A DIFFERENT WEB-BASED GEOCODING SERVICE USING FUZZY TECHNIQUES

P. Pahlavani [a], R.A. Abbaspour [a], A. Zare Zadiny [a], *

[a] Dept. of Surveying and Geomatics Eng., College of Engineering, University of Tehran, Tehran, Iran
(Pahlavani, Abaspour, Zare_zardiny)@ut.ac.ir

Commission VI, WG VI/4

KEY WORDS: Geocoding, Address, Fuzzy, Nearness, Overlay Function

ABSTRACT:

Geocoding - the process of finding position based on descriptive data such as address or postal code - is considered as one of the most commonly used spatial analyses. Many online map providers such as Google Maps, Bing Maps and Yahoo Maps present geocoding as one of their basic capabilities. Despite the diversity of geocoding services, users usually face some limitations when they use available online geocoding services. In existing geocoding services, proximity and nearness concept is not modelled appropriately as well as these services search address only by address matching based on descriptive data. In addition there are also some limitations in display searching results. Resolving these limitations can enhance efficiency of the existing geocoding services. This paper proposes the idea of integrating fuzzy technique with geocoding process to resolve these limitations. In order to implement the proposed method, a web-based system is designed. In proposed method, nearness to places is defined by fuzzy membership functions and multiple fuzzy distance maps are created. Then these fuzzy distance maps are integrated using fuzzy overlay technique for obtain the results. Proposed methods provides different capabilities for users such as ability to search multi-part addresses, searching places based on their location, non-point representation of results as well as displaying search results based on their priority.

1. INTRODUCTION

Geocoding - the process of finding position based on descriptive data such as address or postal code - is considered as one of the most commonly used spatial analyses. Growth of cities on one hand and increase access to internet with the ability to display interactive maps on mobile phones on the other hand, has increased general users interest in geocoding. This subject has motivated online map providers to present geocoding as one of their basic capabilities. Online geocoding services such as Google Maps, Bing Maps, Yahoo Maps, Map Quest, Geocoder US and Open Route Service and so on are among the most popular services of this kind. Despite the diversity of geocoding services, users usually face limitations using them. In linguistic geocoding among humans, proximity to a certain location has critical importance, for example a user wants to find the banks near to a specific location; however in existing geocoding services this concept is not modelled appropriately. In geocoding services, the address entered by the user is only matched with the corresponding address in database in order to search a place. In fact, this search method is based on descriptive data (and not on spatial data) which leads to inability of services to consider spatial proximity in address finding. For example, when a user searches the phrase "Fatemi-Bank" in Google Maps, only the banks that their descriptive address contains "Fatemi" are shown. However, a bank which is in 100 meters from Fatemi Street but doesn't probably have "Fatemi" in its address is not shown for the user (Figure1).

In addition, existing geocoding services enable users to only search one- or two-part addresses while they don't show any results if multipart addresses (over two parts) are entered by users. As well as, there are also some limitations in display results. Existing services, display searching results as a point,

regardless what a user searched (a street, an area or a point of interest). Moreover, geocoding services do not usually consider any priorities when they display the results.

Figure 1. Result of Search "Fatemi-Bank" in Google Maps

Resolving these limitations can enhance efficiency of the above mentioned services. This paper proposes the fuzzy technique in order to deal with the existing limitations. Accordingly, studies have been performed in this field are reviewed first and then a new algorithm is investigated for fuzzy geocoding. Next, a web-based system is designed and implemented according to this proposed method. Finally the most significant results are presented.

2. LITERATURE REVIEW

Various methods have been presented for geocoding so far. In existing geocoding process, the address is found only based on address matching. Therefor many researches have worked on

---

* Corresponding author

efficiency of this matching. For example Peter Christen et.al in [1] described a geocoding system that used a learning address parser based on Hidden Markov Models to separate free-form address into components, and a rule-based matching engine to determine the best set of candidate matches to a reference file. In another case, Daras K et.al in [2] presented a fuzzy matching algorithms for geocoding historical addresses.

These algorithms and methods usually use fuzzy concept only for process the descriptive data (address matching). But there are not any researches that use fuzzy concept on spatial data processing in geocoding. In this paper, fuzzy technique is used in related to spatial data in addition to descriptive data.

