Harnessing Insights from Crowd-Sourced Geotagged Data

Meenakshi Nawal
Department of Computer Science & Engineering, Swami Keshvanand Institute of Technology, Management & Gramothan, Jaipur, 302017 (India)
Email: meenakshi.nawal.02@gmail.com
Received 20.03.2026 **received in revised form** 09.05.2026, **accepted** 12.05.2026
DOI: 10.47904/IJSKIT.16.1.2026.25-31

Abstract- There is a voluminous data generated by people every day on geoportals which is being used tremendously. These geoportals are the life line to many people to get quick information about the location including address, geographical location, category and reviews of the place as well. The geo tagged data available on these portals is useful for better planning to improve in various sectors such as business, public services, event planning and disaster management. However, there is no way to verify the information available on these portals as these are not always contributed by domain experts. This data is collected through crowd sourcing therefore it is an important to make sure the accuracy of this crowd sourced data available on the geoportals. All existing verification methods are based on the crowd sourced information which again can't claim as the reliability of this data. To address this challenge, it becomes vital to explore the insight of geotagged data available on geoportal to understand the features affecting the accuracy most. There must be some ground truth data which can be taken as the reference data while exploring the crowd sourced data. In this paper, the exploratory statistical analysis is provided for crowd sourced geotagged data including the data set preparation using ground truth data as a reference data. In-depth analysis on prepared dataset will lay the foundation for developing automated accuracy measuring models which makes this vast data more useful for planning various citizen centric services specifically in developing nations. This study evaluates the accuracy of crowd-sourced geospatial data available on geoportals by comparing it against verified ground truth data across public service categories including healthcare, education, banking, and police stations. Key parameters assessed include name, address, pin code, geographic coordinates, category, and user type.Findings reveal that only 16% of the dataset was fully accurate**, while** 80% was partially correct **with one or more erroneous parameters (using a 50% match threshold). Hypothesis testing confirmed that both** user type **and** accessible online sources **significantly influence data accuracy, with all parameters found significant at a 5% confidence interval.**

Keywords– Geotagged data, Predictive Model, Supervised Learning, Verification

1. INTRODUCTION

Since last decade, humongous data is being generated globally on geoportal which is now getting used for different kind of activities such as business planning, sales prediction, disaster management planning, public services, as this data has great potential in multi dimensions. Big business house, government policy makers and each individual are dependent on the data contributed by the people at various crowd sourcing geoportals such as Open Street Map, Google Map and Bhuvan Portal etc. When the data has geographical information attached with it, it becomes more valuable which can be used to improve the citizen centric services especially in the developing nations. This data is a key role player to improve the facilities for the people. As this data is contributed by anyone with or without having domain knowledge in that area hence, there is no control on the accuracy and reliability of the data. Therefore, it is a need of the hour to measure the dependability of existing data which is available on the geoportal. There are some meta data associated with each geotagged Point of interest data which includes name of the tagged data, address, geo location coordinate, image, category, review etc. In order to assess the accuracy and reliability of the geotagged data, first we should identify the features affecting the accuracy.

There could be four possible ways to check the accuracy of this data such as 1) Manual verification 2) Crowd sourcing for verification 3) Social media data 4) automated model developed through rigorous data analysis done using historical data. First method is not feasible due to voluminousness of the data and the second and third method is again becoming a bottleneck as this will be dependent on crowd sourcing. Hence the last option is a best suitable choice to measure the accuracy of geotagged data available on geoportals. There is no Point of interest data set available, for which ground truth data is also available to ensure the accuracy level.

Lot of work has been done for truth discoveries for crowd sourced data but these are not related to the POI data available on geoportals [31]. Most of the research work done in this particular area is associated with POI recommendation [10,12,14]. Assessment of the accuracy of POI data is done with the help of social media data which is again a crowd sourced data contributed by common people [24]. Without the ground truth data, no one can evaluate the accuracy confidently.

To lay out the foundation of the automated verification of geotagged data, the data set is organized including the factual ground truth data and corresponding crowd sourced data which is available on the geoportal. In this paper POI data is analyzed in depth including a data set

formulation to the feature identification for building the robust verification model.

