitude value pairs for each successfully geocoded location.

Result evaluation: We evaluated the performance of the geocoders by calculating four evaluation metrics, namely recall, precision, consensus and bias.

- **Recall**: percentage of geocoded locations out of all locations that had been passed into the geocoder.
- **Precision**: percentage of locations geocoded within a 100 km geodesic radius of the tweets' GNSS positions out of the total number of successfully geocoded locations.
- **Consensus**: percentage of locations geocoded within the same country for each pair of geocoders**.**
- **Bias**: percentage of wrongly geocoded locations at continent level.

5 Results

Figure 2 shows the precision and recall values for the nine selected geocoders. It is apparent that geocoders vary in performance. Whilst MapTiler and Bing Maps showed recall values of over 90%, GeoNames, HERE and Geolake exhibited values of less than 60%. With regards to precision, whilst Google Maps returned over 82% of its geocoded locations within a 100 km radius of the tweet's GNSS position, MapTiler, TomTom and HERE returned less than 50% of their geocoded locations within the same radius value. Overall, within a 100 km radius, Google Maps returned the highest precision value (82.50%), followed by Nominatim (67.33%), GeoNames (64.47%) and OpenCage (60.00%).

[11] https://spacy.io/

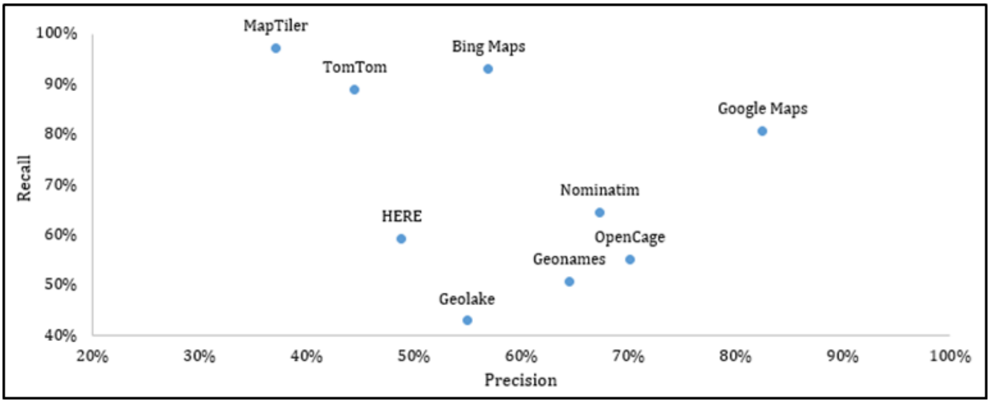


Figure 2: Precision and recall values for all geocoders over a distance of 100 km

In the next step, we computed the consensus between each geocoder pair. Figure 3 shows the correlation matrix of these pairs. Overall, Nominatim and OpenCage returned the highest correlation value (98.24%). Geolake and HERE returned a correlation value of 96.61%. Google Maps returned values of 93.36% with Nominatim, 90.33% with GeoNames, 86.88% with OpenCage, and 85.34% with Bing Maps. The other geocoder pairs had correlation values of less than 85%. HERE and MapTiler showed the lowest consensus (61.5%) of all geocoders.

