Journal Homepage: - www.journalijar.com
INTERNATIONAL JOURNAL OF ADVANCED RESEARCH (IJAR)
Article DOI: 10.21474/IJAR01/6283
DOI URL: http://dx.doi.org/10.21474/IJAR01/6283

*RESEARCH ARTICLE*

# GEO-LOCATION ON TWITTER AND INSTAGRAM BASED ON OSINT TECHNIQUES: A CASE STUDY.

*K. Tzovelekis, V. Kanakaris and D. V. Bandekas.

## Manuscript Info

*Manuscript History*

Received: 11 November 2017
Final Accepted: 13 December 2017
Published: January 2018

*Key words:-*
Twitter; Instagram; social media; OSINT techniques; Geo-Location.

## Abstract

We are living in the era of technology and massive information exchange and social networks are very important part as they drastically changed how we communicate. The use of this technology, inevitably leads to disclosure of personal data on the part of users. Making a distinction between the medium itself (social networking platform) and how it is used by users, we set the research question of how possible is a malicious user to retrieve personal data in an automated undetectable way and use them against an unsuspicious person that uses social media. Answering the above question, this paper had intended to detect the possibility of retrieving information in an automated way of two popular social networks, Twitter and Instagram, without the consent or informing the end user - target. The results shown that our activity on social networks, can be exploited by a malicious user using Open-Source Intelligence (OSINT) techniques and can collect sensitive personal data (user location via tweets / instas from smartphone). This work showcases user activity (geolocated tweets, instas) on a Google Map and connects this activity in chronological order with vectors, using a travel by map animation. Points of interest are clustered. Showed that users of twitter and instagram cannot fully protected from intruders and misusage of personal information, which unhappily keeps to propagate the personality of the medium.

It is worth mentioning that the location activity took place in the two particular social networks, through the targeted user smartphone, reveals the GPS coordinates of the user activity and this was reproduced in a Google Maps type map where we also placed connecting points chronologically. This way a malicious user can track a person's activity and it is possible to predict future location.



## Introduction:-

Social Networks trigger users to rapidly share information with widely dispersed networks. Users are able to post content in a variety of formats, which can be instantly made available to their entire social network 124. Therefore, social networks have become an important platform for dissemination of information, web content discovery, opinion sharing, discussion and debate. The large volume of public data that flows through social network has the potential to deliver valuable new insight to the academic community, marketing agencies, interested in understanding online behavior and monitoring social trends1.

**Corresponding Author:- K. Tzovelekis.**

Twitter is a platform that allows mobile users to embed their precise geo-location into short textual updates known as tweets. The ability to collect and mine location aware tweets opens new research directions by making it possible to study the spatial as well as textual characteristics of online content. It also provides a means for monitoring social trends and online activity on a regional basis. Before the recent smart-phone boom when mobile access to social networks was limited, Twitter enabled anyone with access to a cell-phone to communicate rapidly with widely dispersed networks of people through SMS67 .

As the volume of content shared publicly on social networks continues to grow, the demand for technology that can assist with the collection and mining of this content. Twitter's potential as a tool for research and analysis is underlined by its rapid growth and emergence as a mainstream channel of communication on the Internet. Most existing research into Twitter has focused on social network analysis based on the analysis of usage patterns and the textual content contained within tweets89.

Due to the public nature of most user generated content that flows through its network, Twitter is a particularly useful source for intelligence gathering and large scale data analysis. This is in sharp contrast to Facebook, where user generated content is only made available to pre-selected lists of users41112.

This paper describes the architecture and implementation of a software system that collects data anonymously for Twitter and Instagram. These two platforms are mobile first and their use is more engaging and rich through a modern smartphone. Instagram users can send a photo or a video of 15 seconds and add a comment with limitation of 2200 characters. On the other hand, Twitter users can send only text of 160 characters. Both platforms can tag the location5. The toolkit was designed to provide researchers with access to relevant Twitter and Instagram data in a format suitable for further analysis and data mining. The system integrates modules for data collection, spatial storage of offline data, spatial data retrieval, full-text search, geo-location data, data mapping and export 2.

First, the paper discusses modules for background gathering of Twitter data using the REST API and "geoJSON" in order to extract information regarding location of user and simultaneous conversion of this data to a spatial format. Whilst real-time Twitter data collection features have been integrated into a number of modern software systems, notably, there is a lack of tools for storing the data in a format that enables users to perform advanced spatial queries. Second, the paper outlines the architecture of connecting anonymously to the Internet. Third, the paper introduces the application of "Instagram Real-Time API" which together with "Genymotion" can extract information about the location of the user. These techniques are described as operating at the time of this research took place from July to November 2015. Lastly, the paper describes methodology used for data visualization and export. It concludes with a discussion of test usage of the system and its limitations8910.