This paper or research includes-

1) Data set preparation along the ground truth data and simultaneous Crowd sourced data
2) Statistical Analysis of the prepared dataset
3) Identification of the features affecting the accuracy

2. RELATED WORK

Crowd sourced data available on geoportals ,a large amount of data where it can be used for healthier citizen centric services. Therefore,to analyze this data intensely to harness its power for robust applications is plays a crucial role in the society. Many researchers worked on the identification of land cover classification [15] and commercial area's boundary detection [13,24] using POI data. Twitter data and geotagged images were used to detect spatial area identification and geo location inferences [1,2,18]. Author suggested gazetteer construction using POI data [20]. Methods of detection of missing, hidden and obsolete POI were proposed [3,7,30]. Duplicate label removal techniques were suggested by the authors [9]. Lots of work has been done in the area of POI recommendation [10,12,14]. POI data collection and management methods were discussed [29]. POI data is is hardly ever verified for evaluation, and when it is, crowdsourced data is utilized. [24].Towards the accuracy or verifiability of the POI data a very few work is done. In this paper the insight of the crowd sourced geotagged data for measuring the accuracy is analyzed where the corresponding ground truth data is measured as a reference data. Proposed work towards accuracy measurement is a foundation for the implementing various available techniques to classify the crowd sourced data as per its accuracy.

3. PROBLEM FORMULATION

The brief introductions of the terms are given first before the problem formulation for measuring the accuracy of geotagged data with respect to the ground truth data. This is followed by an explanation of the data set preparation process, and finally the feature selection analysis for accuracy measurement.

Below are the frequently used terms and their interpretation used in this paper.

Crowd sourced data: Data from a common platform where huge number of individuals shares the information.

Geotagged POI data: It is a Point of interest data which comprises geographical data features along with other related data like name, address, category, latitude and longitude. It is used as tagged data interchangeably sat some places.

Ground truth data: This is a location data which is verified and provided by authentic agencies such as government data repositories. This is a reference data in this study.

Verified Geotagged data: A geotagged data is supposed to be verified if some method is applied to check its accuracy which might be either "correct", "incorrect" or "partial correct".

List of notations used in the paper is given in table 1

Table 1: Notations

Notation	Description
GT	Ground Truth
POI	Point of Interest
Lat	Latitude
Long	Longitude
GIS	Geographical Information System
Pincode	Postal Code
CS	Crowd sourced
Sr No	Serial Number

3.1 Problem Statement

Assuming GT is a set of ground truth data and CS is a set of crowd-sourced geotagged data available on the geoportal. F is a set of features associated with the data or in other words the meta data of the tagged data. Res is a set of outcomes or results and n is the total number of records available in the data set. Tar is a set of target values. These sets and their relationship are described below:

$$T = \{g_i\}, where\ i = 1\ to\ n\ ; CS = \{c_i\}\ where\ i = 1\ to\ n\ ; F = \{a_j\}, where\ j = 1\ to\ x \quad \text{(i)}$$

$$Res = \{\text{“}Correct\text{”}, \text{” } Incorrect\text{”}, \text{” } Partial\ Correct\text{”}\};$$

$$Tar = \{t_i\}, where\ i = 1\ to\ n \quad \text{(ii)}$$

σ_j *is the thershold value or* j^{th} *attribute*

$$\forall t,\ f(t_i) \in Res$$

$$f(t_i) = (\text{"}Correct\text{"})\ ,\ iff\ f(g_i, c_i) \geq \sigma_j\ \forall\ a_j\ ,\quad Lets\ call\ it\ X \quad \text{(iii)}$$

$$f(t_i) = (\text{"}Incorrect\text{"})\ ,\ iff\ f(g_i, c_i) \neq \sigma_j\ \forall\ a_j\quad , Lets\ Call\ it\ Y \quad \text{(iv)}$$

$$Tar - (X + Y) = \{r_i\}\ i \geq 1 \leq n\ ,\ \ lets\ call\ it\ Res,\ \ R = \{r_i\}$$

$$R \subset Tar,\ f(r_i) = (\text{"}Partial\ Correct\text{"}) \quad \text{(v)}$$

3.2 Hypothesis for Feature Selection

On the basis of meta data linked with the POI(Point of interest) tagged data, we formulate the below mention hypothesis which helps us to recognize the significance of the features included in the study. Numerical evidence-based features are easily comparable and can be evaluated easily like the difference between two given geolocation or in the postal code but some parameters like impact of user type can't be justified without the detailed statistical study. The below mentioned hypothesis is framed for two such parameters

Hypothesis 1: H0(Null Hypothesis): The accuracy of the crowdsourced geotagged data available on geoportals is unaffected by user type.

