



*International Journal of Engineering and Geosciences (IJEG),
Vol; 5, Issue; 2, pp. 109-119, June, 2020, ISSN 2548-0960, Turkey,
DOI: 10.26833/ijeg.629381*

ACCURACY AND SIMILARITY ASPECTS IN ONLINE GEOCODING SERVICES: A COMPARATIVE EVALUATION FOR GOOGLE AND BING MAPS

Batuhan Kilic¹, Fatih Gulgen^{1*}

¹Yildiz Technical University, Civil Engineering Faculty, Department of Geomatic Engineering, Istanbul, Turkey
(fgulgen/batuhank@yildiz.edu.tr); **ORCID 0000-0002-8754-9017, ORCID 0000-0002-0529-8569**

*Corresponding Author, Received: 04/10/2019, Accepted: 12/11/2019

ABSTRACT: Geocoding is a method used to convert address information into geographical coordinates. It plays a vital role in displaying the relationship between geographic features and semantic information expressed in texts. The objective of this study is to reveal the quality of online geocoding from postal addresses in Turkey provided by Google Maps and Bing Maps services. The quality of geocoding services in urban areas is evaluated using two particular metrics; positional accuracy and address similarity. Positional accuracy measures the distances between point features obtained through the online geocoding and reference data. Address similarity indicates the relationship between two postal addresses based on a similarity index known as the Levenshtein distance. The same performance assessment was also made with the United States' address data to make comparisons and discussions. The results show that services have different geocoding capabilities in both countries because of the differences in the addressing formats.

Keywords: *Geocoding, Google Maps, Bing Maps, Positional accuracy, Levenshtein distance*

1. INTRODUCTION

One of the essential processes used in the geographical analysis is geocoding, which assigns a coordinate pair to the description of a place by comparing the descriptive location-specific elements to those in reference data (Zandbergen, 2008). Researchers often use geocoding services to conduct the geographical analysis in various areas such as epidemiology (McElroy et al. 2003; Ward et al. 2005; Rushton et al. 2006), public safety (Ratcliffe 2004; Bichler and Balchak 2007; Hart and Zandbergen 2013), and traffic accident detection (Qin et al. 2013). There are two types of geocoding tools known as offline or online. Software packages in Geographic Information System (GIS) include offline geocoding tools. Online geocoding tools depend on map services. The popular services which provide online geocoding process use street or rooftop geocoding techniques. Street geocoding utilizes street network data, while the rooftop technique employs address (point) data. Considering that geocoding performance relies heavily on the quality of the reference database, the rooftop geocoding would produce higher accurate results than street geocoding (Roongpiboonsopit and Karimi 2010b).

Several cartographers and GIS specialists studying on geocoding were focused on the creation of model proposals and the assessment of positional quality between the different service results. The model proposal studies usually aimed to solve complex address structures. Li et al. (2010) proposed an address model using a method in China called association rule mining based on machine learning to eliminate the adverse effect of the irregular address structure. Tian et al. (2016) proposed a geocoding service based on an optimized address mapping method by designing a particular address model for China and revealing the differences between the address structures of China and Western countries. The geocoding quality assessment studies also aim to compare the results of different services in different regions. Cayo and Talbot (2003) performed geocoding with MapMarker software using residential addresses. They evaluated the positional accuracy of the obtained points by comparing them with orthophoto maps. Yang et al. (2004) identified the error sources that resulted in geocoding by revealing the pros and cons of commercial (i.e., ArcView, Automatch, and ZIP/GeoLytics) geocoding systems. Karimi et al. (2004) investigated the matching rate and positional accuracy uncertainties arising from geocoding techniques by comparing them with the data collected by the Global Navigation Satellite System (GNSS). Roongpiboonsopit and Karimi (2010a) intended to provide an overview of the appropriate services by examining the matching quality of the online geocoding services: Geocoder.us, Google Maps, MapPoint, MapQuest, and Yahoo. The quality of each service was assessed based on three different measures, defined as matching rate, positional accuracy, and similarity between the services as a result of matching. Zandbergen (2011) compared the geocoding results obtained from three different datasets to determine the influence of road data on geocoding studies. Goldberg et al. (2013) developed an evaluation framework for comparing geocoding systems without mentioning the name because of the non-disclosure

agreements. Chow et al. (2016) compared the results of geocoding services (i.e. desktop geocoding software: ESRI and CoreLogic PxPoint and online geocoding software: Google Maps, Yahoo, Bing Maps, Geocoder.us, Texas A&M University Geocoder and OpenStreetMap) with GNSS metrics by gathering residential addresses in the state of Texas, United States of America (USA). Cetl et al. (2016) compared two online geocoding services (Google Maps and OSM Nominatim) results, taking into account population density in the city of Zagreb, Croatia.