**Related Work:-**

Other researchers and their related work has focused on11:

1. Location acknowledgment of a given user
2. Location acknowledgment of an individual tweet
3. Modeling the spatial density of users

The basic difference between our approach and previous researches from the first category is that they focus on the geo-location of users and tweets that are associated with the provided area, while our goal is to use free apps in order to get information about the geo-location of the user that is used as input. Eisensten et al12 try to solve the user geo-location extraction through geographical topic models by capturing the difference in the language use for a particular theme among user from distant regions. Ahmed et al 13 proposed a scheme of hierarchical structure of the topics by categorizing them and extracting location-specific topics and place users. On the other hand, Cheng et al 14 used the location of phrases and not the topics. Mahmud et al. 15improve the previous method by using Bayes classifier in order to predict the country, state and city of a user. Ren et al. 16 uses an approach that locate each user at a place that associated with the most of his friends.

The second category of relevant research focuses on the geolocation of the tweets. Ikawa et al. 17 tried to estimate the location of a tweet by consorting with expressions that have locations. Li et al. 18 got a better accuracy by identifying Places of Interest (POI). Kinsela et al. 19 use a Query Likelihood model in order to find the location the user and the tweet.

The third group of relevant approaches focuses on modeling the spatial density of users. Cho et al. 20 used a combination model of Gaussians in order to get the diversity of users behavioral patern. Lichman et al. 21 improve the previous approach by using the Kernel Density Estimators (KDE). By this way avoids the data sparsity.

**System Architecture:-**

*Twitter REST API**Error! Reference source not found.**82631*

Twitter is a popular online social media and microblogging service that enables its users to send and read text-based posts of up to 140 characters, known as "tweets". The service was launched on July 2006. It rapidly gained worldwide popularity, with over 140 million active users as of 2012 generating over 340 million tweets daily and handling over 1.6 billion search queries per day. Twitter REST API follows a RESTful API design, meaning that applications use standard HTTP methods to retrieve and manipulate Twitter resources. Many API calls require that the user of the application is granted permission to access their data. Twitter uses the OAuth 2.0 protocol to allow authorized applications to access user data. Resources in the Twitter REST API can be represented using JSON, XML, RSS, or ATOM data formats. It also supports pagination. The API provides HTTP GET and POST methods for reading and creating, updating or destroying resources. The authors of the documentation of Twitter REST API have grouped the methods provided by the API into 20 main categories: Timelines, Tweets, Search, Streaming, Direct Messages, Friends & Followers, Users, Suggested Users, Favorites, Lists, Accounts, Notification, Saved Searches, Places & Geo, Trends, Block, Spam Reporting, OAuth, Help, Legal, Deprecated. Twitter also provides free client libraries for various programming languages including Python, PHP, Ruby, Javascript and Java.232528

The general workflow of our tool is shown in Figure 1. The system is initiated using a crawler service which collects feeds corresponding to keywords provided by the user. The collected feeds are then provided as input to the event detection service, which returns a list of events that occurred during a given interval, each corresponding to a cluster of tweets. Finally, the tweet clusters are localized. Event identification was already reported in our prior work. Briefly, time is slotted. In each slot, microblogs emitted in the corresponding window are collected and keyword tuples in them are identified that occur highly disproportionately in the given time window compared to prior slots. Such tuples are said to have a high information gain. Distinct tuples of high information gain are taken as distinct event signatures. Tweets that contain the same signature are assumed to refer to the same event and are thus put into a distinct cluster. Clearly, the technique is inexact. However, leveraging tweet volume, the inexact nature of the above clustering is overcome by outlier elimination once locations have been associated with individual tweets.27