H1(Alternate Hypothesis): The accuracy of the crowdsourced geotagged data accessible on geoportals is influenced by the type of user.

Hypothesis 2: H0(Null Hypothesis): Available Web sources related to crowd sourced geotagged data has no bearing on how accurate this data is.
H1: The accuracy of the crowdsourced geotagged data accessible on geoportals is influenced by the available Web sources.
To analyze the above-mentioned problem and hypothesis, data available on geoportals needs to be examined in relation to matching ground truth data.

4. METHODOLOGY

Initially, the data set is gathered using crowd-sourced data from the geoportal and ground truth data.

4.1 Problem Statement

There are five major steps involved in this Data set preparation as shown in figure 1. Ground truth data collection and integration, corresponding geotagged data extraction, comparison of these two data sets and the final step is labeling of target.
First step in the data preparation is to collect the ground truth data. The data related to public service sector as education, bank, police station, and hospitals etc. as these data is available with government repository so we can trust the reliability of this data. Although this data doesn't have the consistent format and many times some information is missing. Therefore, we need to remove the missing information records and integrate the data collected from various government repositories for standard parameters. These parameters have the similarities with the meta data of geo tagged data available on the geoportal. The parameters which we considered for ground truth data are named description of the POI, address of the POI, Geo location points (Latitude and Longitude) and category.

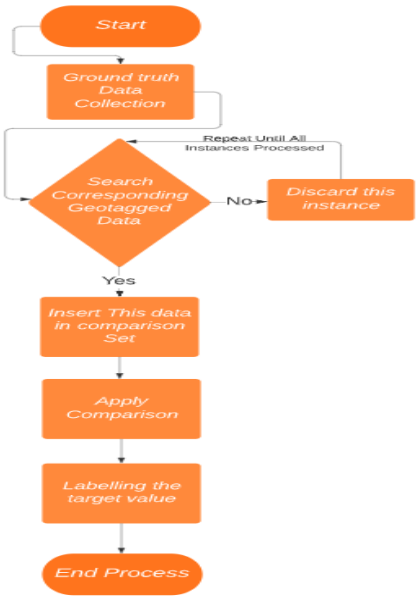

Figure 1: Data Preparation Flow

For corresponding geotagged data, the reverse geocoding API is used which is available on geoportal. Search of the data is done on the basis of named description, address and the postal code. Here few more parameters are considered like contributor's type and supportive web sources. Parameters collected through geoportals and web search are shown in table 2. These are "name of the poi", "address", "category", "latitude-longitude", "user type" and web source count which provide additional claim for the same data.

Table 2: Parameters for GT Data and CS Data

Notation	Description
Name	POIName
Address	POIAddress
Pincode	CSPincode
Category	CSCategorys
GTLat	CSLat
GTLong	CSLong
	UserType
	WebSrcCount

If the poi is not tagged for any record present in the ground truth data, then that record/instance will be discarded. User type could be either general user or local guide or owner. Until we test the first hypothesis, we assume that the user type is also playing significant role in accuracy of the geotagged data. Local guide has more knowledge about his/her locality and so the owner. Therefore, chances of error may be less in these cases as compared to the general user. WebSrcCount is the count of distinct number of sources which provide some information regarding the searched POI data so it provides additional supportive evidence to decide the accuracy of the geotagged data.
Next step is the comparison between these two sets. In both the sets the parameters are merged like name, category, address is the text and latitude, longitude and WebSrcCount are the numeric data. UserType is a nominal data. Therefore, no single method is suitable for comparing different kind of data. Cosine similarity and haversine formula are used for comparing text data and location data respectively.
The final stage of creating a data set is label assignment, which involves giving the response variable a value. The value for the response variable is nominal which is given by *set O* defined in problem statement section. Before deciding on the label value, data comparison and threshold testing should be completed. Algorithms 1 and 2 provide pseudo code for both of these procedures.
Data comparison is done as per explained in algorithm 1 where GT data set , CS data set, an empty comparison result set and the empty threshold result set are provided as the input. This Algorithm return the final result set after applying the threshold values.