Turkey is one of the countries which spread over a wide area and has a high population. The official institutions have tried to reorganize the unusual address expressions in a standard form for many years. However, abundant address components, improper and incorrect implementation of numerating studies, the failure of institutions to use the same address format even in their documents, and institutions' unawareness of new address components bring about confusion related to the addresses in Turkey (Yildirim et al., 2014; Kilic and Gulgen, 2019). Since the irregular expressions obscure understanding the components of an address specified by a user, the geocoding services have trouble matching the input address with the standard reference dataset. Kilic and Gulgen (2017) have compared the postal addresses used in Turkey and the USA within the scope of the geocoding process. They revealed that the geocoding accuracy in Turkey is less because of the following reasons: external door number, incomplete address, misspellings, typographical errors, and incorrect format.

A reference address is composed of a group of address components. The general components of a formatted address used in many countries are address number, street name, town, state/abbreviation, and postal code. Parsing them from an address description is a complicated process because users cannot usually define addresses in a standard form. The difficulty in converting the user's entries to the standard address form affects the quality of the geocoding process. The geocoding quality has been frequently evaluated in countries that have a high standard addressing system. However, the quality of geocoding issues in Turkey has not been adequately investigated yet. This study assesses the quality of the geocoding process used in Google Maps and Bing Maps services for Turkey and the USA. Both services, which are widely used throughout the world, are freely available to everyone. The purpose of this paper is to reveal the quality of geocoding by (1) measuring the positional accuracy between the retrieved and reference data and (2) comparing the text of retrieved with the reference using Levenshtein distance, which is a text similarity technique. While the first type of evaluation is not new, the later evaluation has never before been reported in the literature.

2. STUDY AREA AND REFERENCE DATA

Evaluating the performance of the geocoding process with dependable and accurate data sets for a large geographic area is quite tricky (Roongpiboonsopit and Karimi 2010b). Therefore, two distinct experimental tests in the urban centers of the USA and Turkey with a limited number of reference data that are reliable and

have similar characteristics were implemented.

A point of interest (POI) is a specific point location that someone may find useful or interesting. People use POIs for various purposes, such as navigating between different locations, determining the characteristics of a place, studying urban sociology, analyzing city dynamics, and geo-referencing of the texts (Rodrigues 2010). In this study, the accommodation facilities which supply communication between the dynamics of the cities (Bilgi et al. 2019) were used as the reference data. They are usually stored by point symbols in a POI layer of various online maps (Mulazimoglu and Basaraner, 2019). Accordingly, two test regions were selected: the district of Fatih in Turkey and the city of Miami Beach in the USA. Fatih is located in the historical peninsula of Istanbul, surrounded by the Byzantine city walls, the Golden Horn, and the Sea of Marmara (Figure 1). This region, which has harbored many civilizations throughout its thousands of year's history, is currently the most significant historical tourism center of Istanbul. The other test region, Miami Beach, is a famous coastal resort city in Florida (Figure 2). The city is a major international entertainment and cultural destination. It is a widely visited tourist area and offers various accommodation facilities.

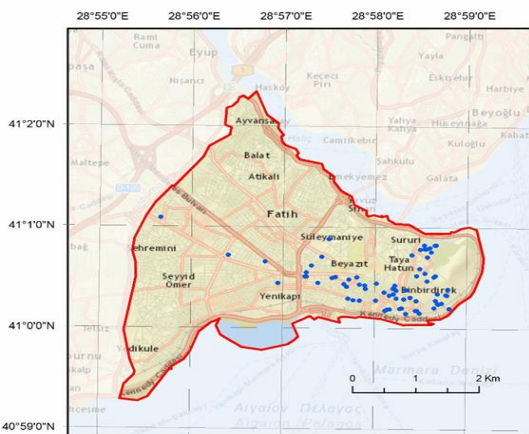


Figure 1. Approximate boundary of the study area within Fatih and accommodation facilities (blue points).

The postal addresses of the accommodation facilities forming the basis of this study were obtained from www.booking.com website automatically by using the web scraping method. Web scraping is a data mining process used to collect standard information stored in a specific location of a web page (Mitchell 2015). There are several commercial web scraping software such as Visual Web Ripper, Web Scraper Plus+, and many more (Rodrigues, 2010). In this study, *Web Content Extractor v.8.4* software developed by Newprosoft Company was used to collect the accommodation data. As a result of the scraping process, name and address information of 250 and 271 hotels with various types of building were obtained from Fatih and Miami Beach, respectively. Three scraped postal addresses are given in Table 1 to demonstrate the general address characteristics of the test regions. The first sample address is from the USA, while the other two are from Turkey.

The majority of scraped addresses contained repetitive and inconsistent information. Therefore, the problematic ones were omitted from the list after

reviewing the entire data. The remaining addresses were considered identifiers of POIs, and their approximate locations were geocoded from the facility name by using Google Maps and Bing Maps services. Then, the advertising signs of the facilities over the street perspectives were checked to confirm the accuracy of the geocoding process based on the facility names (Figure 3a and Figure 4a). Thus, the building footprints of 74 hotels in Fatih and 82 hotels in Miami Beach were verified as reference data for test processes (Figure 3b and Figure 4b).