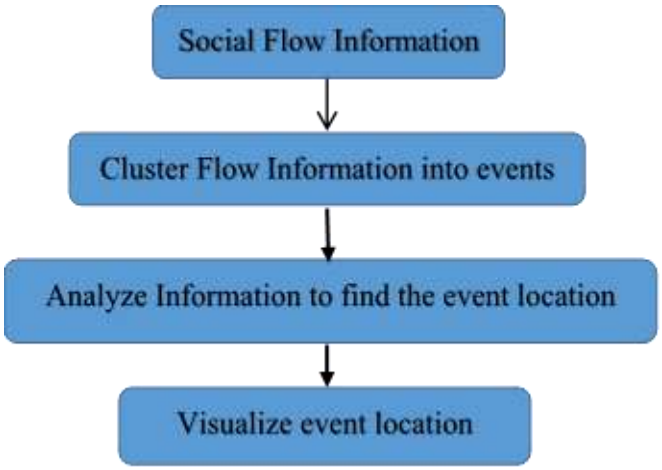



**Fig. 1:-** System Workflow

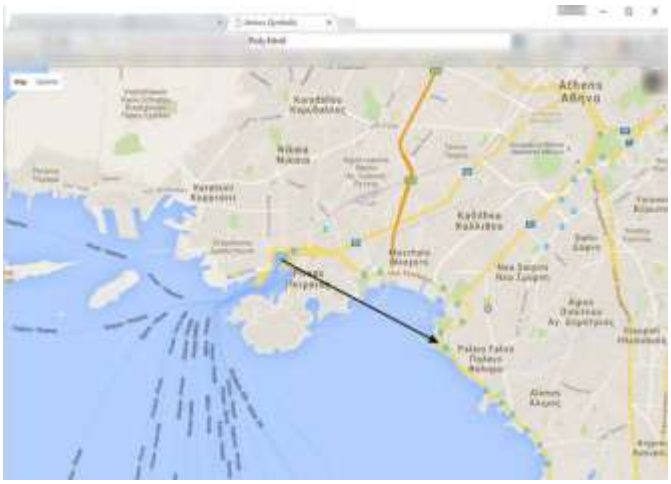
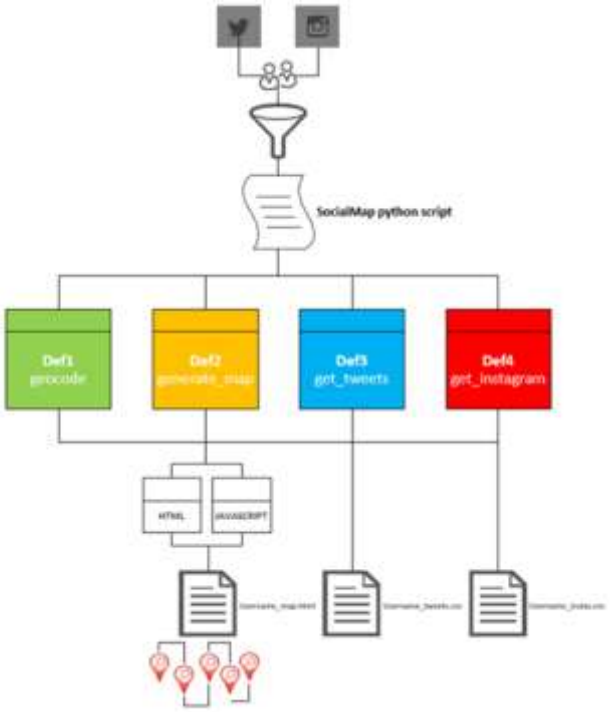


**Fig. 2:-** Google Maps Javascript polyline arrow example




The task of identifying locations associated with an event is challenging due to sparsity of feeds which actually contain location information. In such a scenario the feeds describing physical events usually tend to contain spatial landmarks that can be used as implicit tags in order to determine possible locations. However due to the unstructured nature of these feeds the task requires some preprocessing steps before the spatial information can actually be extracted. The events are then displayed on a graphical map as shown in figures 2 and 3 with markers indicating exact locations. An information pop up box is associated with each event location to offer samples of tweets describing the event in question.

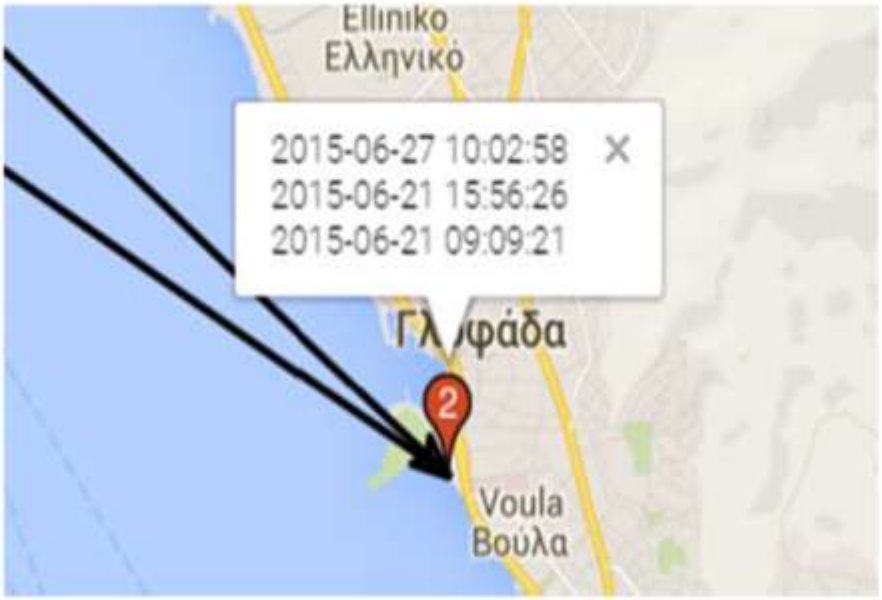



**Fig. 3:-** Gathered records depicted by Google pinpoint Info balloon

Using this integrated library in our python script, isolate the fields of interest of the social media APIs and store the results then in csv files. (Fig. 4)

**Fig. 4:-** JSON data file for Twitter

**Fig. 5** Block diagram of SocialMap python script.

Figure 5 proves the research hypothesis of this python program's objective that a malicious user, utilizing geographic data may withdraw from social media platforms twitter and Instagram for a target user can know almost in real time the location and path of the target over a map. And even anonymously and without leaving traces.