3. PROPOSED METHOD

In this paper fuzzy technique is used in order to deal with the limitations mentioned in introduction. Fuzzy technique can model proximity concept in address finding and remove the limitations in display the searching results. In this paper, nearness to places is defined by fuzzy membership functions and multiple fuzzy distance maps are created. Then these fuzzy maps are integrated using fuzzy overlay technique for obtain the results. Fuzzy overlay is used to determine the locations that best meet the nearness criteria, that is have high likelihood of membership in all fuzzy distance maps. In order to implement the proposed method, a web-based system is designed. The overall structure of this system is displayed in figure 2.
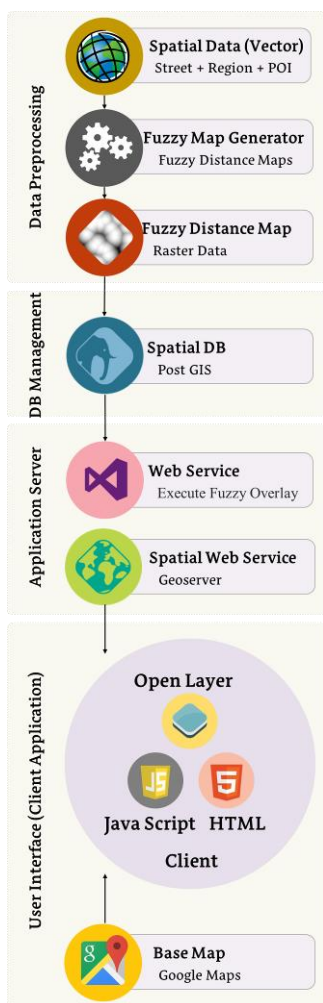
Figure 2. Proposed method

This proposed method contains four main parts. These part are data pre-processing, database management, application server and user interface. In continue each parts of proposed method is explained.

3.1   Data Pre-Processing

Data that used for implementation are related to Tehran Sixth district including network of roads and streets (linear layer), urban areas (polygon layer) and point of interest such as schools, hospitals (point layer). These data are in shape file format. For using data in this method, spatial data must be imported in spatial database but before it, these data need to pre-process. Data pre-processing include of two step. First step is create distance map. In this step, the Euclidean distance map is created based on existing spatial data. These maps show degrees of proximity to the related data discretely. The maximum distance parameter in create of distance map is variable depending on geometric of data. This distance is 50, 70, and 100 meters for point, polyline, and polygon data layers respectively. Figure 3 display an example of distance map for a point layer.
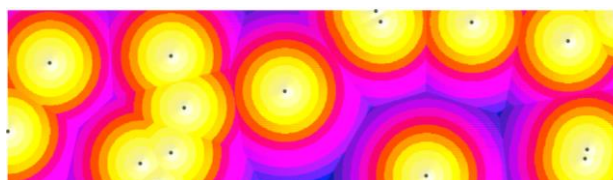
Figure 3. Euclidean Distance Map

Nearness concept is defined with definition of a fuzzy system in step two. This process is performed through introduction of a fuzzy membership function. Introduced membership function is applied on Euclidean distance map. In this implementation for create fuzzy distance map, the Gaussian membership function is selected. A Gaussian membership function is specified by two parameters {c, σ}

$$\text{gaussian}(x;c,\sigma) = \exp[-\frac{(x-c)^2}{2\sigma^2}] \qquad (1)$$

A Gaussian membership function is determined by c and σ; c represents the membership function centre and σ determines the membership function width. In this implementation width and midpoint equal to 0.1 and 0. Figure 4 displays an example of fuzzy distance map for a point layer.
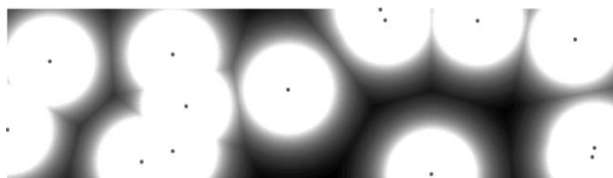
Figure 4. Fuzzy Distance Map

In this system, a windows form application is implemented in Microsoft Visual Studio 2010 to create the distance map and fuzzy distance maps from each vector data and finally save the results as tiff files.

3.2 Database Management

After Data pre-processing phase, for each spatial data including streets, urban areas and point of interest a fuzzy distance map is generated. The address finding process is done based on these raster data. For implementation the web based system, it is necessary to load data in spatial database. The database that used for this implementation is PostGIS. For import raster data (fuzzy distance maps) to PostGIS database, raster2pgsql loading tools is used. Raster2pgsql converts a raster file into a series of SQL commands that can be loaded into a database. The output of this command may be captured into a SQL file, or piped to the psql command, which will execute the commands against a target database. After loading raster data in database, visual check of the raster data is done with Quantum GIS.