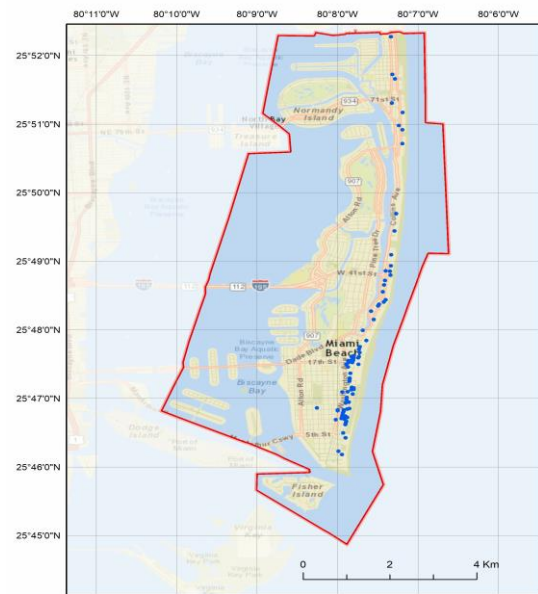


Figure 2. Approximate boundary of the study area within Miami Beach and accommodation facilities (blue points).

Table 1. Examples of postal addresses scraped from www.booking.com website automatically.

Name	Postal Address
Aqua	1530 Collins Avenue, Miami Beach,
Hotel&Suits	FL 33139, United States of America
Angel's	Sultan Ahmet Mh. Amiral Tafdil Sk.
Home Hotel	No:26, 34122 Fatih/Istanbul, Turkey
Hotel	Cankurtaran Mh. Amiral Tafdil Sk.
Romantic	No:17, 34122 Fatih/Istanbul, Turkey



Figure 3. a) Street view and b) Map view of a reference hotel in Fatih.



Figure 4. a) Street view and b) Map view of a reference hotel in Miami Beach.

The address point of a building can be extracted from a combination of various data sources including, but not limited to, parcel maps, building footprints, orthophotos, satellite imagery, and field survey (Roongpiboonsopit and Karimi 2010b). It represents the location of an addressable structure in a jurisdiction and is usually placed directly on top of the specific footprint or directly in front of it (Hart and Zandbergen 2013). These are valid and acceptable representations. In this study, the POIs were placed at the centroid of each digitized polygons. Then, their locations were stored in individual datasets for the test regions. (Figure 3b and Figure 4b). The centroid of a building polygon is the mean position of all vertices in two coordinate directions. If it falls outside the boundaries, the point can be displaced to what is considered a center of gravity within the boundaries.

3. ONLINE GEOCODING FROM POSTAL ADDRESS

Online geocoding is a network-accessible component which is available on the Internet with a Web Application Programming Interface (API) as a service. The geocoding API converts an entry into coordinates and then delivers the result, which includes the coordinates, the address used in geocoding, and the level of accuracy back to the user over the internet (Roongpiboonsopit and Karimi 2010a). During the online geocoding process, a service parses an address data into its natural elements and compares them to the potential candidates within the reference database of service. Then, it identifies and delivers the normalized address and geographical coordinates of the best-matched address as the output. The purpose of parsing is

to segment an input address string into meaningful address elements with exact spatial semantics; in address normalization, any informal or abbreviated address elements are converted into a standard format, and address elements that are misspelled are re-recorded in the correct form (Tian et al. 2016).

In this study, the reference addresses were geocoded using a library for the Google Maps Geocoding API and the Bing Maps Locations API in the Python programming language. As with other online geocoding services, Google Maps and Bing Maps services have different limits and constraints concerning licensing of service usage (URL 1 and URL 2). Both services produce their results using a rooftop geocoding technique for all postal addresses in Miami Beach, while 51 of 74 postal addresses in Fatih are coming from the roof. Of the remaining 23 postal addresses, Bing Maps and Google Maps generated 14 and 8 using the rooftop technique, respectively. There was only one address that is geocoded by the street technique in both services.

The main address components of Google Maps and Bing Maps with the parsed elements belonging to three sample hotels presented in Table 1 were given in Table 2. These services had many parts such as house number, street name, postal code, neighborhood, county, city, state, and country, though some of them were defined in a combined form (e.g., Bing Maps converts the house number and street name into one component).

Table 3 contains the normalized address data and the related geographic coordinates in the standard form after the parsing process. Both services for the first input data returned pretty similar normalized address information, although Bing Maps could not give the parsed elements for neighborhood and county. For the second and third input data, neighborhood, county, and city features were not parsed correctly by the geocoding services because of the complexity or deficiency in sub-region information regarding the addressing system in Fatih. Google Maps converted the postal code element in the third parsed address because of the undefined neighborhood element. Google Maps' normalized format contained a neighborhood element that was incompatible with the postal code information. Bing Maps did not present any neighborhood information which was parsed correctly from input data (Table 3).

Table 2. Postal addresses parsed by services in a standard format.