***Geo-Location***

Geo-Location estimation problem has been studied thoroughly by researchers who propose a variety of solutions to extract user location information from internet social media platforms. These social media platforms include web pages and blogs272829 etc. These works rely on external resources such as gazetteers and databases, to identify the related geographical information. In our work, we do not use any external resource to estimate the geographic location of the user. Also, the work in3031 studies the variation of language usage on Twitter. This can also be used to augment our work to improve the accuracy of predicting user geographic location.

There have been works on: relations between geotags, geo-location estimation in search engine query logs29, user privacy of geotags, predicting geographic location on proximity 31, and a study of private information trials using correlations between different publicly available pieces of information to extract private information about a person. Another recent work involves location prediction of Twitter users based on his/her social network37. The most relevant related work is the content-based approach proposed by Liu et al. 32 to estimate the geo-location of a Twitter user. In our work, we use the Rest API and anonymity on Internet through a process which is described below.

***Anonymity***

Starting our methodology, we use a point where internet access cannot connect back to our identity (eg home or relatives home, our work etc.). We are located in the popular international chain of coffee in the city center, which offers free wireless Internet access. We will use our laptop computer by booting from usb live distribution Tails Linux. This distribution offers anonymity and privacy. It uses the Tor network to route all the traffic on the Internet, in order to ensure anonymity. The full name is Tor: The Onion Router (Tor)46. The word onion (= onion) indicates multiple "layers" used during the operation. The objectives most used are many and not all necessarily ethical or legal. The Tor network is a network with multiple computers online (figure 6).

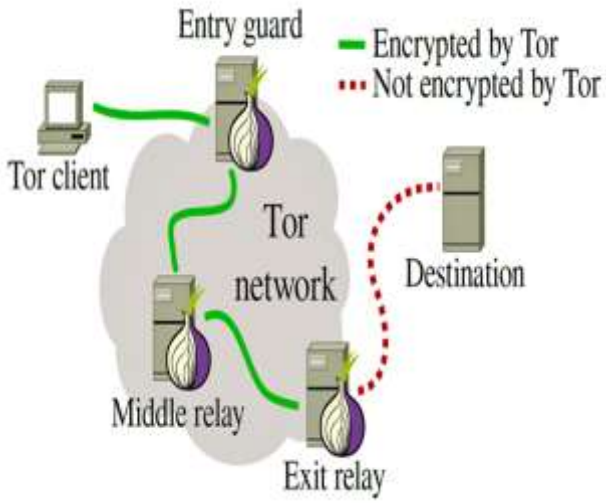



**Fig. 6:-** Routing packets through the TOR network for anonymity.

The number of users connected through Tor, is an important factor as well as the greater the number, the better the overall quality of the network, but even greater individual anonymity that can be achieved. Wanting therefore be some exchange of information from one computer to another, this information will travel encrypted through thousands of alternative paths that give the computers connected to each other. The connection to a node is randomly each time and also is evident the final destination and the starting point in the ip package, ensuring anonymity. Essentially a group of relay nodes that are scattered around the world, are used for mixing and

encryption of Internet traffic in multiple layers (onion layers), so that at the end of the process can be extremely difficult to find from that beginning came the information package. When a user of Tor wants to access a web site, an encrypted request sent by the browser and passed to the Tor network. The first server that receives the request is a «guard» server, which "peel" of the encryption and the request is passed to another randomly selected server. This process repeated until you remove all encryption layers and the last server, the output node, or else exit-node, it forwards the user's browser program request to the actual server hosting the selected site.

Having as a starting point the anonymity through TOR, we created an encrypted mailbox through tutanota.com service. The service during registration doesn't require the declaration of personal data. In addition, the IP addresses of incoming and outgoing email is not recorded in this service, enhancing anonymity. Also offered encryption of emails (subject, content, attachment) and contacts. Encryption and decryption is done on the local computer and not a server (AES 128 bit and RSA 2048 bit). Our email is ready to use within minutes. The email account will be used on all other Online Services that we are going to be registered.

To increase the percentage of anonymity in our methodology, we will use additionally a server cloud with low technical features, in which will be installed a vpn server on Ubuntu 14.04 LTS 32bit. By this way we will add to our tools another level of anonymity. Our connection to the internet it will be according to the following figure.

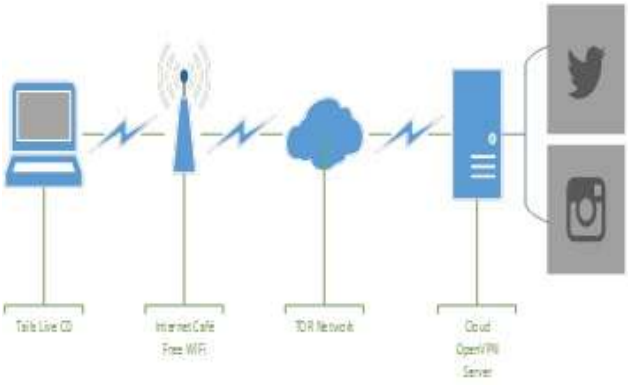



**Fig. 7:-** Anonymous connection to the targets

The next step is to ensure our anonymity during payment for the service of the server cloud without revealing our true identity. The use of credit cards and popular electronic payment methods such as paypal excluded since requiring our identification. The solution for these kind of transactions is the use of electronic currency bitcoin and its derivatives47. According to the page, bitcoin is "a consensual network that enables a new payment system and a completely digital form of money. It is the first decentralized payment network among peers (peer-to-peer) operated by users with no central authority or intermediaries. From the perspective of the user, the Bitcoin is pretty much like cash money on the Internet.