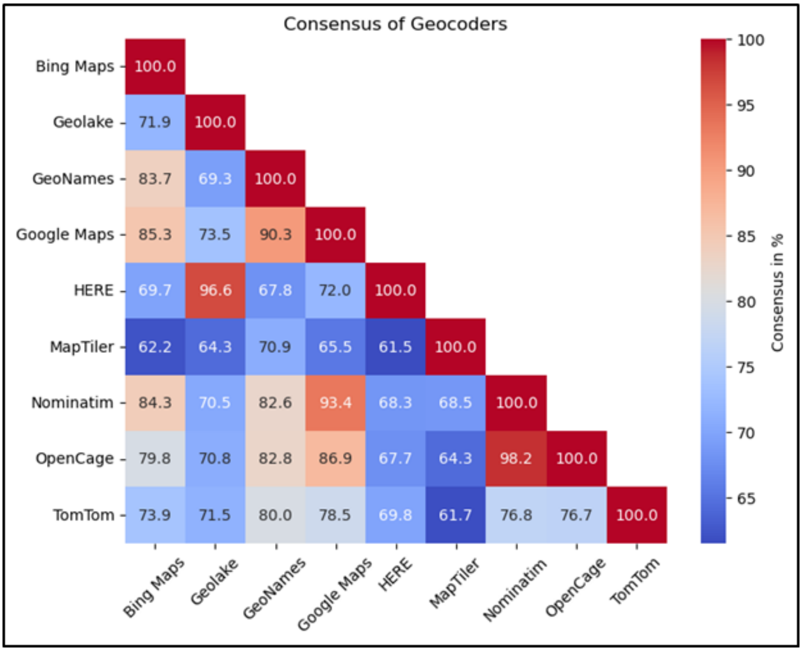
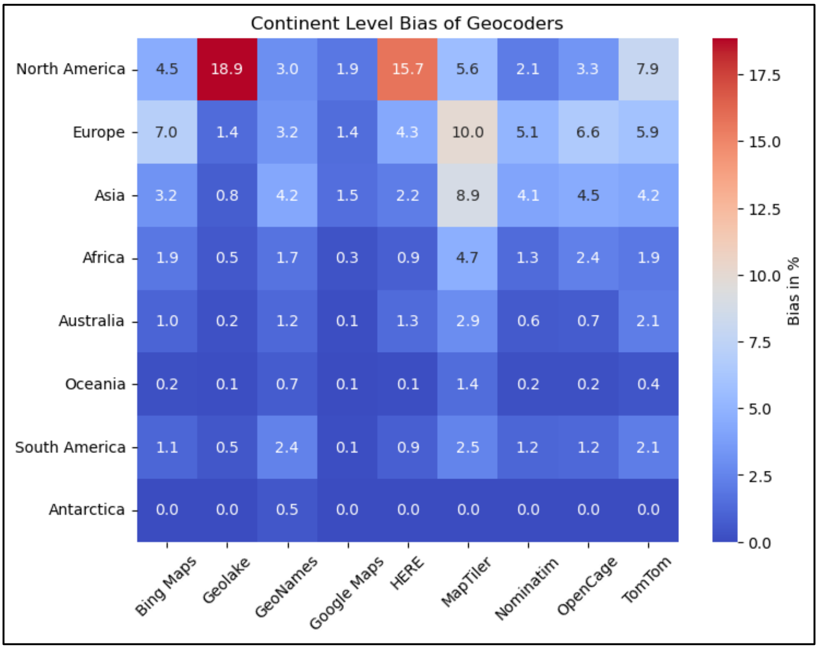


Figure 3: Correlation matrix of country-level consensus values. The values on the tiles indicate the percentage of locations that the corresponding pair of geocoders classified into the same country.

Next, we computed the geocoder bias at continent level (see Figure 2). It is apparent that Geolake and HERE were highly biased towards North America. Geolake had about 14 times more bias and HERE approximately four times more bias towards North America in comparison to other continents. The other geocoders showed a lower magnitude of difference per continent: Google Maps and TomTom were more biased towards North America; Nominatim, MapTiler, Bing Maps and OpenCage were more biased towards Europe, and GeoNames towards Asia.



	Bing Maps	Geolake	GeoNames	Google Maps	HERE	MapTiler	Nominatim	OpenCage	TomTom
North America	4.5	18.9	3.0	1.9	15.7	5.6	2.1	3.3	7.9
Europe	7.0	1.4	3.2	1.4	4.3	10.0	5.1	6.6	5.9
Asia	3.2	0.8	4.2	1.5	2.2	8.9	4.1	4.5	4.2
Africa	1.9	0.5	1.7	0.3	0.9	4.7	1.3	2.4	1.9
Australia	1.0	0.2	1.2	0.1	1.3	2.9	0.6	0.7	2.1
Oceania	0.2	0.1	0.7	0.1	0.1	1.4	0.2	0.2	0.4
South America	1.1	0.5	2.4	0.1	0.9	2.5	1.2	1.2	2.1
Antarctica	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0

Figure 2: Bias of geocoders at continent level. Higher values indicate a higher percentage of locations falsely classified into a particular continent. The continents are sorted by decreasing number of extracted tweets.

6 Discussion and Conclusion

The aim of this paper was to compare geocoder performance on unstructured locations. We compared nine geocoders on 44,747 unstructured locations extracted from a worldwide Twitter dataset. As already observed in previous studies, here too we found **large differences in geocoder performance**. Probably the most intriguing finding to emerge from our study was the high performance of open-data geocoders compared to geocoders with closed data, a finding that will be of interest to researchers seeking to enhance geocoder performance.

With the exception of Google Maps, the three highest-performing geocoders are open-data based, namely Nominatim (67.3%), GeoNames (64.5%) and OpenCage (60.0%). This finding suggests the possibility of enhancing geocoder performance by improving the quality of openly available datasets such as the on-going OpenStreetMap mapping project.

Looking at **geocoder consensus**, it was interesting to see which geocoders had a high degree of agreement with Google Maps, the geocoder with the highest precision. Again, we found that the geocoders based on open data – Nominatim (93.4%), GeoNames (90.3%) and OpenCage (86.9%) – showed the highest levels of consensus.