Services	House Number	Street Name	Postal Code	Neighborhood	County	City	State	Country
Google Maps	1530	Collins Avenue	33139	South Beach	Miami-Dade County	Miami Beach	FL	US
	26	Amiral Tafdil Sk.	34122	-	Fatih	-	Istanbul	TR
	17	Amiral Tafdil Sk.	34122	-	Fatih	-	Istanbul	TR

Table 2. (Continued).

Bing Maps	1530 Collins Ave	33139	-	-	Miami Beach	FL	United States
	Amiral Taftil Sokak 26	34122	Sultan Ahmet	-	Istanbul	Istanbul	Turkey
	Amiral Taftil Sokak 17	34122	Cankurtaran	-	Istanbul	Istanbul	Turkey

Table 3. Postal addresses normalized by services in a standard format.

Services	Name	Address	Latitude and Longitude
Google Maps	Aqua Hotel&Suits	1530 Collins Ave, Miami Beach, FL 33139, USA	[25.78825, -80.13083]
	Angel's Home Hotel	Sultan Ahmet Mh., Amiral Tafdil Sk. No:26, 34122 Fatih/Istanbul, Turkey	[41.00415, 28.97945]
	Hotel Romantic	Cankurtaran Mh., Amiral Tafdil Sk. No:17, 34122 Fatih/Istanbul, Turkey	[41.00424, 28.97989]
Bing Maps	Aqua Hotel&Suits	1530 Collins Ave, Miami Beach, FL 33139	[25.78836, -80.13086]
	Angel's Home Hotel	Amiral Taftil Sokak 26, 34122 Istanbul, Turkey	[41.00421, 28.97962]
	Hotel Romantic	Amiral Taftil Sokak 17, 34122 Istanbul, Turkey	[41.00427, 28.97989]

In Turkish addressing format, the sub-region generally described by a neighborhood name is a crucial element. In some cases, buildings such as accommodation facilities handled in this study may be located within a different sub-region despite being on the same street. This paradox is the most fundamental issue that has not been solved by global geocoding services such as Google Maps and Bing Maps in Turkey. Such examples can be seen in the second and third addresses in Table 3 for both services in Fatih. On the other hand, both services use their particular address formats in Fatih. Bing Maps gives a house number to the structure after the street name and type, yet Google Maps always returns a normalized postal address together with an abbreviation "No:" for the same place used in Bing Maps.

4. QUALITY OF GEOCODING

All geocoding results consisting of normalized addresses and geographic coordinates in both test regions obtained by Google Maps and Bing Maps services are examined below in two sections: positional accuracy and similarity of address description.

4.1 Positional Accuracy

The positional accuracy of the online geocoding is evaluated compared to the reference data and dealt with in two stages:

- Analysis of the distance and the coordinate deviation between the geocoded point of each address and its actual location,

- Analysis of the topological relation between the geocoded point of each address and its building footprint.

In the first stage, the positional accuracy between the geocoded point and the actual location by measuring the straight-line distance in the Euclidean space was calculated. Several authors assumed that the precise location of an address is the centroid of the building footprint on a parcel (Zandbergen and Green 2007; Armstrong and Tiwari 2008; Zandbergen 2008; Roongpiboonsopit and Karimi 2010b; Chow et al. 2016). In this study, latitudes and longitudes were transformed into projected coordinates to calculate the positional deviation between the geocoded and centroid points. For Fatih and Miami Beach, the Gauss Kruger in ITRF96 datum and the Florida State Plane in NAD83 datum were the coordinate systems, respectively. The deviation for easting and northing coordinates was calculated by using Eq. (1) and Eq. (2), respectively.

$$\Delta E_i = E_{ref} - E_i \quad (1)$$

$$\Delta N_i = N_{ref} - N_i \quad (2)$$

where ΔE_i and ΔN_i refer to deviations in the easting and northing coordinates, E_{ref} and N_{ref} are the easting and northing coordinates of the centroid, E_i and N_i are the easting and northing coordinates of the geocoded point, respectively. The positional deviation ΔS_i between the two points is calculated from the values ΔE_i and ΔN_i by using Eq. (3).

$$\Delta S_i = \sqrt{\Delta E_i^2 + \Delta N_i^2} \quad (3)$$

The calculated deviations in Fatih and Miami Beach are shown on the graphs in Figure 5 and Figure 6, respectively. The horizontal coordinate in all charts represents the rank, which increases depending on the positional deviation value calculated for each accommodation facility. The vertical coordinates of ΔE and ΔN graphs show positive or negative deviations

between the geocoded point and the centroid in the easting and northing directions, respectively. The vertical coordinate of ΔS refers to the values of positional deviation. According to the deviation graphs, Bing Maps' geocoding results have a more significant difference than Google Maps' results in Fatih. The results of both services in Miami Beach are reasonably close to each other (Figure 5 and Figure 6).