Practically there are three ways for someone to obtain bitcoins:
• Mining directly in bitcoin (cpu, gpu, asic mining).
• Buy a bitcoin via an ATM with euro.
• Mining in alternative electronic currency conversion via exchange service in bitcoin.

The first method is neither efficient nor profitable for someone that has to run the mining project in multiple cpu / cores and multiple graphics cards in the same system or to do scale out, namely to use multiple computers, where there is no need to have the same composition or processing power to run cooperatively the mining project. At the time was written this research as mentioned earlier, it was not cost effective this method.

The second method is relatively new, concerns the use of automated banking machines, but with bitcoins as a medium of exchange and euro, but it does not offer anonymity.

The third method, which is also selected on this methodology is the production of alternative cryptocoins using cpu and gpu and then exchanged into bitcoins. It should be noted that even the specific way it is not economically

profitable since the value of current required for the discovery of cryptocoins is greater than the value of cryptocoins, just give us the undetectable payment system which will use it to buy online services that will help us in our cause.

The digital currency that has been chosen to do mining, is the Monero (XMR). It is a secure, private, undetectable digital currency. Alongside the product is open source, which offers secure transactions using a distributed consensus network (P2P), where each transaction is encrypted. It is safe because the transactions are not publicly visible in the global transaction file (blockchain).

After the collection of well-respected amount of bitcoins, then we searched to find cloud servers services that could accept bitcoins instead of other payment methods (paypal, cretid card) in order to maintain our anonymity. The provider that accepted the bitcoin payment had the same features like any other big cloud server providers (amazon, google, digital ocean etc).

Instagram Real-Time API

Instagram on 2014, has 200 million monthly users, with over 20 billion photos shared, an average of 1.6 billion likes each day, and an average of 60 million photos posted per day 45. Instagram used from a smartphone as an application, which represents different practices and interests, like shared experiences promoted through social media, including the instant publication of statuses and images from the scene of the experience, tagging friends present, commenting on others' content; promoting of small amount of information like photos, 15 second videos and tweets.

The Instagram API3940 provides a search hook dedicated solely to tags, and this offers an immediate comparative opportunity. Calling for the Instagram API for selected tags then provides results akin to similar Twitter projects, although with different metadata, resulting in further methodological questions.

An Instagram API 45 (figure 5) query provides a wealth of metadata for relevant media shared on the platform. For each media object matching the tag query, the API also returns not just its unique identifier (id) and links to the low and standard-resolution versions of the content (whether image or video), but also metadata including usernames, time and date of creation, caption, comments (and user and time information for comments), tags, likes, and location information when a user has geotagged their media. Such data allows for quantitative and qualitative analyses, whether counting the amount of content over time, users, or tags, mapping media based on location data, or looking at the content of the media and their captions. The content of Instagram is more dynamic data points than tweets, though. Again, each image or video is its own data point. However, if a user responds to an image by leaving a comment, that becomes an extra information to the original data point.

Filtering out comments and pondering the media posted on Instagram, since media might attract many or less comments, there is no consistency between data points: while the results can be stored in a database, analysing variable comment threads which may change over the course of the data capture is a new methodological concern which does not affect Twitter research.

Seeking for a specific tag will retrieve information about media published with the relevant tag. Nevertheless, the same search will also provide results about the media which contains the tag in comments even if the original caption does not feature the tag, and if the comment and tag were published by the original user. Including the tag can give to the media a publicity that didn't have previously, adding the differences required for the possible uses and intentions around tagging.

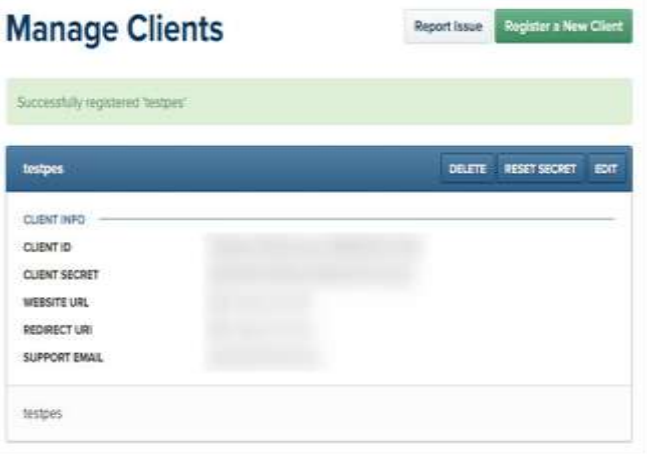



**Fig. 8:-** Instagram API Keys

With Instagram Realtime API4142, someone can check in real time the activity of users, tags and locations of media. So when there is a search of a user via Instagram Real-Time API then information is exported about the new publications. When there is a search regarding labels the exported information according the label then published in the respective media. Through the searching of locations there are new notifications about photos or videos uploaded and tagged with a specific location. Also data can be exported in connection with new photos posted to an arbitrary location.

