Original Article

Research and Optimize Geocoding with the User's Coordinates on VMap

Tran Van Trung*, Tran Anh Tu, Nguyen Phuong Thao, Luu Quang Thang, Pham Bao Son, Bui Quang Hung

VNU Center of Multidisciplinary Integrated Technologies for Field Monitoring University of Engineering and Technology, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam

Received 08 November 2020 Revised 07 February 2021; Accepted 29 March 2021

Abstract: In this article, we present an optimal method for searching the address on VMap based on the current location and the entire view of the users. The article will focus on understanding and researching factors to optimize user's search, enhancing the user’s experience on VMap. The article used data from OpenStreetMap, open source of Pelias geocoder to run test suites. Specifically, based on comparison results with the best rated digital map platform today, Google Maps, we calculated the effectiveness of the method. The proposed method helps to diversify search results with the user's current location, providing relevant search results in case the user's view does not have much influence on search optimization (such as user's view too large or too small). From there, as the optimal basis for searching of many other locations other than the user's current location and view, location bias priority will be enhanced. Besides, evaluation methods need to be performed on multiple users to evaluate effectiveness. This method was applied for VMap and showed optimistic results.

Keywords: QUERE, geocoding, user’s coordinates, digital maps, VMap, GIS.

1. Introduction

Map means a general model representing geographical features at a certain scale, according to mathematical rules, with a conventional symbol system, and based on the results of processing information and data acquired from the survey price [1]. According to a statistic about number of monthly map-users in the US in April 2018, top 5 popular map services are: Google Maps (72%), Waze (12%), Apple Maps (11%), Mapquest (10%), and Google Earth (2%) [2, 3]. However, those map services still have not provided official services for Vietnam and have many limitations when being deployed in Vietnam: i) Low accuracy; ii) Incomplete number of address data; iii) Limited functionality; iv) Numerous legal constraints; and v) High maintenance costs. In Vietnamese market, the top 5 companies providing map services are: VietMap (with 2.5
million addresses), CocCoc (with 1.2 million service points), DDG (with 1 million addresses), Map4D and Vietbando. Map services in Vietnam are incomplete and have not met the user's demand for: i) Accuracy; ii) Number of addresses data; iii) User interface and experience not satisfied; iv) Non detailed features; e) Having no effective Software Development Kit (SDK). According to the Vietnam Population and Housing census of the General Statistics Office of VietNam in 2019, the total number of households in Viet Nam is 26,870,079 [4]. A statistic in August 2020 announced that Google Maps has 1.2 million location displayed data, VietMap has at most 2.5 million address data. Vietnam needs a map platform that meets the needs of users for map features and functions, improves the accuracy with a large amount of address data, and has a friendly interface that is suitable for the user’s experience in different fields.

The Vietnam Digital Map Project-VMap-aims to create a platform for collecting and sharing map data (coordinates) of landmarks and addresses nationwide, which is the platform for businesses to use to create applications in many fields: education, culture, tourism,... [5]. Currently, VMap has 23.4 million address data collected from 6 sources: i) Ministry of Natural Resources and Environment; The Department of Survey, Mapping and Geographic Information Vietnam (headquarters of People's Committee at all levels; residential landmarks, streets; borders, boundaries); ii) Ministry of Culture, Sports and Tourism (accommodation); iii) Ministry of Health (medical institutions); iv) Ministry of Education and Training (institution education and training); v) Vietnam Post (VNPost) (post office); vi) Field data collected by more than 120,000 postal workers and youth union members. VMap provides the following services: i) Mapping; ii) Geocoding; iii) Routing. VMap is focusing on developing and integrating algorithms to optimize service and enhance user’s experience.

Address search is a geocoding component. The input data is the user’s typing keywords that they want to query, and the result is a list of matching addresses based on many factors. We did a survey with 180 test cases. This test suite is used to test on Google Maps, Vietbando, VMap which included: different keywords cases, different languages cases (Vietnamese and English), different typing cases (for Vietnamese: signed and unsigned), the user’s view changed cases, zoom level changed cases (zoom in and zoom out). Then, we determined 4 factors used to optimize the search results, including: i) Natural language similarity; ii) Semantic; iii) User’s history; and iv) Spatial coordinates (user’s location and user’s view).