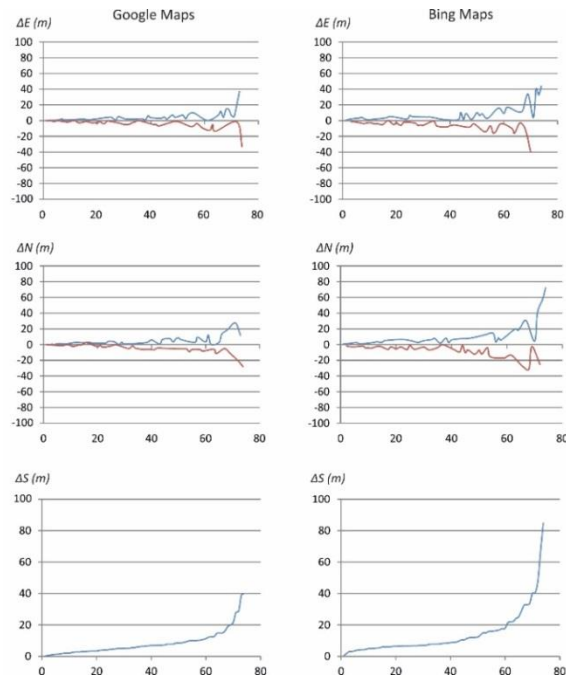


Figure 5. Easting, northing, and positional deviations for Fatih: blue and red lines show the positive and negative deviations, respectively.

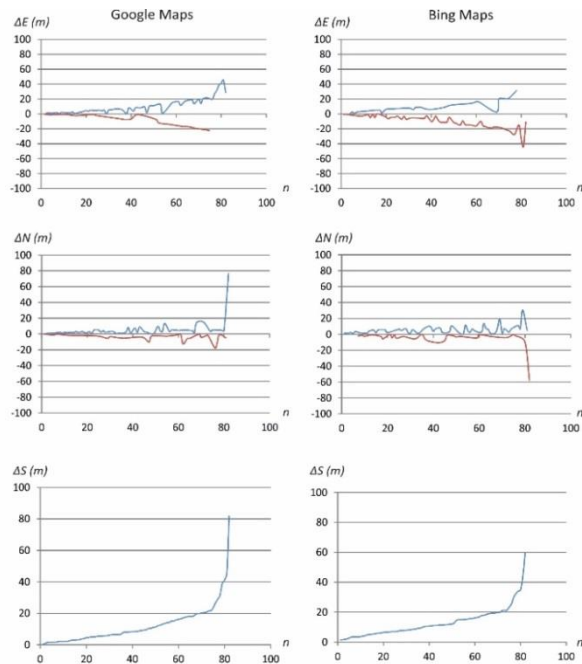


Figure 6. Easting, northing, and positional deviations for Miami Beach: blue and red lines show the positive and negative deviations, respectively.

Statistically, several measures known as minimum, maximum, median, mean, and norm are used to reveal the quantity of error between the geocoded point and the exact location. The minimum and maximum are the least and the highest samples of a set of observations. The median and mean are the conventional measures of central tendency in summary statistics. Equation (4) shows the norm value, which is the total magnitude of observation values. It increases depending on the number of observations in a set.

$$Norm = \sqrt{\sum_i^n x_i^2} \quad (4)$$

where x refers to each observation value, and n

Table 4. Statistical measures calculated from deviation values.

Region	Measure	ΔS (m)		ΔE (m)		ΔN (m)	
		Google Maps	Bing Maps	Google Maps	Bing Maps	Google Maps	Bing Maps
Fatih	Min	0.1	0.4	0.1	0.3	0.1	0.3
	Max	39.7	85.1	37.3	44.3	28.6	72.6
	Median	6.4	8.2	3.1	5.2	4.4	5.8
	Mean	8.3	13.9	4.9	8.1	5.7	9.8
	Norm	97.5	171.2	67.8	105.4	70.1	134.9
Miami Beach	Min	0.3	1.4	0.1	0.1	0.2	0.1
	Max	82.0	59.7	45.3	43.7	76.9	58.7
	Median	8.4	10.9	6.8	8.2	2.8	3.5
	Mean	12.2	12.9	10.0	10.4	4.8	5.4
	Norm	155.3	146.5	126.2	120.6	90.5	83.3

On the other hand, when Google Maps' statistical values were examined, it is found that the amount of deviations within Fatih is lower than the differences within Miami Beach. The reason for better results of geocoding in Fatih is the fact that most of the building geometry in Miami Beach consists of non-rectangular and complex shapes, and also an enormous amount of buildings in Fatih has simple geometry. Besides, the statistical values are insufficient to make decisions in the test region, where Bing Maps results better than Google Maps.

In the second stage, the positional accuracy between the geocoded point and building footprint was measured by considering the point&polygon topological relationships. The point-in-polygon algorithm, which is an essential operation to determine whether a point is inside a complex polygon (Haines, 1984), was employed. Likewise, Roongpiboonsopit and Karimi (2010b) used this algorithm to verify the correctness of geocoded locations at the building level.

When geocoding an address using the rooftop technique, one can expect that the point will place within the building footprint. The point shown in blue in Figure 7 is adequately geocoded. However, this expectation is

represents the number of observations.

