

Locations Ranking using Page Rank Algorithm

Ghaidaa A. Al-Sultany*¹, Asraa A. Abd Al-Ameer²

^{1,2}Department of Information Network, Babylon University, Babel, Iraq
*Corresponding Author E-mail: ghaidaa.almulla@itnet.uobabylon.edu.iq

Abstract

Nowadays, large amount of user-generated data can be obtained from social media (e.g., Instagram and Flickr). People sharing their travel experiences with geo-tagged photos through these media, and the photo itself has important information like title, tags and location. These types of data provide a new perspective for us to understand the contexts of users. In this paper we focused on tourism service by collecting and analyzing geo-tagged photos from social media to identify the most popular tourist places and rank them based on user location. We used PageRank algorithm that ranks locations based on the relation between the locations and a set of tags that were discovered by a weighting scheme. Thus, intelligent location-based tourist services can be provided to people. The work experiments have shown encouraging results in terms of ranking the given locations.

Keywords: Geo-tags resources, Social media data, Information ranking, Tourist attraction, Smart cultural tourism

1. Introduction

In the past few years, information and communication technology tools and applications (e.g., mobile devices) are widespread and many of smart applications have been diversely used [1]. The benefits of social media and expansion abilities human exploration can be boosted by the mobile base and web service [2]. The use of social media is a growing phenomenon in contemporary society and its platform used as a means of communications and sharing information [3]. The users are capable to share media anywhere and at anytime to cooperatively communicate and interact with other users. They share their data like photos and videos and exchange their experiences with other users through it [2]. Using these data from social media can improve the utilization of it and provide good personalized services to users [1].

The large amounts of data shared on social media being geo-tagged reflect activities and interests of people's, thus offering capabilities to analyze and evaluate our world. By associated geographic information, these social-media items enable us to understand what geographic areas people are interested in. Mining this dataset can provide greatly valuable to a several set of applications including improving health [4], city management [5] and more. The data of social media (e.g., Flickr, Twitter, Instagram) called geo-tagged when it is associated with geographic coordinates, usually expressing where the item was generated [6].

Basically, many types of information can be collected from social networking services. Therefore, social media has been considered as a source of tourist resources [7]. Consequently, the behaviors of tourists can be observed through social media geo-tagged big data, thus providing backing for the applications such as intelligent travel recommendations and tourism resource development [8].

In this paper, we show an idea to exploit geo-tagged photos extracted from SNS (e.g., Flickr) to provide a service to people about the most attractive tourist places with respect to their location.

From defining a user's location the system can extract, collect the data (consisting of photos and tags) that are linked to nearby that location and analyze them as an agile way for providing service to people that guiding and help them as much as possible in the time of their traveling.

The paper is organized as follows. In Section 2, introduces some studies related to the background of using geo-tagged resources to provide specific services to people. Section 3 shows the methodology of our work that consists of data representation, tags similarity and explains PageRank algorithm and how we used it to rank tourist locations to provide a tourist service to users. Section 4 shows the results and experiment discussion of our work. Section 5 draws the conclusion of our study.

2. Related Work

Recently, there are many studies that benefit from the social network service resources such as geo-tagged photos which aim to provide a specific tourist service to users and help in the development of tourism. Where geo-tagged resources were used by the authors in [8], they proposed CFSFDP, a new clustering method in discovering the tourist attraction and they added zoning and standardization step in the process of spatial clustering to make their theory more flexible to the clustering scenes when there are large differences in regional density. In addition and to set the tag vector for each initial cluster, they used TF-IDF method and then merged adjacent semantically similar clusters by performing vector similarity calculation which enables more efficient in tourist attraction discovering that has an irregular shape and a large geographical area. Their experimental results displayed that their approach is much higher in the accuracy of the classification, has better flexibility in the case of an irregular density distribution and enables distinguishing adjacent high-density areas. In [9] the authors showed an idea to exploit the tags attached on geographical photos that published by users in SNS to help finding the most familiar locations for enjoying in specific food by using traditional dishes keyword. In order to use geo-tagged resources for ranking locations, they used LoCHITS algorithm and

modified the value of nodes with relevant locations and tags. Besides, they extend LocHITS to LocHITSS algorithm by using similarity between the set of tags of each location and the set of common tags. Furthermore, in order to calculate the weight of tags and apply them into the LocHITSTF algorithm, they used term frequency of tags in each location. The experiments showed that the GeoHITS algorithm converge quickly, and the results had high ranking values.