Google maps JavaScript API
The Google maps API43 provides the functionality of google maps with fast and direct way. It is a web application service that is provided by Google and offers street maps and navigation services to websites or mobile applications.

The services offered are:
• Create and display the map.
• Import markers (pinpoints), polygons, popups, polylines (figure 7), Info Window.
• Event handlers.
• Geocoder: Coordinates / address translation service.
• Direction: Route and package design. route navigation (by car, by public transport or on foot).
• Identifying business in countries around the world.

To Google Maps API is free for commercial use, provided that the application in which it is used will be accessible to the public without charge the user for each access, and do not produce more than 25,000 accesses map per day43. Of course there are premium packages for a fee covering increased needs in applications and use. For the case of SocialMap python script, the specifications of the free package is more than enough.

The question in the Google Maps API is done by sending an HTTP GET request to the Web application, and returns the response in XML or JSON format. The application described in this document uses JSON messages as we have mentioned. To use the Google Maps API from an application is required to obtain a "key" (Google Maps API Key) from the creator of the application that uses it and the introduction of the application code.

The Google Maps JavaScript API capabilities we use to SocialMap script are:
• Markers (pinpoints).
• InfoWindow with activation of the marker touch event (On Click).
• Polylines.
• Moving the map center.

The Google maps geocoding API44, refers to the function of converting postal addresses to geographic coordinates, which then we can illustrate with the help of Google maps JavaScript API.

In cases where users of Twitter and Instagram, use the service via their smartphone, but having disabled the function of GPS, we can use the parameter of location tags (instead of coordinates tags), to obtain a less accurate estimate of their location (within about 1 Km). This way we have more number of records related to location, and therefore have more places on the map user.

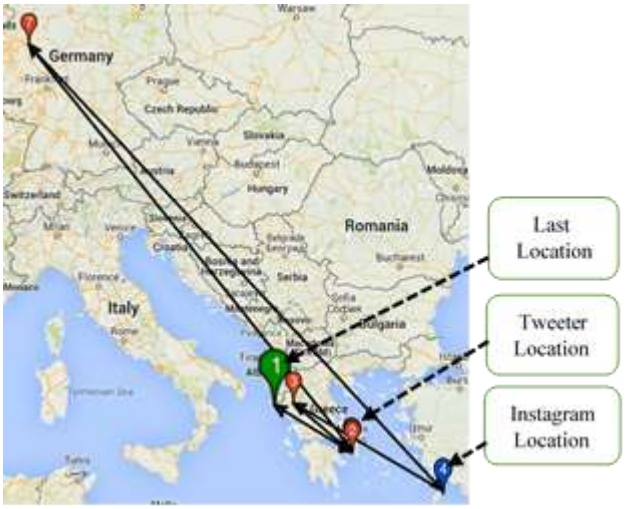



**Fig. 8:-** Example of polyline use

## Conclusion:-

The basic research question was argued in this article, it had to do with whether it is permissible for a malicious user to monitor, without the consent of the target user, the activity of the two major social networks Twitter and Instagram and draw conclusions, data relating to the location of each point in time.

Open-source intelligence (OSINT) is intelligence collected from publicly available sources. In the intelligence community, the term "open" refers to overt, publicly available sources. Following OSINT techniques and methods analyzed in the respective chapters we have managed to gain access to Twitter, Instagram and Google APIs without revealing our true identity and purpose and without something can be linked back to us at least easily.

We can fully protect us as end users from such malicious actions? Probably not 100%, because the very nature of social media itself, requires the creation of content by the user himself, which unfortunately tends to convey the personality of the medium. The use of common sense, the separation and the preservation of sensitive personal information and raise awareness on security issues (security awareness), can help to significantly reduce our exposure to such attacks.

On proposals for improvements of this work in the future, we could mention the support of more social networking platforms and the creation of a graphic web interface / application with friendly user operating environment. Furthermore, they could be integrated functions to create automated reports per user view, for a given specific time period, and the corresponding degree of exposure. Another possibility - extension that could be incorporated is predicting the future position of the target based on past history and the frequency of visits to a particular site. Finally, could the particular job, be adapted from big companies, which would like to determine the degree of threat of leakage of important information from their own employees (insider threats) or to capture the psychological - the social employee profile in critical positions, as inferred from their presence in social media.