The minimum, maximum, median, mean, and norm values of the deviations are given in Table 4. The absolute values are used when calculating the minimum, median, and mean. According to Table 4, all Google Maps' statistical values in Fatih are lower than Bing Maps' values. These values prove that Google Maps' geocoding results have higher accuracy than Bing Maps' results in Fatih. The statistical significances of the deviations in Miami Beach are quite varied, and the benefits of neither Google Maps nor Bing Maps are superior over the other. They represent that the positional accuracies of both services are approximately similar to each other in Miami Beach.

not always possible. The red spot is improper because the geocoding service locates it outside the building footprint. The numbers and percentages of point-in-polygon analysis in each test region used in this study are presented in Table 5.



Figure 7. Geocoding points located inside and outside the digitized buildings.

The results of the point-in-polygon analysis show

that the number of geocoded points obtained adequately from the Google Maps service is more than that obtained from the Bing Maps service in both regions. Similarly, the results of Miami Beach are better than those of Fatih. The differences in percentages between Miami Beach and Fatih are approximately 5% on Google Maps and 15% on Bing Maps.

Table 5. Point-in-polygon analysis for geocoding services.

Region		Google Maps		Bing Maps	
		Inside	Outside	Inside	Outside
Fatih	Number	61	13	44	30
	%	82.4	17.6	59.5	40.5
Miami Beach	Number	72	10	61	21
	%	87.8	12.2	74.4	25.6

On the other hand, the best geocoding location for an address is a place close to the front edge of the building footprint. The front side is determined by considering the actual status of the building's main entrance door. Considering that the proper geocoding point is on a building's footprint boundary, if a geocoded point is within the footprint, the address can be assumed to be geocoded correctly. It means that the location of the geocoded point can be varied from the centroid to the footprint of a building. The quantity of error for a geocoding point, which is outside the footprint, is calculated from the closest distance between the location and the related footprint edges (Figure 7). Table 6 provides some statistical values for the improper geocoding points in each test region. These show that geocoding results obtained from Google Maps are of higher quality than Bing Maps, and the results in Miami Beach are more reliable than the results in Fatih.

Table 6. Statistical measures calculated for improper (outside) geocoded points.

	Distances (m)			
	Fatih		Miami Beach	
	Google Maps	Bing Maps	Google Maps	Bing Maps
Minimum	0.6	0.1	0.3	0.7
Maximum	28.2	60.5	6.3	8.5
Median	5.0	6.6	1.8	2.5
Mean	9.3	11.1	2.1	3.0
Norm	48.1	92.8	8.6	17.3

4.2 Address Similarity

The similarity rate between two strings can be calculated by using a similarity measurement method

known as string metric in information theory, mathematics, and computer science. The most widely known string metric is the Levenshtein distance. The minimum number of operations gives the Levenshtein distance between two string vectors, and that needed to transform one text into the other, where an action is an insertion, deletion, or substitution of a single character (Levenshtein 1966). In this study, The Levenshtein distance as an indicator of similarity between two postal addresses was used. The algorithm detects the address similarity in two stages. In the first, the similarity percentage between the postal address of an accommodation facility obtained from the scraped website and the best-matched address in each service database is calculated. Then, a calculation of the similarity percentage is done between the results of both services. Figure 8 shows the similarity graphs of computed values for both test regions. According to the charts containing scraped data, the similarity in Fatih for Google Maps changes approximately between 50-85% and between 35-82% for Bing Maps. The similarity in Miami Beach for Google Maps and Bing Maps is roughly between 67-80% and 60-75%, respectively. The increased difference between the highest and lowest percentage values (i.e., scale expanding on the vertical axis of the graphs) points out the issues that are growing related to the use of standard address format. Differences between the maximum and minimum values in the charts reveal that the address standard in Miami Beach is higher than that in Fatih. The similarity between the addresses obtained from both services is approximately 37-73% in Fatih and 66-95% in Miami Beach. Also, the average value calculated for both services in Miami Beach is hugely higher than the others (Table 7). These percentages indicate that Google Maps and Bing Maps return the addresses in a similar standard format for Miami Beach. However, the address format used by each service has different standards in Fatih due to the complications in the identification of the address components in Turkey.

Table 7. Similarity percentages for Google Maps, Bing Maps geocoded addresses, and reference addresses.

	Reference addresses		Google Maps
	Google Maps	Bing Maps	Bing Maps
Fatih	65.3	64.6	61.2
Miami Beach	72.3	66.9	93.2