There have been studies on the factors affecting geocoding and the application of these factors in solving practical problems. Natural language processing is one of the most popular factors used to optimize the search results. Studies show that including the normalization of address data and address parsing have a positive effect on geocoding [6, 7]. Küçük Matci and Avdan had present a study seeking to address problems such as inconsistencies in address formats, including inaccurate numbering systems, misspellings, the use of abbreviations and a lack of data that refers to the geocoding process by way of a standardization process [7]. To that end, it employs a method that decomposes addresses used as input data in geocoding, identifies spelling mistakes and abbreviations, and reorganizes the addresses through the Natural Language Process (NLP). Geocoding in Google API and ArcGIS is performed on the re-formatted addresses and the results compared to the non-standardized results. Lambert, Francisco and Lerner had presented invention pertains to enhancement or refinement of estimated locations based upon user-specific information, extracted from a number of user-related sources, including the web browser history, search history, maps history, address book, e-mail archives and calendar entries, used to refine an initial location estimate of the user [8]. This may be done by determining whether one or more positions provide a more accurate position of the user than the initial estimate, used to provide enhanced driving directions to the user.
Spatial coordinates are used in optimal searches as a location bias. Locations that are close to a given coordinate will be ranked higher in the search results. Several types of coordinates are used as the priority of the user's current location, the preference of the user's view center,... Only Google Maps has used the user’s current location to optimize the search results and is appreciated by users, so it has become popular. The method presented in this article combined user’s current location and user’s view factors to optimize VMap's search results.

2. Related Work

2.1. Survey, Compare VMap and Google Maps

In Viet Nam, Google Maps is the most popular map service. This article made a comparison between Google Maps platform (Web browser version) and VMap platform (Web browser version) before applying the optimal method to evaluate the influence of the user's location in optimizing search results. The places used in the evaluation suite are: i) common places such as “trường học” (school), “bệnh viện” (hospital), “chợ” (market),...; ii) Famous places in Viet Nam such as “Lăng Bác” (Uncle Ho's Mausoleum), “Cung đình Huế” (Hue Palace),...; and iii) Administrative areas such as “thành phố Hà Nội” (Hanoi city), district,... The article also calculated the distribution of search results of Google Maps and present VMap (applied the optimal method) to compare the similarity in search results between the two platforms.

Survey with the user’s current location in Hanoi while the user’s view with the center point in Ho Chi Minh City return 2 cases of results: i) Case 1, zoom level is 5 (user’s position is visible), Google Maps search results (Figure 1.a) are distributed near the user’s location, while VMap's search results (Figure 1.b) are distributed near the center point of the user’s view; and ii) Case 2, zoom level is 8 (user’s position is invisible), Google Maps search result (Figure 1.c) is distributed near the center point of the user's view and VMap's result (Figure 1.d) is the same.

The survey’s results showed that Google Maps optimizes address search based on the user’s current location (point-optimized) and the user’s view (region-optimized). Vmap applied an optimal method using the center point of the user’s view (point-optimized), without using the user's current location and region-based optimization.

a) Google Maps searches by user’s location.
b) VMap does not search by user’s location.

c) Google Maps searches by user’s view.

d) VMap searches by user’s view.

Figure 1. Address search survey with Google Maps (a, c) and VMap (b, d).
Google Maps prioritizes search results by user’s location in the following cases: i) The user’s view has great coverage that does not make much sense in optimizing the search results, ii) The users appear in the view, iii) The user’s position is invisible but close to the view.

At present, VMap cannot combine two optimal factors at the same time for searching the address. The regional optimization (entire user’s view) still had not applied in VMap's address search engine. When the view is changed to large or small, or the zoom level is changed, the results are returned without any change. Searching results are focused only in the user’s view and not returned results in the user’s current location. VMap's search results are not really diverse, have not met the search needs in many cases and still have not enhanced the address search experience for users.