## References:-

1. Giordano, Andrea, Giandomenico Spezzano, Harry Sunarsa, and Andrea Vinci. "Twitter to integrate human and Smart Objects by a Web of Things architecture." In Computer Supported Cooperative Work in Design (CSCWD), 2015 IEEE 19th International Conference on, pp. 355-361. IEEE, 2015.
2. Cho, Hae-Ryong, and Min Choi. "Replacing Socket Communication by REST Open API for Acquisition Tax Analyzer Development." In Advanced Information Networking and Applications Workshops (WAINA), 2014 28th International Conference on, pp. 462-468. IEEE, 2014.

3. Jaiswal, Ayush, Wei Peng, and Tong Sun. "Predicting time-sensitive user locations from social media." In Advances in Social Networks Analysis and Mining (ASONAM), 2013 IEEE/ACM International Conference on, pp. 870-877. IEEE, 2013.
4. Das, Tushar Kanti, D. P. Acharjya, and M. R. Patra. "Opinion mining about a product by analyzing public tweets in Twitter." In Computer Communication and Informatics (ICCCI), 2014 International Conference on, pp. 1-4. IEEE, 2014.
5. Lim, B.H., Lu, D., Chen, T. and Kan, M.Y., "# mytweet via Instagram: Exploring User Behaviour across Multiple Social Networks". In Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 113-120, ACM, 2015.
6. McCreadie, Richard, Ian Soboroff, Jimmy Lin, Craig Macdonald, Iadh Ounis, and Dean McCullough. "On building a reusable Twitter corpus." InProceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval, pp. 1113-1114. ACM, 2012.
7. Rusk, David, and Yvonne Coady. "Location-based analysis of developers and technologies on github." In Advanced Information Networking and Applications Workshops (WAINA), 2014 28th International Conference on, pp. 681-685. IEEE, 2014.
8. Cieszko, Michal, and Jaroslaw Legierski. "Graph Based Messaging APIs-concept and implementation." In FedCSIS, pp. 925-932. 2014.
9. Giridhar, Prasanna, Tarek Abdelzaher, Jemin George, and Lance Kaplan. "Event localization and visualization in social networks." In Computer Communications Workshops (INFOCOM WKSHPS), 2015 IEEE Conference on, pp. 35-36. IEEE, 2015.
10. Lin, Jie, and Robert G. Cromley. "Evaluating geo-located Twitter data as a control layer for areal interpolation of population." Applied Geography 58 (2015): 41-47.
11. D. Kotzias, T. Lappas, D. Gunopulos. "Home is where your friends are: Utilizing the social graph to locate twitter users in a city." Information Systems (2015).
12. J. Eisenstein, B. O'Connor, N.A. Smith, E.P. Xing, A Latent variable model for geographic lexical variation, in: Proccedings of the 2010 Conference on Empirical Methods in Natural Language Processing, EMNLP'10,2010, pp. 1277-1287.
13. A. Ahmed, L. Hong, A.J. Smola, Hierarchical geographical modeling of user locations from social media posts, in: Proceedings of the 22nd International Conference on World Wide Web. WWW '13, 2013, pp. 25–36.
14. Z. Cheng, J. Caverlee, K. Lee, You are where you tweet: a content-based approach to geo-locating twitter users, in: Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM '10, 2010, pp. 759–768.
15. J. Mahmud, J. Nichols, C. Drews, Where is this tweet from? Inferring home locations of twitter users, in: J.G. Breslin, N.B. Ellison, J.G. Shanahan, Z. Tufekci (Eds.), ICWSM, The AAAI Press, Dublin, Ireland, 2012.
16. K. Ren, S. Zhang, H. Lin, Where are you settling down: geo-locating twitter users based on tweets and social networks, in: Information Retrieval Technology, Lecture Notes in Computer Science, vol. 7675, 2012, pp. 150–161.
17. Y. Ikawa, M. Enoki, M. Tatsubori, Location inference using microblog messages, in: Proceedings of the 21st International Conference Companion on World Wide Web, WWW '12 Companion, 2012, pp. 687–690.
18. W. Li, P. Serdyukov, A. de Vries, C. Eickhoff, M., Larson, The where in the tweet, in: Proceedings of the 20th ACM International Conference on Information and knowledge management, CIKM '11, 2011, pp. 2473–2476.
19. S. Kinsella, V. Murdock, N. O'Hare, "I'm eating a sandwich in Glas- gow": modeling locations with tweets, in: Proceedings of the 3rd International Workshop on Search and Mining User-Generated Contents, SMUC '11, 2011, pp. 61–68.
20. E. Cho, S.A. Myers, J. Leskovec, Friendship and mobility: user movement in location-based social networks, in: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Dis- covery and Data Mining, KDD '11, 2011, pp. 1082–1090.
21. M. Lichman, P. Smyth, Modeling human location data with mixtures of kernel densities, in: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '14, ACM, New York, NY, USA, 2014, pp. 35–44. URL ⟨http://doi.acm. org/10.1145/2623330.2623681⟩.
22. Paredes-Valverde, Mario Andrés, Giner Alor-Hernández, Alejandro Rodr'guez-González, and Gandhi Hernández-Chan. "Developing Social Networks Mashups: An Overview of REST-Based APIs." Procedia Technology 3 (2012): 205-213.
23. Shankar, Pravin, Yun-Wu Huang, Paul Castro, Badri Nath, and Liviu Iftode. "Crowds replace experts: Building better location-based services using mobile social network interactions." In Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on, pp. 20-29. IEEE, 2012.