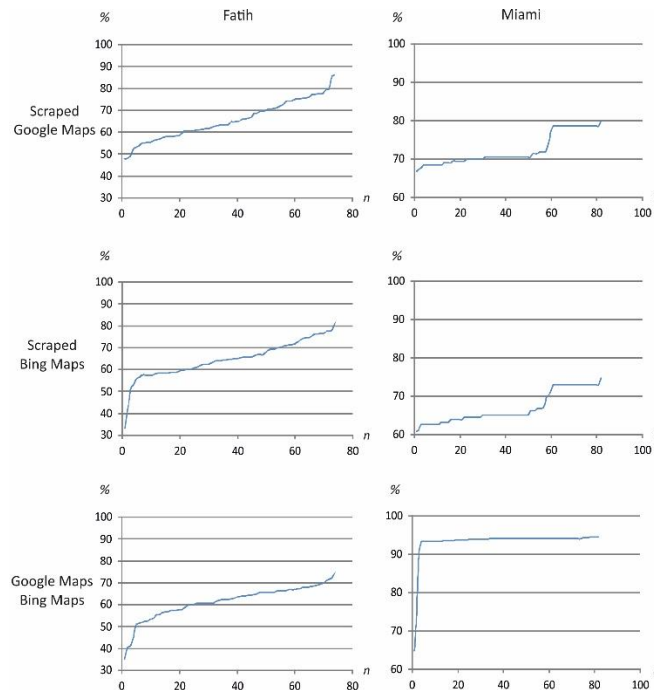


Figure 8. Address similarities for geocoding between a) Scraped and Google Maps data, b) Scraped and Bing Maps data c) Google Maps and Bing Maps data.

The similarity values between the postal addresses obtained with the geocoding of both services in Table 7 have significant differences of approximately 30% for Fatih and Miami Beach, individually. The main reason for the deficiency of geocoding in Fatih is that normalized addresses of both services cannot provide proper and sufficient information for sub-regions such as neighborhood, county, and city. Besides, Google Maps can return a neighborhood name that has not been parsed from a postal address, and uses various abbreviations for road types, while Bing Maps does not present any neighborhood information and abbreviations. As both services use similar abbreviations in Florida, such as Avenue: Ave, Street: St, Florida: FL, and present the addresses following the USA standard format, similarity values of their geocoding results are significantly high. The main factor that prevents reaching a similarity rate of 100% between the results in Google Maps and Bing Maps is the fact that Bing Maps service does not present country information in its normalized addresses.

5. CONCLUSION

The postal address is one of the most crucial input data used in the geocoding process. In this study, geocoding from postal addresses are comparatively analyzed through Google Maps and Bing Maps services. Test conducted in two individual study regions suggests that those services have different geocoding capabilities in Turkey and the USA. The positional accuracy of geocoding in the USA is almost similar for Google Maps and Bing Maps online geocoding services, whereas, in Turkey, Google Maps provides a few better-geocoding results than Bing Maps does.

One of the most critical factors that increase the

quality of geocoding is the actuality of standard addresses stored in the database of services. The addresses in databases of each service are defined according to their standards. The address formats of both services in the USA have almost similar standards, yet they may be quite different in Turkey (e.g., the regular use of an abbreviation or the order of a parsed element in a normalized address). The discrepancies among geocoding services in Turkey concerning address standardization still exist among the public institutions. Many of them use their address format and share information with other stakeholders. It is a complex problem that needs to be solved in collaboration with the official institutions and local governments in Turkey as soon as possible.

One of the main problems for the online geocoding process is to determine the actual location of the point belonging to an address on a geospatial database. The correct location is the front edge of a building, depending on the actual status of the main door used to enter the building. However, the services mark the address points of the buildings generally, regardless of the main entrances. In the future, the authors aim to carry out a study on deriving the entrances of buildings automatically to determine the actual location of address points provided by the services more appropriately. Thus, the quality of the geocoding will be improved.

REFERENCES

- Armstrong, M.P. and Tiwari, C. (2008). Geocoding methods, materials, and first steps toward a geocoding error budget. In Rushton G, Armstrong M P, Gittler J, Greene B R, Pavlik C E, West M M, and Zimmerman D