Geolake and HERE had the overall **highest geocoder bias** values, of 18.9% and 15.7% respectively. Both wrongly assigned a larger percentage of locations to North America, proving that some geocoders are biased towards certain geographical regions (Whitsel et al., 2006). We would like to stress that, whilst geocoder bias indicates in this case bias towards continents (i.e. assigning a location to the wrong continent), areas with low or zero bias do not indicate that the geocoder performs well in those areas. It could be that the geocoder fails to cover those regions and thus does not return any location (either correct or incorrect). This might be the case for Antarctica and Oceania, where the bias was mostly zero.

With regards to the overall performance of each geocoder, we would like to point out that the values reported in this study relate exclusively to our input dataset. Changes in input data, by correcting location mentions (Clemens, 2018), dropping some locations, or using locations extracted from a different dataset, would produce different results (Whitsel et al., 2006).

Although we aimed for a robust analysis, there were **limitations** in our workflow that may have affected the geocoders' precision and recall values. We filtered remote locations (i.e. locations that do not coincide with the tweeters' positions) using a list of keyword phrases adopted from Serere, Resch and Havas (2023). This method is not robust, however, and has a high miss rate which impacts geocoder precision. Future studies should consider more robust approaches, such as grammatical or syntax-based filtering.

We used a syntax-based model, spaCy, to extract location mentions, which has the advantage of not being affected by spelling errors. A disadvantage, however, is that there is a substantial risk of false-positive locations, which then impacts the geocoder recall values. Future studies could consider cross-checking extracted locations by fuzzy-matching the extracted locations to a gazetteer before geocoding.

In conclusion, the results of our study can be used as a guide for selecting the geocoder to use for unstructured locations. Using the right geocoder to geotag social media posts correctly would help increase the robustness of various spatial applications such as disaster mapping, mapping of refugee movements, or user sentiment analysis. Furthermore, geocoder providers could also make use of our findings to see where they might improve their services.

References

Clemens, Konstantin. 2018. Enhanced Address Search with Spelling Variants. Pp. 28–35 in Proceedings of the 4th International Conference on Geographical Information Systems Theory, Applications and Management. SCITEPRESS - Science and Technology Publications.

Das, Rahul Deb, and Ross S. Purves. 2020. Exploring the Potential of Twitter to Understand Traffic Events and Their Locations in Greater Mumbai, India. IEEE Transactions on Intelligent Transportation Systems 21(12):5213–22. doi: 10.1109/TITS.2019.2950782.

Karami, Amir, Rachana Redd Kadari, Lekha Panati, Siva Prasad Nooli, Harshini Bheemreddy, and Parisa Bozorgi. 2021. Analysis of Geotagging Behavior: Do Geotagged Users Represent the

Twitter Population? ISPRS International Journal of Geo-Information 10(6):373. doi: 10.3390/ijgi10060373.

Karimi, Hassan A., Matej Durcik, and William Rasdorf. 2004. Evaluation of Uncertainties Associated with Geocoding Techniques. Computer-Aided Civil and Infrastructure Engineering 19(3):170–85. doi: 10.1111/j.1467-8667.2004.00346.x.

Lovasi, Gina S., Jeremy C. Weiss, Richard Hoskins, Eric A. Whitsel, Kenneth Rice, Craig F. Erickson, and Bruce M. Psaty. 2007. Comparing a Single-Stage Geocoding Method to a Multi-Stage Geocoding Method: How Much and Where Do They Disagree? International Journal of Health Geographics 6:1–11. doi: 10.1186/1476-072X-6-12.

Owusu, Claudio, Yu Lan, Minrui Zheng, Wenwu Tang, and Eric Delmelle. 2017. Geocoding Fundamentals and Associated Challenges. Pp. 41–62 in Geospatial Data Science Techniques and Applications.

Roongpiboonsopit, Duangduen, and Hassan A. Karimi. 2010. Quality Assessment of Online Street and Rooftop Geocoding Services. Cartography and Geographic Information Science 37(4):301–18. doi: 10.1559/152304010793454318.

Serere, Helen Ngonidzashe, Bernd Resch, and Clemens Rudolf Havas. 2023. Enhanced Geocoding Precision for Location Inference of Tweet Text Using SpaCy, Nominatim and Google Maps. A Comparative Analysis of the Influence of Data Selection. Ed. by P. Vellucci. PLOS ONE 18(3):e0282942. doi: 10.1371/journal.pone.0282942.

Singh, Jyoti Prakash, 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(1–2):737–57. doi: 10.1007/s10479-017-2522-3.

Whitsel, Eric A., P. Miguel Quibrera, Richard L. Smith, Diane J. Catellier, Duanping Liao, Amanda C. Henley, and Gerardo Heiss. 2006. Accuracy of Commercial Geocoding: Assessment and Implications. Epidemiologic Perspectives and Innovations 3(May 2014). doi: 10.1186/1742-5573-3-8.