As another approach other authors focused more on practical and structural forms of Flickr mining rather than algorithmic and technical forms of it such as in [10], the researcher proposes a framework that used geo-tagged photos in Queensland, Australia (attractive tourist destination) to extract associative points-of-interest patterns. The framework consist of two popular data mining techniques: making for the points-of-interest detection a clustering, and association rules mining for associative points-of-interest patterns. While the authors in [11] find the tourist attractions and places from the geo-tagged photos that have geographical location information and extracted representative photos and corresponding text labels, able the users to use text or pictures to search tourist attractions places.

3. Methodology

In our work we focus on defining the problems and fundamental issues that may face when want to provide to tourists some helpful information associated to their trip, by using geo-tagged social media data we can identify tourist places from user's position and rank them. To ranking locations, we depend on the relationships between tags and locations. The problem of ranking location focus on data that can obtained from SNS by using tags that attached with it determine the features of locations. Where a tag (t_i) is a label attached to photo in order to identify or describe the its contexts, location (l_j) is a name of a country, city, or a place that used for ranking and it is identified by a collection of tags over social media. The relation defines the link between a tag (t_i) and a location (l_j), so if the tag (t_i) is contained in location (l_j) or in other words, if location (l_j) has been tagged with tag (t_i), this means that there is a relation (link) between the tag and the location.

We use a collection of geo-tagged photos from SNS and applying ranking method in order to identify and find the better location for the user. The tags that appear on many locations should be chosen for using in this work. Besides, we inspired from [1] that to find the best location for the user by ranking method, we use a collection of tags that occurs in many locations. Because there is no meaning for some tags with the other location that did not contain them. We need to use a collection of common tags. The common tags are defined as :

$$\mathbb{T}^c = \{t_i | t_i \in \mathbb{T}, |\mathbb{L}^{t_i}| \geq \alpha\} \tag{1}$$

Where t is a tag, l is a location, \mathbb{T} is a collection of tags, \mathbb{L} is a collection of locations and \mathbb{L}^t is a collection of locations contain tag t .

In social media, there are large scale use for tags from which it can be possible to apply a process to select the most popular tags that has relationship closed to the keyword. The workflow of our STS (Social Tourist Service) is shown in the Figure 1.

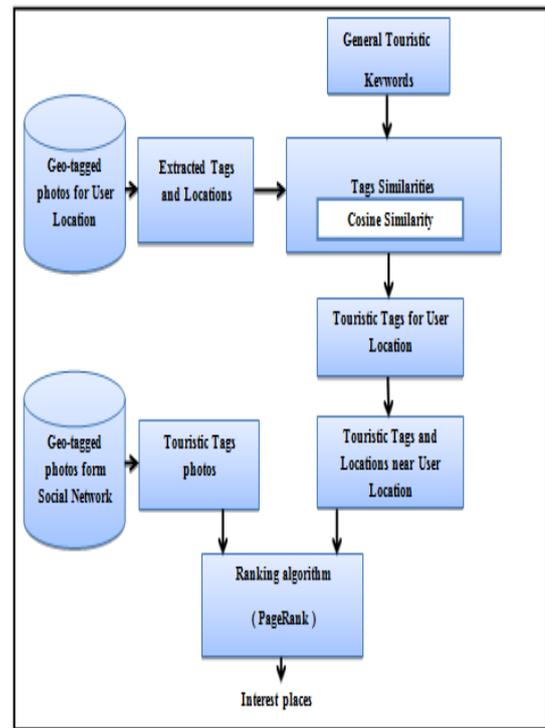


Fig. 1: The proposed research diagram

3.1 Data Representation

Our system used dataset taken from [12]. This dataset crawled from Flickr and it consists of a sample of geo-tagged photos with the corresponding file metadata that consist of image metadata in CSV format that contain information's fields about all images. Each field define a specific information on the image. 5000 geo-tagged photos have been taken from the dataset for testing this work. The selected dataset has been apart into two parts, the first one has contained a set of locations that taken from photos with its tags collected based on their closeness to user location or given area. The latitude and longitude of the location where the photo was taken using Google map API service have been utilized to get the locations that near the user or given area. The second part of the dataset collected based on using a keyword related to tourism through calculating the cosine similarity [13] between set of most common tourist keywords and the tags of all photos that contain a set of tourist tags.