- L (eds) Geocoding Health Data: The Use of Geographic Codes in Cancer Prevention and Control, Research and Practice. Boca Raton, FL, CRC Press: 11–35.
- Bichler, G. and Balchak, S. (2007). Address matching bias: Ignorance is not bliss. Policing: An International Journal of Police Strategies & Management, 30 (1), 32–60.
- Bilgi, S., Gulnerman, A., Arslanoğlu, B., Karaman, H. and Ozturk, O. (2019). Complexity measures of sport amenities allocation in urban area by metric entropy and public demand compatibility. International Journal of Engineering and Geosciences, 4 (3), 141–148.
- Cayo, M.R. and Talbot, T.O. (2003). Positional error in automated geocoding of residential addresses. International Journal of Health Geographics, 2 (10).
- Cetl, V., Kliment, T. and Jogun, T. (2016). A comparison of address geocoding techniques—a case study of the city of Zagreb, Croatia. Survey Review, 50 (359), 97–106.
- Chow, T.E., Dede-Bamfo, N. and Dahal, K.R. (2016). Geographic disparity of positional errors and matching rate of residential addresses among geocoding solutions. Annals of GIS, 22 (1), 29–42.
- Goldberg, D.W., Ballard, M., Boyd, J.H., Mullan, N. Garfield, C., Rosman, D., Ferrante, A.M. and Semmens, J.B. (2013). An evaluation framework for comparing geocoding systems. International Journal of Health Geographics, 12 (1), 50.
- Haines, E. (1994). Point in polygon strategies. Graphics gems IV, 994, 24–26.
- Hart, T.C. and Zandbergen, P.A. (2013). Reference data and geocoding quality: Examining completeness and positional accuracy of street geocoded crime incidents. Policing: An International Journal of Police Strategies & Management, 36 (2), 263–294.
- Karimi, H.A., Durcik, M. and Rasdorf, W. (2004). Evaluation of uncertainties associated with geocoding techniques. Computer-Aided Civil and Infrastructure Engineering, 19 (3), 170–185.
- Kilic, B. and Gulgen F. (2017). Coğrafi Kodlama için Adres Standardizasyonu. 16. Türkiye Harita Bilimsel ve Teknik Kurultayı. 3-6 May, Ankara, Turkey.
- Kilic, B. and Gulgen F. (2019). Türkiye’de Kullanılan Posta Adres Bilgilerinde Uyum Düzeylerinin Belirlenmesi. Harita Dergisi, 161, 26–34.
- Levenshtein, V.I. (1966). Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. Soviet Physics Doklady, 10 (8), 707–710.
- Li, B., Zhang, X. and Chen, Y. (2010). Automatic construction and visualization of address models. In 2010 Sixth International Conference on Natural Computation (Vol. 6, pp. 2894-2897), IEEE.
- McElroy, J.A., Remington, P.L., Trentham-Dietz, A., Robert, S.A. and Newcomb, P.A. (2003). Geocoding addresses from a large population-based study: lessons learned. Epidemiology, 14 (4), 399–407.
- Mitchell, R. (2015). Web scraping with Python: collecting data from the modern web. O’Reilly Media, Inc.
- Mulazimoglu, E. and Basaraner, M. (2019). User-centred design and evaluation of multimodal tourist maps. International Journal of Engineering and Geosciences, 4 (3), 115–128.
- Qin, X., Parker, S., Liu, Y., Graettinger, A.J. and Forde, S. (2013). Intelligent geocoding system to locate traffic crashes. Accident Analysis & Prevention, 50, 1034–1041.
- Ratcliffe, J.H. (2004). Geocoding crime and a first estimate of a minimum acceptable hit rate. International Journal of Geographical Information Science, 18 (1), 61–72.
- Rodrigues, F. (2010). POI Mining and Generation. PhD Thesis. Faculty of Sciences and Technology (FCTUC) Department of Informatics Engineering - University of Coimbra. Portugal.
- Roongpiboonsopit, D. and Karimi, H.A. (2010a). Comparative evaluation and analysis of online geocoding services. International Journal of Geographical Information Science, 24 (7), 1081–1100.
- Roongpiboonsopit, D. and Karimi, H.A. (2010b). Quality assessment of online street and rooftop geocoding services. Cartography and Geographic Information Science, 37 (4), 301–318.
- Rushton, G., Armstrong, M.P., Gittler, J., Greene, B.R., Pavlik, C.E., West, M.M. and Zimmerman, D.L. (2006). Geocoding in cancer research: a review. American Journal of Preventive Medicine, 30 (2), S16–S24.
- Tian, Q., Ren, F., Hu, T., Liu, J., Li, R. and Du, Q. (2016). Using an Optimized Chinese Address Matching Method to Develop a Geocoding Service: A Case Study of Shenzhen, China. ISPRS International Journal of Geo-Information, 5 (5), 65.
- Ward, M.H., Nuckols, J.R., Giglierano, J., Bonner, M.R., Wolter, C., Airola, M., Mix, W., Colt, J.S. and Hartge, P. (2005). Positional accuracy of two methods of geocoding. Epidemiology, 16 (4), 542–547.
- Yang, D.H., Bilaver, L.M., Hayes, O. and Goerge, R. (2004). Improving geocoding practices: evaluation of geocoding tools. Journal of Medical Systems, 28 (4), 361–370.
- Yildirim, V., Yomralioglu, T., Nisanci, R. and Inan, H. (2014). Turkish street addressing system and geocoding challenges. Proceedings of the Institution of Civil Engineers, 167 (2), 99–107.

Zandbergen, P.A. and Green, J.W. (2007). Error and bias in determining exposure potential of children at school locations using proximity-based GIS techniques. *Environmental Health Perspectives*, 115 (9), 1363–1370.

Zandbergen, P.A. (2008). A comparison of address point, parcel and street geocoding techniques. *Computers, Environment and Urban Systems*, 32 (3), 214–232.

Zandbergen, P.A. (2011). Influence of street reference data on geocoding quality. *Geocarto International*, 26 (1), 35–47.

URL 1. Google Maps Platform – Geocoding API documentation, <https://developers.google.com/maps/documentation/geocoding/start?hl+=de>, accessed on 3 October 2019.

URL 2. Bing Maps Dev Center – Find location by address (geocode), <https://www.bing.com/api/maps/sdk/mapcontrol/isdk/searchbyaddress>, accessed on 3 October 2019.