The optimal method presented in the article that developed for Vmap was: i) Applying VMap's point-of-point search optimization method for the user’s location; ii) Building an optimal method of address search according to the user’s entire view; iii) combining 2 above methods to optimize address search results.

2.2. Current Status of VMap's Search System and Proposed Improved Method

One of the input data of the search system is a text string entered through the online interface of the VMap on the user’s client. Input string may contain keywords, words or phrases about the desired location that users want to search or suggest. Another input data is the user’s view center point coordinates, including the longitude and latitude as decimal numbers. From this center point, it indicates the interested areas of users, or identifies relevant areas similar to the user’s demand.

The performed process is: i) standardized inputs: parsing the address into administrative levels and searching objects, standardized longitude in the range (180, -180) and latitude in the range (90, -90); ii) Query from standardized data and searching; iii) Packing the search results as JSON, return and display on the user’s interface. The query is JSON format and sent directly to the search engine. The Elasticsearch search engine searches and calculates relevant points, distances between locations.

Figure 2. Sequence diagram of improved VMap search system.
The output of the search system is a list of suggested and searched results, that satisfies the user’s keywords and filter conditions. The output is also sorted by related factors, packed in layers, added with necessary data fields such as distance, or type of unnecessary data, error data or loss data. Results are prioritized according to the user’s view center point: the further location, the lower priority.

The innovative system standardizes the user’s current position coordinate data and the coordinates of 4 angles of the user’s view. Then, the system runs the search query with the coordinates of the user’s current location and the coordinates of the user’s view. Other steps in the process remain the same as VMap’s search system architecture before. Details about methods of standardizing and creating queries will be presented in the next section.

3. Proposed Method

3.1. The Optimal Searching Method by User’s Location Coordinates on the Vmap

Optimizing search by coordinates uses a pair of longitude-latitude as the input. The old version of VMap used the view center point and improved to the current user’s location. Results are the relevance score of the locations according to the user’s current location calculated in Elasticsearch by Decay Function [9] by using the Exponential function [10]. Closer locations will have higher scores while further locations will have lower scores (Figure 3).

Formula of exponential function for optimize search by point (*):

\[ S_i = \exp\left(\lambda_i \cdot \max(0, D)\right) \]

where

\[ \lambda_i = \frac{\ln(\text{decay}_i)}{\text{scale}_i} \]

and

\[ D = \text{geo}_\text{distance}_i - \text{offset}_i \]

The parameters of the formula are kept the same with \( \text{offset}_i = 0\text{km} \) (defined to compute the decay function for ranked location with a distance greater than the defined \( \text{offset}_i \)), \( \text{scale}_i = 50\text{km} \) (defines the distance from user’s current location plus offset at which the computed score will equal \( \text{decay}_i \)), \( \text{decay}_i = 0.5 \) (defines how ranked location are scored at the distance given at \( \text{scale}_i \)), \( \text{geo}_\text{distance}_i \) (geographical distance from the ranked location to the user’s current location), all measured in kilometers.
3.2. Construct the Optimal Method by the Entire View on the Vmap

Region-optimized search is different from point-optimized search in terms of the number of input parameters and parameter values of the relevance score formula. The input parameter of the region-optimized formula is the coordinates of 4 corners of view. From those coordinates, it will be used to calculate the center point and the circumference radius (R) that is the $scale_2$ parameter. Since the user’s view is usually an isosceles quadrilateral, the center and radius of the circumference of the view can be easily determined. Parameters $offset_2$ and $decay_2$ are kept with values of 0km and 0.5. The $geo\_distance_2$ parameter is geographical distance from the ranked location to the user’s view center point coordinates. All parameters are measured in kilometers. The outputs are locations according to the relevant thresholds. The locations in the view and near the center of the view have high priority.

Reapplying the formula of exponential function for optimize the search by view center:

$$S_2' = \exp(\lambda_2 \cdot \max(0, D))$$

where

$$\lambda_2 = \ln(\frac{Decay_2}{scale_2})$$

and

$$D = geo\_distance_2 \cdot offset_2$$

The scoring formula of the region optimization gives an equal score for all locations within the same area of the user’s view radius. The scores of locations in the outer radius area are half that of locations in the nearest inner radius. The radius areas are limited by radii that are multiples of the $scale_2$ parameter, like $[0, scale_2), [scale_2, 2\cdot scale_2)$ ...