24. Togias, Konstantinos, and Achilles Kameas. "An ontology-based representation of the Twitter REST API." In Tools with Artificial Intelligence (ICTAI), 2012 IEEE 24th International Conference on, vol. 1, pp. 998-1003. IEEE, 2012.
25. Bhat, F., M. Oussalah, K. Challis, and T. Schnier. "A software system for data mining with twitter." In 2011 IEEE 10th International Conference on Cybernetic Intelligent Systems (CIS). 2011.
26. B. Huberman and D. R. F. Wu. Social networks that matter: Twitter under the microscope. First Monday, 14, 2009.
27. J. Eisenstein, N. Smith and E. Xing, A Latent Variable Model for Geographic Lexical Variation, EMNLP '10 Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing.
28. Vis, Farida. "A critical reflection on Big Data: Considering APIs, researchers and tools as data makers." *First Monday* 18.10 (2013).
29. Romsaiyud, Walisa. "Detecting emergency events and geo-location awareness from twitter streams." *The International Conference on E-Technologies and Business on the Web (EBW2013)*. The Society of Digital Information and Wireless Communication, 2013.
30. Jurgens, David, et al. "Geolocation prediction in Twitter using social networks: a critical analysis and review of current practice." *Proceedings of the 9th International AAAI Conference on Weblogs and Social Media (ICWSM)*. 2015.
31. Ghahremanlou, Lida, Wanita Sherchan, and James A. Thom. "Geotagging twitter messages in crisis management." *The Computer Journal* (2014): bxu034.
32. Liu, Chun, ed. *Principle and application progress in location-based services*. Springer, 2014.
33. Bruns, Axel, and Stefan Stieglitz. "Towards more systematic Twitter analysis: Metrics for tweeting activities." *International Journal of Social Research Methodology* 16.2 (2013): 91-108.
34. Daume, Stefan. "Mining Twitter to monitor invasive alien species—An analytical framework and sample information topologies." *Ecological Informatics* 31 (2016): 70-82.
35. Zhu, Zack, Ulf Blanke, and Gerhard Tröster. "Recognizing composite daily activities from crowd-labelled social media data." *Pervasive and Mobile Computing* (2015).
36. Di Pietro, G., Aliprandi, C., De Luca, A. E., Raffaelli, M., & Soru, T. (2014, August). Semantic crawling: An approach based on Named Entity Recognition. In *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on* (pp. 695-699). IEEE.
37. Rahimi, Afshin, Trevor Cohn, and Timothy Baldwin. "Twitter user geolocation using a unified text and network prediction model." *arXiv preprint arXiv:1506.08259* (2015).
38. Cha, Miriam, Youngjune Gwon, and H. T. Kung. "Twitter geolocation and regional classification via sparse coding." *Proceedings of the 9th International Conference on Weblogs and Social Media (ICWSM 2015)*. 2015.
39. Ferrara, Emilio, Roberto Interdonato, and Andrea Tagarelli. "Online popularity and topical interests through the lens of instagram." In Proceedings of the 25th ACM conference on Hypertext and social media, pp. 24-34. ACM, 2014.
40. Highfield, Tim, and Tama Leaver. "A methodology for mapping Instagram hashtags." First Monday 20, no. 1 (2014).
41. Peng, Rong, Dong Sun, and Wei-Tek Tsai. "Success factors in mobile social networking application development: case study of instagram." In Proceedings of the 29th Annual ACM Symposium on Applied Computing, pp. 1072-1079. ACM, 2014.
42. Reuter, Christian, and Simon Scholl. "Technical Limitations for Designing Applications for Social Media." In Mensch & Computer Workshopband, pp. 131-139. 2014.
43. Google Maps.. Available from https://developers.google.com/maps/documentation/javascript/ [Accessed 28 August 2015]
44. Google Geocoding API. Available from https://developers.google.com/maps/documentation/geocoding/intro [Accessed 28 August 2015]
45. Instagram., Blog. Available from http://blog.instagram.com/post/8756150468/a-real-time-api-for-next-generation-apps , [Accessed 1 October 2015]
46. Weinberg, Zachary, Jeffrey Wang, Vinod Yegneswaran, Linda Briesemeister, Steven Cheung, Frank Wang, and Dan Boneh. "StegoTorus: a camouflage proxy for the Tor anonymity system." In Proceedings of the 2012 ACM conference on Computer and communications security, pp. 109-120. ACM, 2012.
47. Bitcoin. (2014) Available from https://bitcoin.org/el/faq#what-is-bitcoin . [Accessed 25 September 2015].