3.3 Application Server

When user want to find an address in this system, client application sends a request for application server. This request contains the address. Target address including one or more parts that these parts are separated by '-'. Application server selects data related to each part of address by an address matching. "Like" command is used for selection the spatial data corresponding to each part of address entered by user. While using this command, there is no need to enter the exact name equal to what exists in database.

In this paper, an ASP Web Service is implemented as application server. This web service connects to Post GIS and selects the data (fuzzy distance map) related to each part of address and finally integrated these fuzzy distance maps using fuzzy overlay technique. Fuzzy overlay, combine fuzzy membership raster's data together, based on selected overlay function. The following lists described the most commonly used fuzzy overlay function:

- The Fuzzy AND overlay function will return the minimum value of the sets the cell location belongs to. This technique is useful when we want to identify the least common denominator for the membership of all the input criteria. Fuzzy AND uses the following function (Equation 2) in the evaluation:

$$\mu_{AND} = \min(\mu_1, \mu_2, ..., \mu_n) \qquad (2)$$

- The Fuzzy OR overlay function will return the maximum value of the sets the cell location belongs to. This technique is useful when we want to identify the highest membership values for any of the input criteria. Fuzzy OR uses the following function (Equation 3) in the evaluation:

$$\mu_{OR} = \max(\mu_1, \mu_2, ..., \mu_n) \qquad (3)$$

- The Fuzzy Product overlay function will, for each cell, multiply each of the fuzzy values for all the input criteria. The resulting product will be less than any of the input, and when a member of many sets is input, the value can be very small. It is difficult to correlate the product of all the input criteria to the relative relationship of the values. Fuzzy Product uses the following function (Equation 4) in the evaluation:

$$\mu_{Multiplication} = \prod_{i=1}^{n} \mu_i \qquad (4)$$

- The Fuzzy Sum overlay function will add the fuzzy values of each set the cell location belongs to. The resulting sum is an increasing linear combination function that is based on the number of criteria entered into the analysis. Fuzzy Sum uses the following function (Equation 5) in the evaluation:

$$\mu_{Summation} = 1 - \prod_{i=1}^{n} (1 - \mu_i) \qquad (5)$$

- The Fuzzy Gamma function is an algebraic product of Fuzzy Product and Fuzzy Sum, which are both raised to the power of gamma. The generalize function is as Equation 6. When Gamma is 1 the result is the same as Fuzzy Sum. When Gamma is 0 the result is the same as Fuzzy Product.

$$\mu_{Combination} = (\mu_{Summation})^{\gamma} \times (\mu_{Multiplication})^{1-\gamma} \qquad (6)$$

In this implementation, for integration of all Fuzzy distance maps, Gamma overlay function is used with different values for γ coefficient. For each γ value, a new result is obtained. In this implementation γ coefficient can be changed by the user optionally. Application server perform Gamma function on selected raster data (fuzzy distance maps) by ST_MapAlgebra command. ST_MapAlgebra, returns a one-band raster given one or two input raster's, band indexes and one or more user-specified SQL expressions. An algebraic expression involving the two raster's and Post GIS defined functions/operators that will define the result pixel value. After executing the ST_MapAlgebra command, a raster is generated and store in database. Finally, Geoserver as a spatial web service received the target raster from Post GIS and sent it for client.

3.4 User Interface

For implementation the client application, the Open Layers 3 is customized. Open Layers is an open source java script library to load, display and render maps from multiple sources on web pages based on HTML5 and CSS3. Figure 5 display a snapshot of client application.
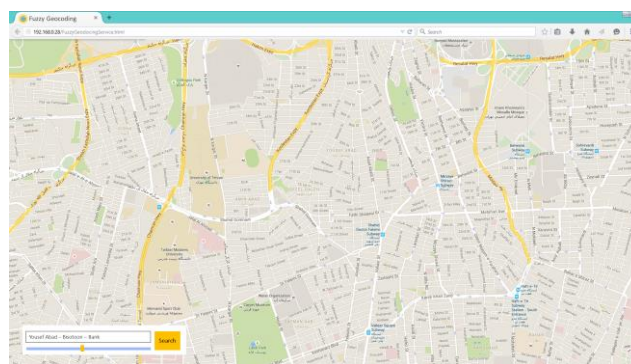

Figure 5. User Interface of Web-based Geocoding System

User entered the address in textbox and select γ coefficient. Client application sent address for ASP web service and receive result from Geoserver as a WMS Layer. Finally the result is displayed on base map (Google Maps).