The formula calculates the exponential score of the $decay_2$ parameter and the integer of the quotient between $geo\_distance$ and $scale_2$.

As expected, the ranked locations within radius of $scale_2$ from user’s view center point will have score of 1 ($decay_2^0$), locations within radius of $2 \cdot scale_2$ and outside of radius $scale_2$ from user’s view center point will have score of 0.5 ($decay_2^1$), and so on. The formula for calculating the additional parameters for the region-optimized formula:

$$S_2' = decay_2^{\max\left(0, \left\lfloor \frac{D}{scale_2} \right\rfloor \right)}$$

The regional optimization searching formula (**):

$$S_2 = \frac{S_2 + S_2'}{2}$$

![Figure 4. Relevance score distribution by the user’s entire screen view.](image-url)
Score distribution by the user’s entire view is shown by the dashed yellow line in Figure 4, points within the same radius R will have a slight difference, and points within different radius R will have higher, or lower difference that made dashed lines. The locations near the center of the view will have higher priority.

3.3. The Combined Optimal Method of Address Search

The presented method optimizes address search is based on optimal results according to the user’s current location (*) section 3.1) and the user’s entire view (**, section 3.2). In other words, the relevance score of a location includes the scores based on the user’s current location and the entire user’s view. The user’s current location takes higher priority in search optimization. Optimal address search formula by current location and view of users:

\[
S = S_1 + S_2
\]

Table 1. Cases of the desired results of the article

| No | User’s view | Distance | Desired results |
|----|-------------|----------|-----------------|
|    |             |          | User’s location | User’s view |
| 1  | Large       | Large    | X               |
| 2  | Large       | Medium   | X               |
| 3  | Large       | Small    | X               |
| 4  | Medium      | Large    | X               |
| 5  | Medium      | Medium   | X               |
| 6  | Medium      | Small    | X               |
| 7  | Small       | Large    | X               |
| 8  | Small       | Medium   | X               |
| 9  | Small       | Small    | X               |

The locations with the high relevance score will be returned first. The desired result will set priority to the search result according to the user’s current location. In all cases, the location near the users will be returned first. The second priority is the user’s view-points in the view will be returned first. If both the current location and the view of users do not match, matched results on other regions will be returned.

In this article, the view is divided into three main cases. Large panoramic view with large zoom, from 1 to 5 (low level of detail, see all of Vietnam), medium panoramic view with medium zoom, from 6 to 10 (medium level of detail, with city details visible), low panoramic view, with small zoom, 11 onwards. At the same time, the distance from the user’s current position to the user’s view is divided into three cases. Large distance (more than 600km), medium (from 150km to 600km), and small (less than 150km). The results showed that in some cases of the visible user’s current position in the view, it will prioritize the user’s current location. This article shows typical cases which cover most of the user’s search cases. There are 9 cases in total which can cover almost the results of the new address-search method.

4. Experiment

The article integrates address search optimization methodology that combines user’s location and user’s view into VMap's test environment. The accuracy and effectiveness of the method were assessed using the evaluation suite, which compared the search results between Google Maps maps and VMap. The article built 50 test cases for the evaluation suite, based on the Google Maps mapping platform, by taking the user’s current location (Hai Phong city), with view center point of the view respectively are in the city Ho Chi Minh City, Hue City, Thua Thien - Hue, and Hanoi City. The views sequencely covers all of Vietnam, entire region (North, Central, and South, respectively), entire city, a district of that city, a commune, ward or town, the search...
keyword is "kfc". Each time the view is changed, the circumscribed circle of view will change, but the user’s current position will remain. The results returned by Google Maps will be used to compare with the results returned from the improved method to calculate the similarity.