3.2 Tags Similarity

In the process of ranking, we use tourist tags to decision a list of ranking tourist locations and that has been done by PageRank algorithm. So, to get a list of tourist tags, we calculate the cosine similarity measure between each tag for each location with the general tourist keywords to get the tourist tags only.

After selecting a collection of tourist tags, we calculated the number of locations that contain the tourist tag to get a collection of common tags that occurred in many locations, and that depend on the tags of locations to compute the ranking value for each location. Thus, that will change the weight of vertices (tags and locations) and make the result of ranking by PageRank more accurate.

Each location will be affected by its tags and vice versa; each tag will be affected by its location. Thus, we need to compute the weight of tags in each location; and we inspired that from [1] where the number of occurrence for each tourist tag in the location will count and will compare with the total number of tags in that location. It's calculated as follows :

$$w_{ij} = \frac{p(t_i, l_j)}{\max\{p(t_k, l_j) | t_k \in \mathbb{T}^{l_j}\}} \tag{2}$$

With $\varphi(t_i, l_j)$ is the count of occurrence of tag t_i in location l_j and the maximum occurrence of t_k (t_k is a tag in the collection of location l_j tags). In our work, we used the relationships between tags and locations such that value of each tags in each location and value of each location that contain the tag.

3.3 Formalization of Location Ranking

PageRank algorithm is proposed by Sergey Brin and Lawrence Page, it's one of the most commonly ranking algorithm that used for ranking the various pages and its currently used by Google search engine [14].

From studying results of [15], we have proposed the PageRank algorithm to find the best location, in which each node is defined by a tag or a location. The grade of a vertex of a graph in graph approach, is the count of edges that incident to the vertex. In directed-graph, there are two kinds of degree of vertex p : In-degree "the count of links that points to vertex p " and Out-degree "the count of links from p point to other points". Differently, the count of In-degree and Out-degree of vertex in an undirected graph is equal.

In this paper and for ranking locations, we depend on the relation between tags and locations and an undirected graph we used to represent data $G = \langle N, L \rangle$ where N represented the collection of nodes corresponds to a tag t_i or a location l_j and L is the collection of edges $(t_i, l_j) \in L$ indicates location l_j contains tag t_i . To calculate values for nodes, we found that we should chose the tags that occur in various locations because if only one tag show in a location for just a few times then it has barren for another locations ranking.

At the beginning and to get the most interested locations, the initial approximation of PageRank for locations ($\forall v \in N$) would begin with an estimated PageRank of $1/n$ (n is the number of location in the place) and the PageRank of tags would begin with an estimated PageRank of $1/m$ (m is the number of tourist tags) as follows:

$$PR(l_j) = \frac{1}{n} \tag{3}$$

$$PR(t_i) = \frac{1}{m} \tag{4}$$

Then, for each iterations the PageRank of all locations will depend on the PageRank of tourist tags that it contained and the probability of transitioning from tag to location and that will calculated by the formulas as follows:

$$PR(l_j) = \sum_{i=1}^m PR(t_i, l_j) \cdot p_{ij} \tag{5}$$

In addition to that, for each iterations the PageRank of all tags will depend on the PageRank of location that contained in it and the probability of transitioning from location to tag and that will computed by the formulas as follows:

$$PR(t_i) = \sum_{j=1}^n PR(l_j, t_i) \cdot p_{ji} \tag{6}$$

Because the edges in the graph that we deal with it in our work may have weight (p_{ij}), the probability of transitioning from tag to location (p_{ij}) and the probability of transitioning from location to tag (p_{ji}) is defined in these cases as follow [16]:

$$p_{ij} = \frac{w_{ij}}{\sum_{j \in Out(i)} w_{ij}} \tag{7}$$

$$p_{ji} = \frac{w_{ji}}{\sum_{i \in Out(j)} w_{ji}} \tag{8}$$

In this study we used undirected graph to rank location based of its tags. So, the weight of edges between the tag and location and vice versa is the same. At the end, the final PageRank of tags and locations will calculate as follows:

$$PR(t_i) = \frac{d}{m} + (1 - d) \cdot \sum_{j \in In(i)} PR(l_j) \cdot p_{ji} \tag{9}$$

$$PR(l_j) = \frac{d}{n} + (1 - d) \cdot \sum_{i \in In(j)} PR(t_i) \cdot p_{ij} \tag{10}$$