4. EXPERIMENTAL RESULT

The following scenarios show some instances of system outputs:

4.1 Scenario 1

Suppose that a user wants to search "Karim khan – Bank" in the system. Figure 6 displays the output of this search for $\gamma = 0.7$.
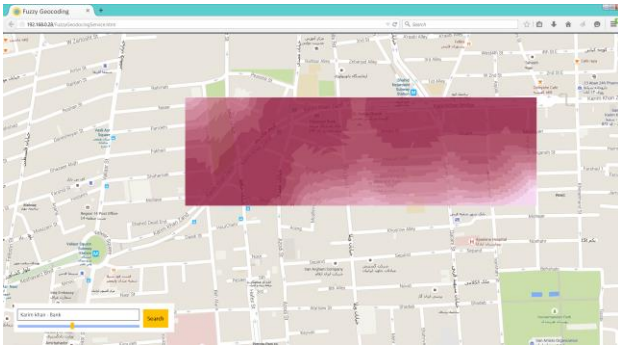

Figure 6. Output of Scenario 1

4.2 Scenario 2

For evaluation, the outputs of proposed method can be compared with output of most commonly used online geocoding services such as Google Maps, Yahoo Maps, Tehran.ir maps and Open Route Service. Result of find "Karim khan - Bank" in these geocoding services is considered in figures 7-10.
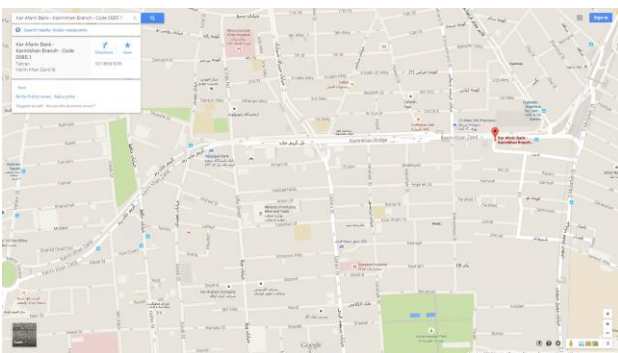
Figure 7. Output of Search Address in Google Maps
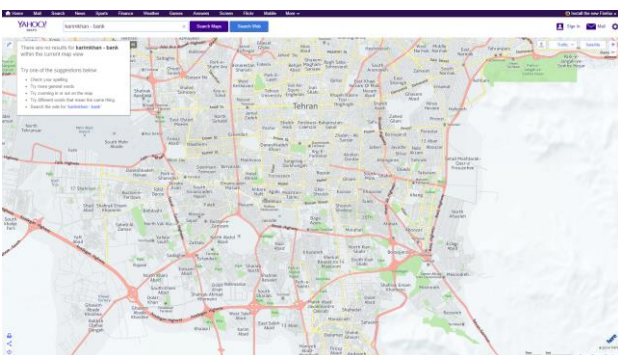
Figure 8. Output of Search Address in Yahoo Maps
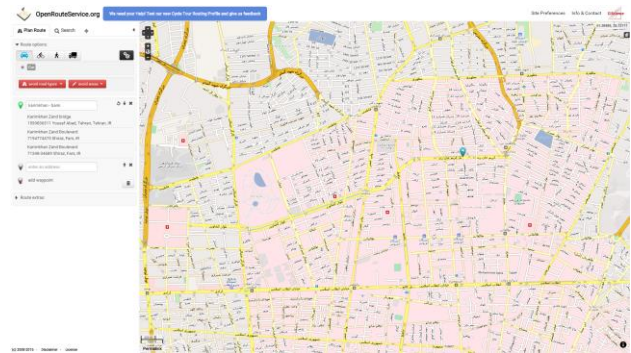
Figure 9. Output of Search Address in Open Route Service

Figure 10. Output of search address in Tehran.ir maps

4.3 Scenario 3

User can change γ coefficient and receive different outputs from the application service. The figure 11 indicates system outputs for different γ coefficient.