The comparison results between Google Maps and VMap are illustrated in the two below tables. With comparison factors about search results based on the current location is in Hai Phong city, the user’s view in Ho Chi Minh City and Hanoi, including 4-corners’ coordinate of the view. Deviation of the view center point is the distance from the center point of the circumscribed circle of the view, to the center point of view that is displayed to the users by the map. Radius R is the radius of the circumscribed circle of the user’s view, because the user’s view is of an isosceles quadrilateral shape. Search results of Google Maps and VMap in the cases are shown in Table 2 and Table 3.

### Table 2. Results with Ho Chi Minh City view

| User’s Location (Lat - lon) | User’s view | User’s view coordinates (Upper left, lower right coordinates) | Google Maps’s results | VMap’s results |
|----------------------------|-------------|-------------------------------------------------------------|-----------------------|---------------|
| Hai Phong 21.0045767 106.5910793 | Ho Chi Minh | [24.500417, 76.724863], [-2.972366, 136.666265] | Hai Phong, Hai Duong, Quang Ninh, Hanoi | Hai Phong, Hai Noi, Ho Chi Minh |
| Hai Phong 21.0045767 106.5910793 | Ho Chi Minh | [12.550029, 102.949226], [9.086662, 110.441902] | An Giang, Can Tho, Ho Chi Minh | Ho Chi Minh |
| Hai Phong 21.0045767 106.5910793 | Ho Chi Minh | [11.225709, 105.757606], [10.358310, 107.632149] | Ho Chi Minh, Binh Thuan, Binh Duong | Ho Chi Minh |
| Hai Phong 21.0045767 106.5910793 | Ho Chi Minh | [10.810576, 106.637028], [10.756446, 106.754101] | Ho Chi Minh | Ho Chi Minh |
| Hai Phong 21.0045767 106.5910793 | Ho Chi Minh | [10.796728, 106.666296], [10.769621, 106.724832] | Ho Chi Minh | Ho Chi Minh, Hai Phong |

### Table 3. Results with Hanoi City view

| User’s Location (lat - lon) | User’s view | View’s coordinates (Upper left, lower right coordinates) | Google Maps’s results | VMap’s results |
|----------------------------|-------------|---------------------------------------------------------|-----------------------|---------------|
| Hai Phong 21.0045767 - 106.5910793 | Hanoi | [33.480434, 75.820652], [7.682087, 135.805999] | Hanoi, Hai Duong, Hai Phong, Quang Ninh | Hai Phong, Hanoi, Da Nang |
| Hai Phong 21.0045767 - 106.5910793 | Hanoi | [22.711287, 102.088960], [19.420959, 109.581636] | Hanoi, Hai Duong, Hai Phong, Quang Ninh | Hai Phong, Hanoi, Da Nang |
| Hai Phong 21.0045767 - 106.5910793 | Hanoi | [21.457236, 104.897340], [20.633119, 106.771882] | Hanoi, Hai Duong, Hai Phong | Hai Phong, Hanoi, Da Nang |
| Hai Phong 21.0045767 - 106.5910793 | Hanoi | [21.049934, 105.8205987], [21.024178, 105.864566] | Hanoi | Hai Phong, Hanoi, Da Nang |
| Hai Phong 21.0045767 - 106.5910793 | Hanoi | [21.040055, 105.827970], [21.033637, 105.842615] | Hanoi | Hanoi |
Search result similarity is assessed by the rate of the distribution of search results of the Vmap compared to Google Maps. The higher rate indicates that the search method has updated better to the user's search demand. Experimental results show that: i) Vmap's address search method has optimal factors according to user’s location and the entire user’s view; ii) The search results are effective, having high similarity up to 70% when compared to the address search results of Google Maps; and iii) Can combine optimal search in other locations (outside the user’s location and the user’s view). The results include the exact locations according to the user’s current location, the user’s view, or a combination of the two above results. The VMap’s results deviation from Google Maps were divided into two cases. Case 1, the results are in different positions, account for 25% of search results. In case 2, the results are at the same position but the distance is different, the difference in distance is not significant. The reason for the cases with different results and deviation is that the data used in the address search is quite different. Google Maps will have more complete and accurate data. This article used OpenStreetMap data, which is only for experimental methods, so there are changes and differences as mentioned above.