Where n is the total number of locations, m is the total number of tags and d is a damping factor which can be set to between 0 and 1, $d = 0.85$ in this paper.

4. Result and Experimental Discussion

In this section, we introduced the dataset that collected from SNS [12] and used in our work and shows implementation steps to rank locations based on tags of them and its results.

4.1 Dataset

To perform our work with real-world data, a data of geo-tagged photos on a world-wide scale is presented. The data contains a sample of 5000 geo-tagged photos posted and shared by many users on over the world and crawled from Flickr with the corresponding metadata like title, tags, latitude, longitude, date taken, date uploaded and other attached features belong to that photos.

4.2 Experimental Results

In our work, we used two dataset, the first for the locations around the given area, we assume that the user is in a New Orleans city which is one of Louisiana state in America country, using the position of the city (latitude and longitude) the application collected a set of locations found in it by using one of Google map API services which is Geocoding API service, the locations that collect its depend on latitude and longitude metadata of photos that collect from Flickr and its considered dataset 1. The second for tourist words (tags) extracted from geo-tagged photos that posted and shared by many users on over the world when they want to tags or comment through Flickr and that by using keyword related to tourism and its considered dataset 2.

Two different cases were experienced to test the results of our work before working on the 5000 geo-tagged photos, in each case there is a different number of tourist locations and non-tourist locations. A 500 different photos (tourist and non-tourist photos) were taken from different places around the world, 31 photos from New Orleans city which we want to find the tourist location in it and the other from different locations around the worlds. The 31 photos were captured and shared from 20 different location in New Orleans. In order to evaluated the results, two measures precision and recall have been used and in the F-measure the two measures are used together to provide one measurement for the evaluation of our system as show below [1]:

$$p = \frac{|\text{relevant}|}{|\text{predicted}|}$$

$$r = \frac{|\text{relevant}|}{|\text{actual}|}$$

F-measure

Case 1 : In this case, 31 different photos from New Orleans city that captured and share from 20 different location have been used. From the 20 locations, 17 of them its tourist location and the rest (3 locations) it's not. The result gave the 17 tourist locations and two non-tourist location as shown in Table 1 based on tags relevant to tourism that collect from social network that shown in Table 2.

Table 1: The rank of tourist places

The Locations	Locations Rank Value	Touristic/Not Touristic
1725 St Roch Ave	0.06374373294351214	Touristic
343 Carondelet St	0.05903317144347555	Touristic
425 Basin St	0.05804262640382418	Touristic
5100 Canal Blvd	0.05780526628352904	Touristic
2900 Loyola Ave	0.056617381462497396	Touristic
102 City Park Ave	0.05542947975743634	Touristic
401 St Peter St	0.05428507363539069	Touristic
1 Canal St	0.051934072891302155	Touristic
733 Bourbon St	0.05189291420918382	Touristic
317 N Rampart St	0.051882888048774986	Not Touristic
800 Decatur St	0.050292397660818715	Not Touristic
698 Loyola Ave	0.049554242436851303	Touristic
Aquarium Dr	0.049551278442396325	Touristic
Collins Diboll Cir	0.04952696167338881	Touristic
90057 Diagonal Dr	0.04951506158874967	Touristic
2604 Coliseum St	0.04950188575578165	Touristic
5016 Willow St	0.047148192963958856	Touristic
Mardi Gras World	0.047122627228623294	Touristic
1320 Mardi Gras Blvd	0.047120745170504824	Touristic

Table 2: Tourist tags (Dataset 2)

Tourist Tags
cemetery - grave - river - trip - vault - water - tourist - photo - travel - trees - mississippiriver - garden - tour - tourism