ALGORITHM 1: DataSet Comparison

```
Input: CS_dataset, GT_Dataset, Comparison_resultSet,
threshold_rsult_set
Variable: name_match, pin_match, location_match,
Address_match, category_match
        for the ith iteration(i=1,2,….,n) do
name_match
=calculateCosineSimilarity(CSpoiNmae,GTPoiName)
address_match =calculateCosineSimilarity(CSaddress,
GTaddress)
category_match =calculateCosineSimilarity(CScategory,
GTcategory)
pin_match =calculateCosineSimilarity(CSpincode,GTpincode)
location_match
=calculateDiffDistance(CSLat,CSLong,GTLat,GTLong)
setValueInComparisonSet(name_match, address_match,
category_match ,pin_match, location_match)
  endfor
threshold_result_set  =checkThershold(Comparison_resultSet)
        output threshold_result_set
```

Algorithm 2 provides threshold value checking. It uses the comparison result set as input and returns the result set after the threshold value check.

ALGORITHM 2: Threshold Checking

```
Input: Comparison_resultSet
Output: threshold_result_set
        for the ith iteration(i=1,2,….,n) do
    if(name_match >=0.5 , “yes”:”no”)
    if(address_match >=0.5 , “yes”:”no”)
    if(category_match >=0.5 , “yes”:”no”)
    if(location_match<=1,”yes” : “no”)
    setValue(threshold_rsult_set)
  endfor
output  threshold_rsult_set
```

If all data sets are identical, there should ideally be a 100% match in every parameter included in the GT and CS data sets; nevertheless, it is practically impossible to get a perfect match in both data sets. Therefore, some threshold value is taken in each type of comparison like in case of text data comparison, For name, address, category, and pin matching, we assume that a cosine similarity score greater than 0.5 is meant to be matched. Similarly, for the purposes of this investigation, a location difference of less than one kilometer is acceptable. Therefore, each comparison parameter's output is derived as a nominal value based on these threshold values, which can be either "yes" or "no," as specified in algorithm 2.

3's logic method is used to determine the response variable's final value for each occurrence after obtaining the matching parameters as the nominal values.

ALGORITHM 3: Label Assignment

```
Input: threshold_rsult_set , empty  labelSet
Output: labelSet
        for the ith iteration(i=1,2,….,n) do
    if( all the parameters in threshold_rsult_set(i) == “yes”)
        setlabel(“Correct”)
    elseif(all the parameters in threshold_rsult_set(i) ==
“no”)
        setLabel(“incorrect”)
    else   setLabel(“partial correct”)
  endfor
output labelSet
```

Outcome of the algorithm 3 would be the final target label values which could be either “correct”, “incorrect” or “partial correct”. That’s how we prepared the supporting dataset which can be further used to analyzed for testing the hypothesis stated in section 2. The statistical analysis of the data set is analyzed in the following section.

4.2 Data Analysis

The statistical methods are applied to know the significance of the selected parameters on the response variable. General statistics of the data set is shown in table 3. Total 700 records of different categories are used and after removing the incomplete or the missing data during the data cleaning process, 439 records were available for further use. 16% records of this dataset are showing the completely correct data as these were labeled as “correct” and rest of the data have discrepancies which includes incorrect and partial correct records where incorrect means all the parameters having the mismatch while partial correct has the mismatch in one or more parameters .

Table 3: Statistics of the data set

Parameter Name	Label	Count	Weight	Distinct
PinMatch	Yes	335	335	2
	No	104	104	
LocationMatch	Yes	303	197	2
	No	136	98	
AddressMatch	Yes	230	146	2
	No	209	149	
CategoryMatch	Yes	303	232	2
	No	136	63	
WebSrcAvail	Yes	290	183	2
	No	149	112	
CSUserType	Owner	131	131	3
	Local Guide	121	121	
	GeneralUser	187	187	
Target	Incorrect	7	7	3
	Correct	70	70	
	Partial Correct	362	362	

Attribute distribution plot is shown in figure 2, which demonstrates that the data is not normally distributed. It also shows that classes are not balanced as there are very few records falls under the “correct” class. In such cases, non-parametric tests are suitable for further analysis. The data set is categorical in nature and all the events are mutually exclusive, therefore, a Chi-Square non parametric test is applied to understand the significance of the parameters on the response variables which is explained in the below subsection.

Chi-Square Test:

This statistical test aids in examining the connection between categorical variables and is used for hypothesis testing.According to the null hypothesis of this test, there is no impact of independent variables on the target parameter. The results of specified test are shown in table 4.