The search result distribution of the combined method was compared with the old searching method of VMap and Google Maps (Figure 5) in two cases: i) The user’s location in Hanoi, the view at the Hue, the zoom level is 6, the result distribution of Google Maps (Figure 5.i) by user’s location, of the combined method (Figure 5.iii) by user’s location, of VMap (Figure 5.v) is by view; and ii) User’s location in Hanoi, view center point of Ho Chi Minh, zoom level is 8, Google Maps result distribution (Figure 5.ii) by view, of combined method (Figure 5.iv) by view, VMap's (Figure 5.vi) by view. From the results distribution seen in Figure 5, the combined search method of this article has better search results, meets the user’s search needs and is closer to the search results of Google Maps.

![i) Result of Google Maps - case 1.](image)
ii) Result of Google Maps - case 2.

iii) Result of combined method - case 1.
iv) Result of combined method - case 2.

v) Result of VMap - case 1.
The new method’s search results, based on the current location and the view of user’s, were able to combine multiple coordinate points into the optimal location search. Specifically, the two factors with the most special meaning are the user’s current position and the user’s view. The search results have clearly demonstrated the effectiveness of the method through comparison with the most popular map today, Google Maps, the results are up to 70% matches. The results show that the new method is highly practical and can be applied to VMap platform.

5. Conclusion

The article presents optimal address search method for VMap based on the combination of the user’s current location and the user’s entire view. The results of the method were tested by a test suite built on VMap’s test environment. Experimental results were evaluated, compared with the search results of Google Maps address and similarity of results is up to 70%.

In the future, we will do a practical implementation method for the VMap platform. Many datasets will be used to test and evaluate the effectiveness of the method. Along with that, we will conduct an assessment survey based on the actual search wishes of the users on the VMap platform.

Acknowledgements

This work has been supported by Vietnam National University, Hanoi (VNU), under Project No. QG.18.63.

References

[1] The National Assembly of Vietnam, Law of Surveying and Cartography, 2018, https://thuvienphapluat.vn/van-ban/tai-nguyen-moi-truong/Luat-Do-dac-va-Ban-do-354638.aspx/, 2018 (accessed on: October 26th, 2020) (in Vietnamese).

[2] Top U.S. Mapping Apps by Reach 2018, Statista, https://www.statista.com/statistics/865419/most-
[3] Top U.S. Mapping Apps by Users 2018, Statista, https://www.statista.com/statistics/865413/most-popular-us-mapping-apps-ranked-by-reach/, 2018 (accessed on: October 20th, 2020) (in Vietnamese).

[4] Results of the Population and Housing Census at 0:00 on April 1, General Statistics Office of Vietnam, https://www.gso.gov.vn/du-lieu-so-lieu-thong-ke/2019/12/kei-qua-tong-dieu-tra-dan-so-va-nha-o-thoi-diem-0-gio-ngay-01-thang-4-nam/, 2019 (accessed on: October 26th, 2020) (in Vietnamese).

[5] Ministry of Information and Communication, Vmap: The Vietnamese Digital Map Has Officially come Into Operation (in Vietnamese).

[6] F. Wu, X. Wang, A Geocoding Algorithm for Natural Language Address with Neighborhood Properties, in Proceedings of 2nd International Conference on Computer Science and Network Technology, ICCSNT 2012, 2012, pp. 807-810, https://doi.org/10.1109/ICCSNT.2012.6526053.

[7] D. K. Matci, U. Avdan, Address Standardization Using the Natural Language Process for Improving Geocoding Results, Comput, Environ, Urban Syst., Vol. 70, 2018, pp. 1-8, https://doi.org/10.1016/j.compenvurbsys.2018.01.009.

[8] C. Lambert, S. Francisco, G. Lerner, Refining Location Estimates and Reverse Geocoding Based on User Profile Inventors, Browsing History, May, 2010.

[9] Elastic, Script score Query, Elasticsearch, Elastic, https://www.elastic.co/guide/en/elasticsearch/reference/current/query-dsl-function-score-query.html, 2018 (accessed on: October 26th, 2020) (in Vietnamese).

[10] I. Stewart, D. Tall, The exponential function, in Complex Analysis, 2012, pp. 8294.