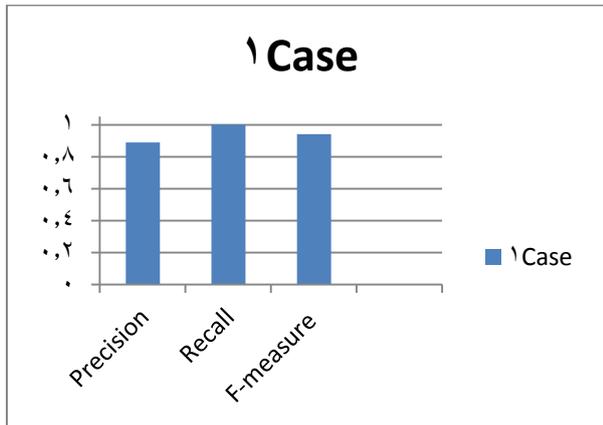


Fig. 2: Performance measure for case 1

Case 2: In this case, 31 different photos from New Orleans city that captured and share from 20 locations have been used. From the 20 locations, 14 of them are tourist locations and the rest (6 locations) are not. The result gave the 14 tourist locations and four non-tourist locations as shown in Table 3 based on tags relevant to tourism that collect from social network that shown in Table 4.

Table 3: The rank of tourist places

The Locations	Locations Rank Value	Touristic/Not Touristic
1725 St Roch Ave	0.0693876101165103	Touristic
425 Basin St	0.06273799374822389	Touristic
5100 Canal Blvd	0.06246092639954528	Touristic
2900 Loyola Ave	0.06107558965615227	Touristic
102 City Park Ave	0.05969025291275927	Touristic
1 Canal St	0.0564300352573853	Touristic
401 St Peter St	0.05642870260543905	Touristic
2604 Coliseum St	0.05374396135265701	Touristic
800 Decatur St	0.05374396135265701	Not Touristic
Collins Diboll Cir	0.05336007673913605	Touristic
90057 Diagonal Dr	0.0533595047898433	Touristic
733 Bourbon St	0.05335942624967543	Touristic
5016 Willow St	0.050292384690543056	Not Touristic
1320 Mardi Gras Blvd	0.05029098310627568	Touristic
Mardi Gras World	0.05029075514396734	Touristic
433 Bolivar St	0.04999289570900823	Not Touristic
8025 St Charles Ave	0.04999289570900823	Not Touristic

Table 4: Tourist tags (Dataset 2)

Tourist Tags
cemetery - grave - tomb - flower - photo - vault - water - tourist - travel - trip - mississippiriver - garden - tour - tourism

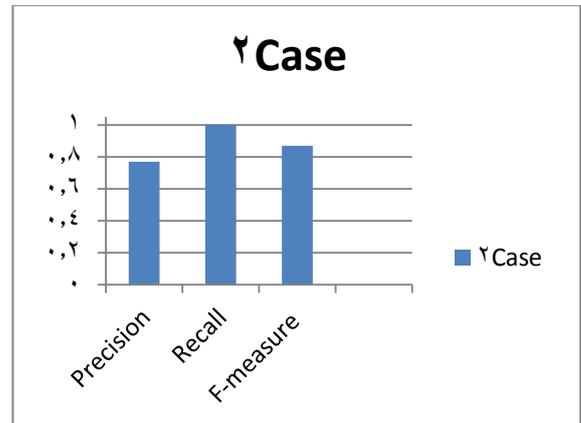


Fig. 3: Performance measure for case 2

Besides, our system has been implemented on data containing a sample of 5000 geo-tagged photos posted and shared by many users over the world and crawled from Flickr. The first dataset contains a set of locations that are near the user location; we assume that the user is in New Orleans city which is one of Louisiana states in America country. The first dataset is described in Table 5.