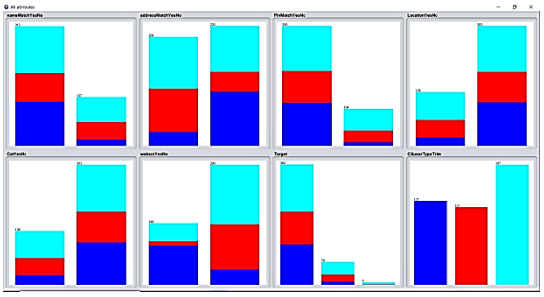

Figure 2: Attribute Distribution Plot of Name ,Address,Pin,Location,Category,Websrc,Target,UserType

Table 4: Chi-Squared Test Result

Parameter Name	Df	p-Value
PinMatch	2	0.004195787
LocationMatch	2	0.0003893918
AddressMatch	2	0.000000004175053
CategoryMatch	2	0.0003893918
WebSrcAvail	2	0.000102969
PoiNameMatch	2	0.0008566566
CSUserType	2	0.001601104

Significance of the parameters are tested for p-value set as less than 0.005 and observed that all the parameters involved are showing the significance on the target so the null hypothesis will be rejected. Therefore, all the parameters selected earlier, are kept for further model building. P-Value plot is shown in figure 3.

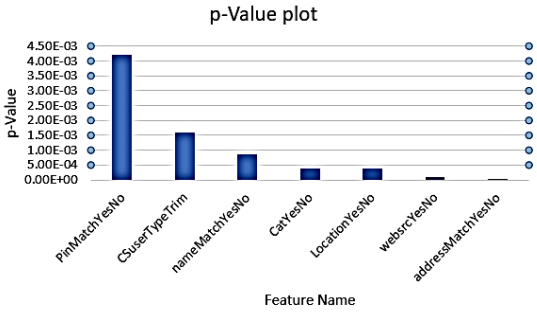

Figure 3: p-Value Plot with all features

Chi-square test and the p value help us to decide whether we accept the null hypothesis or not. The p value is obtained for all the involved parameters like pincode, address, name, category, websrc count, usertype etc. There are only 5% chances where the stated null hypothesis can be true. The results of acceptance or rejection of null hypothesis is discussed in following section.

5. IMPLEMENTATION & RESULTS

To analyze the relationship between accuracy of geotagged data and the metadata parameters, the dataset is prepared with the collection of the relevant data as per the requirement. Different tools were used to make the analysis more accurate and elaborative. For data collection, the reverse geocode API provided by Google, OSM and Mapmyindia is used. These API were called from a Java application and for some cases Postman Application was used. Ground truth data is extracted from Open Government Data (OGD) Platform India.

5.1 Results & Discussion

The required dataset is collected and analyzed data from the perspective of accuracy measurement. Few features are identified which could contribute in determining the correctness of the poi data. To understand the impact of these features, the hypothesis is stated in section 2. Testing of this hypothesis is given below.

Hypothesis 1: First hypothesis is related to find the impact of user type on the accuracy of the geotagged data.Three different types of users are exists in data set such as “owner”, “local guide” and “general user”. Distribution graph for these user types in the dataset is shown in below figure 4.

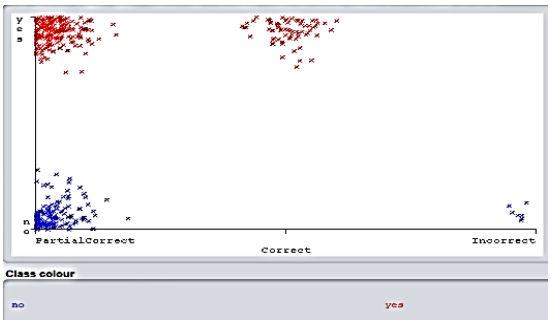

Figure 4: Distribution graph for CS user type

Owner is the business owner who has tagged his/her own business location on the geoportal. Local guide is the contributor living in the same vicinity whereas general user could be anyone other than these two types of contributors. Correlation plot is given in figure 5.