5. CONCLUSION

As it was mentioned in introduction, users face some limitations when they use available online geocoding services. This paper proposes the idea of integrating fuzzy technique with geocoding process to resolve these limitations. Accordingly, a new algorithm is proposed for geocoding, while a web-based system is also implemented. The system provides the following capabilities for users:

- The ability to search multi-part addresses: Users can search a multi-part address including district, street, and a specific location's name in the system.
- Searching places based on their location: With defining nearness concept in this service, it is possible to search address based on nearness to a specific location.
- Non-point representation of results: As it was observed in implementation outputs, search results are shown as an area for the user.
- Displaying search results based on their priority: A specific priority has been defined based on different degrees of transparency and according to implementation outputs so that highlighted areas represent high priority search results.

Integration of this service with existing geocoding services available can enhance them. Moreover, defining a weight for each part of an address can provide more accurate and closer results to what users want. Furthermore, system speed

improvement can be suggested as a research focus for future investigations.
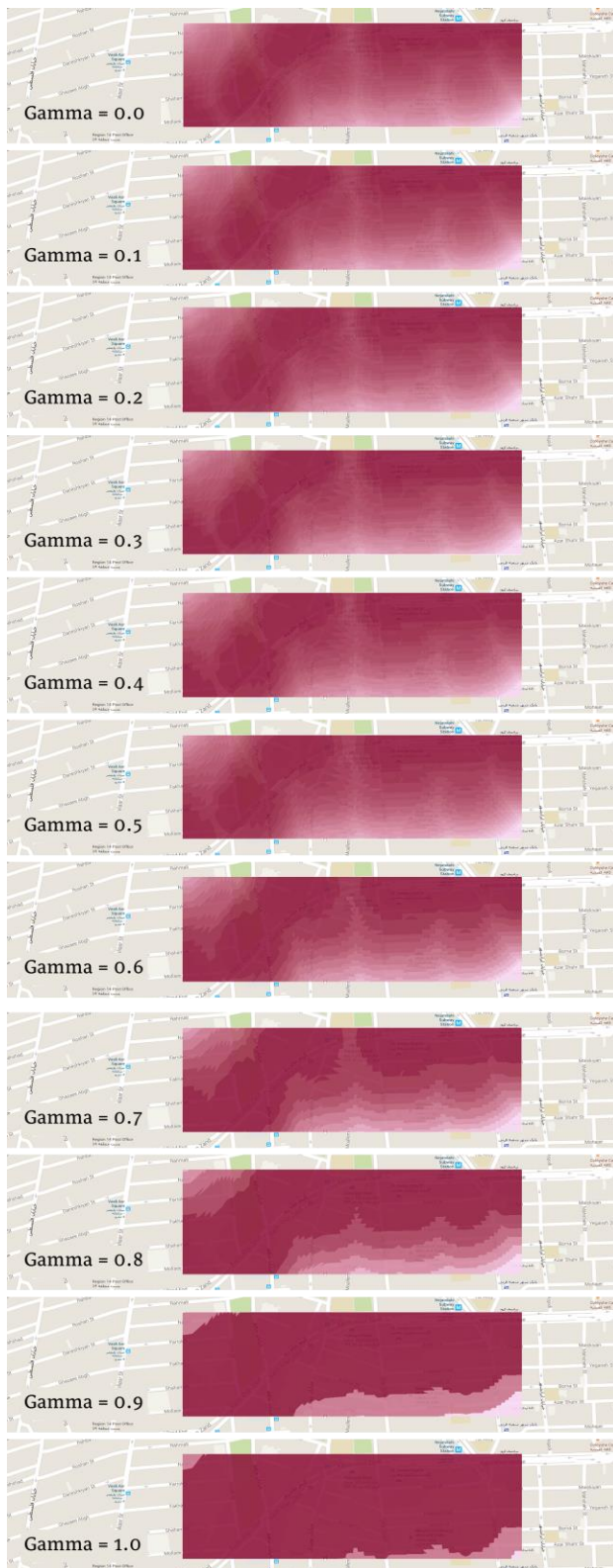

Figure 11. Different Outputs for Different γ (from 0 to 1)

REFERENCES

Peter Christen, Tim churches and Alan Willmore. A Probabilistic Fuzzy Geocoding system based on a National Address File.

Daras K, Feng Z and Dibben C. 2014. HAG-GIS: A spatial framework for geocoding historical addresses.

PostGIS 2.1.9dev Manual, the PostGIS Development Group Chapter 5. Raster Data Management, Queries, and Applications

Web Sites:

- https://maps.yahoo.com
- http://openrouteservice.org
- https://maps.google.com
- http://map.tehran.ir
- http:// openlayers.org

Revised July 2015