Table 5: Dataset 1 (The names of locations in New Orleans city with its tags)

Location name	Location tags
2900-2998 Loyola Ave	cemetery- grave- louisiana -neworleans -vault- crypt - fuchs
102 City Park Ave	door - cemetery - rust - Louisiana- neworleans - chain doorknob - padlock - crypt - forlorn - grave - warning - tomb - gravesite
5100 Canal Blvd	flowers - cemetery - grave - gold - golden - louisiana - paint - neworleans - vault -bouquet - crypt - grave headstone - tomb -brotherhood - boilermakers - shipbuilders
425 Basin St	flowers - blue - roses - strange - grave - Louisiana - purple neworleans - eerie - vault - crypt - cemetery - cult vault - priestess - voodoo - marieleveau - ochre - soldier - memorial - revolution - revolutionarywar - black - dark - magic - occult - sinistergravedeathcryptneworleanslouisianaeeriemacabre - death - macabre
1725 St Roch Ave	strange - grave - death - louisiana - god - neworleans - cemetery - religion - jesus - eerie - graves - macabre - hdr - gravesite - bigeassy - stroch - camposanto
1914 Tea Room Dr	bird - nature - animals - zoo - louisiana - wildlife - neworleans - parakeet - audubonzoo- orangutan - monkey - primate - feline - jaguar - growl - cat - alligator - bayou - albino - gator - swamp - tropical - red - colorful - toucan

The second dataset contains common tourist tags that are used by many users around the world when they published geo-tagged photos in Flickr as described in Table 6.

Table 6: Dataset 2 (Tourist tags with the number of occurrence)

Tourist tags	Number of locations that contain it	Tourist tags	Number of locations that contain it
nature	110	bird	34
landscape	78	mountain	33
sea	68	beach	30
blue	54	bridge	30
architecture	50	cemetery	29
scenery	48	animals	29

wood	46	louisiana	28
red	44	flowers	27

The data that result from dataset1 and dataset2 have been taken to get the essentialtags for ranking the locations that belong to tourism. The locations are tagged by the people when they have been visiting it. So, all locations is described by a collection of tags.To filter data in each location, the tags that are related to tourist will be used.So, from the two dataset we noticed that each location have a set of tourist tags as shown in Table 7.

Table 7: The locations with its tourist tags

Location name	Tourist tags
2900-2998 Loyola Ave	cemetery- louisiana
102 City Park Ave	cemetery - louisiana
5100 Canal Blvd	flowers – cemetery - louisiana
425 Basin St	flowers – blue – louisiana
1725 St Roch Ave	cemetery - louisiana
1914 Tea Room Dr	bird - nature - animals - louisiana

The rank of tourist places depend on user’s favorites by tagging the photos that they share on social media. By the relation between the locations and tourist tags , we use the data in dataset2 (Table 6) to get the term frequency of the tourist tag in all locations of dataset 1 (Table 5) . Undirected graph used to represent data where all vertex represent a location or a tag, and all edge represent the relationship between them. Here, we only taking into consideration the relationship tag–place or place–tag , without taking the relationship tag–tag or place–place. By applying PageRank algorithm depending on the relation between tags and locations we will get the most attractive tourist places and tourist tags in New Orleans city. Page rank algorithm will remove the locations that not contain any tourist tag and remove the tags that don't found in any location ,So that will give a more accurate result as shown in Table 8 .

Table 8: The rank of tourist places

Rank of locations	PageRank value for the locations
1725 St Roch Ave	0.18164971507234723
1914 Tea Room Dr	0.16600514487352738
102 City Park Ave	0.16451425775244655
425 Basin St	0.16451424071901233
5100 Canal Blvd	0.1616583215129468
2900 Loyola Ave	0.16165832006971925

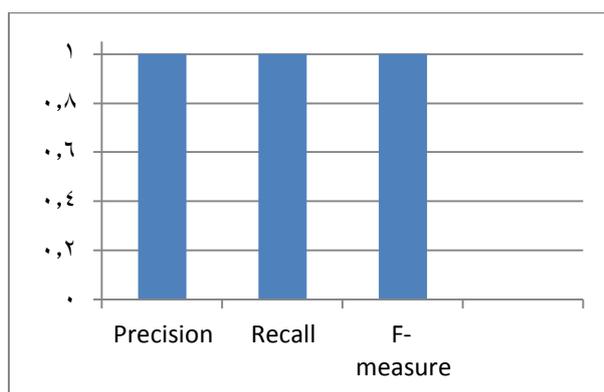


Fig. 4: Performance measure for the system

5. Conclusion

In this paper,a benefit has been shown from social network data. By given a particular geolocation,tourists can be provided with the most attractive tourist places in their area based on geo-tagged resources in social media. To ranking locations based on geo-tagged resources, we proposed PageRank algorithm and modified the values of the locations based on the tourist tags that contained it.By that,impressive information can find for tourists and

providing effective information to the agents of travelling and that helping in the development of tourism.