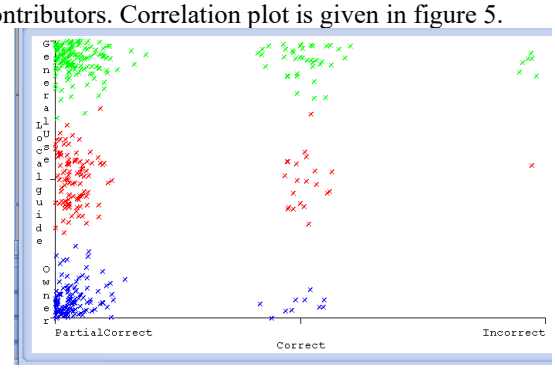

Figure 5: Correlation graph between user type and the target data

As we can observe from the correlation plot, owner has not contributed any incorrect data, local guide also committed very less incorrect data. Maximum incorrect information was tagged by the general user. P value for user type is 0.001601104 which is less than 0.05 ,Therefore, this data is adequate to refute the null hypothesis, which demonstrated that user type had no effect on accuracy.

Hypothesis 2: In order to determine the accuracy of the geotagged data, a second hypothesis is put out to assess the importance of the new online information. The data is examined in relation to the goal values and the web-

source count in order to test this hypothesis. Figure 6 displays a scatter plot illustrating the correlation.

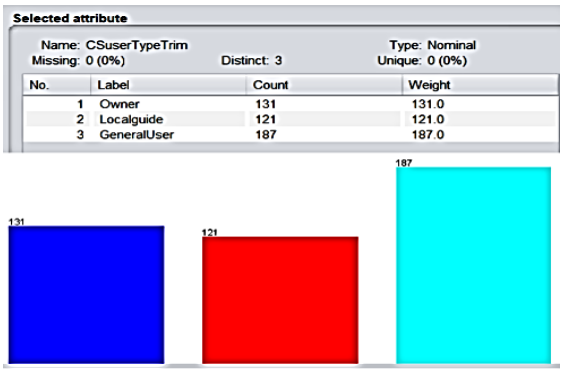

Figure 6: Correlation graph between Web source count and target data

The likelihood of more accurate tagged data improves when the same information about the sought geotagged data is available. Poi name, address and pin code were matched in the searched data and found the distinct number of sources which gives the same information. p value for this parameter is 0. 000102969 which is also less than 0.05. As per the results the null hypothesis is rejected and the alternate one is accepted.

6 .CONCLUSIONS

In this research, first the data set is prepared to analyses the crowd sourced data available on geoportals with the help of existing ground truth data. We are assuming that the ground truth data is already verified. Comparative study was done using ground truth data as the reference data mostly related to the public services like healthcare, education, bank, police station etc. considering the basic parameters such as “Name”, “Address”, “Pin code”, “latitude and Longitude”, “category”, “user type” and the supporting data available on the web. It is found that completely correct data is very less which is around 16% of the entire dataset. However, 80% data is partial correct where one or more parameters are wrong. Although the threshold is taken 50% for the matching criteria otherwise the proportion of inaccuracy may high. The hypothesis is investigated in order to determine the influence of user type and accessible online sources on the geotagged data. Testing for both the hypothesis showed the evidences to reject the null hypothesis and we concluded that user type has an impact of the accuracy of the data so the web sources. All the considered parameters were found significant for the 5% of the confidence interval. However, these results need be evaluated on larger datasets and with more factors to produce a robust system capable of predicting the accuracy of crowd-sourced geotagged data.

REFERENCES

[1] A. Alfarrarjeh, S. H. Kim, A. Deshmukh, S. Rajan, Y. Lu and C. Shahabi, "Spatial Coverage Measurement of Geo- Tagged Visual Data: A Database Approach," 2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM), Xi'an, China, 2018, pp. 1-8, doi: 10.1109/BigMM.2018.8499062.

[2] B. Devkota and H. Miyazaki, "An Exploratory Study on the Generation and Distribution of Geotagged Tweets in Nepal," 2018 IEEE 3rd International Conference on Computing, Communication and Security (ICCCS), Kathmandu, Nepal, 2018, pp. 70-76, doi: 10.1109/CCCS.2018.8586827.

[3] Chang, CH., Chuang, HM., Huang, CY. et al. Enhancing POI search on maps via online address extraction and associated information segmentation. Appl Intell 44, 539–556 (2016). https://doi.org/10.1007/s10489-015-0707-5

[4] Jurgens, T. Finethy, J. Mccorriston, Y. T. Xu, and D. Ruths, “Geolocation Prediction in Twitter Using Social Networks: A Critical Analysis and Review of Current Practice,” pp. 188–197. doi=10.1.1.700.5386

[5] D. Z. H. Arenga, J. C. Dela Cruz, F. M. A. Lorilla and P. A. Tangian, "Cloud-Based Flora Repository System with Geo-Location Mapping for Mt. Hamiguitan Sanctuary Exploration," 2018 IEEE Region Ten Symposium (Tensymp), Sydney, NSW, Australia, 2018, pp. 51-56, doi: 10.1109/TENCONSpring.2018.8692056.

[6] H. Arenga and M. Page, “Cloud-based Flora Repository System with Geo-Location Mapping for Mt. Hamiguitan Sanctuary Exploration,” 2018 IEEE Reg. Ten Symp., pp. 51–56, 2018. doi:10.1109/IALP.2018.8629117

[7] Hsiu-Min Chuang and Chia-Hui Chang. 2015. Verification of POI and Location Pairs via Weakly Labeled Web Data. In Proceedings of the 24th International Conference on World Wide Web (WWW '15 Companion). Association for Computing Machinery, New York, NY, USA, 743–748. DOI: https://doi.org/10.1145/2740908.2741715

[8] Hu, Sheng & he, zhanjun & Wu, Liang & Yin, Li & Xu, Yongyang & Cui, Haifu. (2019). A framework for extracting urban functional regions based on multiprototype word. Computers Environment and Urban Systems. 80. 10.1016/j.compenvurbsys.2019.101442. https://doi.org/10.1016/j.compenvurbsys.2019.101442

[9] Jingbo Zhou, Shan Gou, Renjun Hu, Dongxiang Zhang, Jin Xu, Airong Jiang, Ying Li, and Hui Xiong. 2019. A Collaborative Learning Framework to Tag Refinement for Points of Interest. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19). Association for Computing Machinery, New York, NY, USA, 1752–1761. DOI: https://doi.org/10.1145/3292500.3330698

[10] Jun Suzuki, Yoshihiko Suhara, Hiroyuki Toda, and Kyosuke Nishida. 2019. Personalized Visited-POI Assignment to Individual Raw GPS Trajectories. ACM Trans. Spatial Algorithms Syst. 5, 3, Article 16 (September 2019), 28 pages. DOI: https://doi.org/10.1145/3317667

[11] L. I. N. Li and W. E. I. Hu, “Visualization of Location-Referenced Web Textual Information Based on Map Mashups,” IEEE Access, vol. 7, pp. 40475–40487, 2019. doi: 10.1109/ACCESS.2019.2907570.

[12] L. Liu, J. Li and S. Ye, "Research on the Method of Constructing POI Data Model Based on AP-MRO," 2018 IEEE 4th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 2018, pp. 1730-1733, doi: 10.1109/ITOEC.2018.8740599.

[13] Lifang, Zhu & Yixin, Chen & Yang, Liu & Yue, Zhang & Jing, Wang. (2017). "POI data applied in extracting the boundary of commercial centers," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, China, 2017, pp. 44-47, doi: 10.1109/ICBDA.2017.8078693.

[14] M. N. Y. Utomo, T. B. Adji and I. Ardiyanto, "Geolocation prediction in social media data using text analysis: A review," 2018 International Conference on Information and Communications Technology (ICOIACT), Yogyakarta, Indonesia, 2018, pp. 84-89, doi: 10.1109/ICOIACT.2018.8350674.

[15] M. O. Pratama and A. M. Arymurthy, "Automatic land cover classification of geotagged images using ID3, Naïve Bayes and Random Forest," 2017 International Conference on Advanced Computer Science and Information Systems (ICACSIS), Bali, Indonesia, 2017, pp. 245-250, doi: 10.1109/ICACSIS.2017.8355041.

[16] N. G. P. S. Estimation, S. Jiang, Y. Kong, and Y. Fu, "Deep Geo-

Constrained Auto-Encoder for Non-Landmark GPS Estimation," in IEEE Transactions on Big Data, vol. 5, no. 2, pp. 120-133, 1 June 2019, doi: 10.1109/TBDATA.2017.2773096.

[17] N. Yoshikatsu, "Geographic names on old maps of early 20th century toward a spatio-temporal gazetteer: A study on their accuracy in Northeast Thailand," 2017 Pacific Neighborhood Consortium Annual Conference and Joint Meetings (PNC), Tainan, 2017, pp. 98-103, doi: 10.23919/PNC.2017.8203528.