References

- [1] T. T. Nguyen, D. Camacho, and J. E. Jung, "Identifying and ranking cultural heritage resources on geotagged social media for smart cultural tourism services," *Pers. Ubiquitous Comput.*, vol. 21, no. 2, pp. 267–279, 2017.
- [2] I. Memon, L. Chen, A. Majid, M. Lv, I. Hussain, and G. Chen, "Travel Recommendation Using Geo-tagged Photos in Social Media for Tourist," *Wirel. Pers. Commun.*, vol. 80, no. 4, pp. 1347–1362, 2015.
- [3] L. Townsend and C. Wallace, "Social Media Research: A Guide to Ethics," *Univ. Aberdeen*, pp. 1–16, 2016.
- [4] A. Sadilek, H. Kautz, and V. Silenzio, "Modeling Spread of Disease from Social Interactions," *Proc. Sixth Int. AAAI Conf. Weblogs Soc. Media*, pp. 322–329, 2012.
- [5] C. Tech, A. Krebs, C. Tech, R. Schwartz, A. Langdon, and C. Tech, "CityBeat: Real-time Social Media Visualization of Hyper-local City Data," pp. 167–170, 2014.
- [6] D. Flatow, M. Naaman, K. E. Xie, Y. Volkovich, and Y. Kanza, "On the Accuracy of Hyper-local Geotagging of Social Media Content," *Proc. Eighth ACM Int. Conf. Web Search Data Min. - WSDM '15*, pp. 127–136, 2015.
- [7] J. J. Nguyen TT, Hwang D, *Intelligent Distributed Computing III*, vol. 237. Madrid, Spain, 2009.
- [8] X. Peng and Z. Huang, "A Novel Popular Tourist Attraction Discovering Approach Based on Geo-Tagged Social Media Big Data," *ISPRS Int. J. Geo-Information*, vol. 6, no. 7, p. 216, 2017.
- [9] T. Tri Nguyen and J. J. Jung, "Exploiting geotagged resources to spatial ranking by extending HITS algorithm," *Comput. Sci. Inf. Syst.*, vol. 12, no. 1, pp. 185–201, 2015.
- [10] I. Lee, G. Cai, and K. Lee, "Exploration of geo-tagged photos through data mining approaches," *Expert Syst. Appl.*, vol. 41, no. 2, pp. 397–405, 2014.
- [11] Y. Gao, J. Tang, R. Hong, Q. Dai, T.-S. Chua, and R. Jain, "W2Go," *Proc. Int. Conf. Multimed. - MM '10*, p. 123, 2010.
- [12] H. Mousselly-Sergieh, D. Watzinger, B. Huber, M. Döller, E. Egyed-Zsigmond, and H. Kosch, "World-wide scale geotagged image dataset for automatic image annotation and reverse geotagging," *Proc. 5th ACM Multimed. Syst. Conf. - MMSys '14*, pp. 47–52, 2014.
- [13] N. Pradhan, M. Gyanchandani, and R. Wadhvani, "A Review on Text Similarity Technique used in IR and its Application," *Int. J. Comput. Appl.*, vol. 120, no. 9, pp. 29–34, 2015.
- [14] N. V. Pardakhe and R. R. Keole, "Analysis of Various Web Page Ranking Algorithms in Web Structure Mining," *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 2, no. 12, pp. 4798–4803, 2013.
- [15] M. Batra, S. Sharma, A. Prof, D. Of, C. Applications, and M. Rachna, "Comparative Study of Page Rank Algorithm With Different Ranking Algorithms Adopted By Search Engine For Website Ranking," *Int. J. Comput. Technol. Appl.*, vol. 4, no. 1, pp. 8–18, 2013.
- [16] C. C. Aggarwal, *Recommender Systems*. 2016.