[18] P. Li, H. Lu, N. Kanhabua, S. Zhao and G. Pan, "Location Inference for Non-Geotagged Tweets in User Timelines," in IEEE Transactions on Knowledge and Data Engineering, vol. 31, no. 6, pp. 1150-1165, 1 June 2019, doi: 10.1109/TKDE.2018.2852764.

[19] P. Taylor, K. Janowicz, and C. Keßler, “The role of ontology in improving gazetteer interaction,” no. November 2014, pp. 37–41, 2008. https://doi.org/10.1080/13658810701851461

[20] S. Gao, L. Li, W. Li, K. Janowicz, and Y. Zhang, “Constructing gazetteers from volunteered Big Geo-Data based on Hadoop,” CEUS, vol. 61, pp. 172–186, 2017. https://doi.org/10.1016/j.compenvurbsys.2014.02.004

[21] S. S. A. Abbas, S. K. Ajin and P. D. Sundar, "Realization of multimodal geo-tag using ARM-A53 with python," 2017 International Conference on Intelligent Computing and Control (I2C2), Coimbatore, 2017, pp. 1-6, doi: 10.1109/I2C2.2017.8321905.

[22] S. Y. Ssin, J. E. Zucco, J. A. Walsh, R. T. Smith and B. H. Thomas, "SONA: Improving Situational Awareness of Geotagged Information using Tangible Interfaces," 2017 International Symposium on Big Data Visual Analytics (BDVA), Adelaide, SA, 2017, pp. 1-8, doi: 10.1109/BDVA.2017.8114625.

[23] Sanket Kumar Singh, Davood Rafiei, Strategies for Geographical Scoping and Improving a Gazetteer, Proceedings of the 2018 World Wide Web Conference, April 23-27, 2018, Lyon, France .DOI:10.1145/3178876.3186078

[24] T. Bui, Y. Han, S. Park and S. Park, "Detection of POI boundaries through geographical topics," 2015 International Conference on Big Data and Smart Computing (BIGCOMP), Jeju, 2015, pp. 162-169, doi: 10.1109/35021BIGCOMP.2015.7072827.

[25] U. A. Deshpande , M. A. Radke, N. Gautam, A. Tambi , "Geotagging Text Data on the Web—A Geometrical Approach," in IEEE Access, vol. 6, pp. 30086-30099, 2018, doi: 10.1109/ACCESS.2018.2843814.

[26] W. Niu, J. Caverlee, H. Lu and K. Kamath, "Community-based geospatial tag estimation," 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), San Francisco, CA, USA, 2016, pp. 279-286, doi: 10.1109/ASONAM.2016.7752246.

[27] X. Qian, Y. Zhao and J. Han, "Image Location Estimation by Salient Region Matching," in IEEE Transactions on Image Processing, vol. 24, no. 11, pp. 4348-4358, Nov. 2015, doi: 10.1109/TIP.2015.2462131.

[28] Y. Liu, Runqiang Li, Kaichen Chen, Yihong Yuan, Lingli Huang and Hao Yu, "KIDGS: A geographical knowledge-informed digital gazetteer service," 2009 17th International Conference on Geoinformatics, Fairfax, VA, USA, 2009, pp. 1-6, doi: 10.1109/GEOINFORMATICS.2009.5293495.

[29] Yuanrong He, Yuanmao Zheng, Jian Deng and Huoping Pan, "Design and implementation of a POI collection and management system based on public map service," 2016 Fourth International Conference on Ubiquitous Positioning, Indoor Navigation and Location Based Services (UPINLBS), Shanghai, China, 2016, pp. 197-200, doi: 10.1109/UPINLBS.2016.7809971.

[30] Yue Cui, Liwei Deng, Yan Zhao, Bin Yao, Vincent W. Zheng, and Kai Zheng. 2019. Hidden POI Ranking with Spatial Crowdsourcing. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD '19). Association for Computing Machinery, New York, NY, USA, 814–824. DOI: https://doi.org/10.1145/3292500.3330844

[31] Sharma M, Bundele M, Bothale V, Nawal M. Fine-tuned Predictive Model for Verifying POI Data. International Journal of Advanced Computer Science and Applications. 2021;12(7).