Intelligent Environments 2020
C.A. Iglesias et al. (Eds.)

doi:10.3233/AISE200065

Towards a Web Tool for the Analysis of Twitter Profiling Information

Feliciano J. Gómez-Fernandez[a] and Fernando Terroso-Sáenz[a,1]
[a] *Universidad Católica de Murcia (UCAM), Murcia, Spain*

Abstract. In these days and age, Online Social Networks (OSNs) are common ways to keep us in touch with relatives, friends and colleagues. As a result, we are used to share many personal information and opinions in these platforms. However, this might be dangerous for certain types of users, specially the underage ones. For that reason, the present work introduces a preliminary version of a web tool that, on the basis of the posts shared by a user in Twitter, infers and extracts certain personal information from these posts and present it to the user. The goal is to raise awareness among OSN users about how easy it is to uncover aspects of their private life from the data that they share in an Internet platform.

Keywords. Online Social Networks, privacy, data analysis

1. Introduction

Nowadays, online social networks (OSN) are widely spread in our society. As matter of fact, Facebook has around 2,449 million active users per month whereas Twitter has roughly 340 million active users per month [1].

Given this scenario of massive use of OSNs, one of the key drawbacks of these platforms has to do with the privacy of users, with special interest in the youngest ones because they are more vulnerable. They are not aware that the information they share in these platforms, in either a deliberately or unconsciously manner, could be used to cause them some kind of harm with the right processing. For example, 64% of children between 9 and 16 years old say they include a photograph of their face in their OSN profile, 50% say they include their surname and 29% their school [2].

For this reason, the present work introduces "What Do They Know About Me" (WDTKAM), a novel web application to show the valuable personal information that can be inferred about a person through the analysis of the content that he posts in the OSN Twitter. It is true that there are several existing tools that provide reports about the usage profile of Twitter users. However, most of them focus on providing just quantitative information that maybe not fully understandable by first-time users. Therefore, the goal of WDTKAM is to provide a solution whose analytical outcome can be easily understandable by a wide range of users and, thus, encourage them to adopt basic safety rules in their interaction with OSNs.

The remainder of the paper is structured as follows. Next, an overview the relevant related work is put forward in section 2. Section 3 is devoted to describing in detail the logic structure and the processing stages of the tool. Then, section 4 discusses the main

[1] Fernando Terroso-Sáenz, Universidad Católica de Murcia (UCAM), Campus de los Jerónimos, 30107, Murcia, Spain; E-mail: fterroso@ucam.edu.

results of the performed experiments. Finally, the main conclusions and the future work are summed up in section 5.

2. Related Work

In the literature we can find several web-based tools that carry out different types of analysis of Twitter profiles, publication of tweets, followers and friends, geo-localization of profiles, etc. Here we list the foremost existing solutions based on the surveys put forward in [3, 4, 5].

- foller.me[2]: this application a provides valuable information about any public Twitter profile, collecting almost real-time data on topics, mentions to other profiles, hashtags, followers, location, etc. It is worth noticing that the location-extraction feature is not currently supported in the trial version.
- accountanalysis.app[3]: this web application allows to evaluate Twitter accounts, showing information about profile automation, number of published retweets or which websites the target account frequently links. It also shows statistical data about daily-publication rates in a graph-based format that compares days of the week with hours of the day, as well as type of tweet, publication interface, hashtags classification and cited users.
- twitonomy[4]: this tool, apart from offering useful information to companies in the marketing field, also offers relevant information on user profiles such as statistical data on usage and information of mentions to other profiles, hashtags, followers. It is quite similar to foller.me.
- followerwonk[5]: this tool has some free functionalities but the detailed account reports can be only generated by paid version of the tool. Like *twitonomy* and *foller*, it shows statistical data about the target profile, such as hours of use by types of tweets, running time of the followed profiles since their creation, etc. As a novelty, this tool provides a map with approximate locations extracted from the location field of the followed users' profiles. This might lead to some inaccuracies because this field is text free.
- Twitter Analytics[6]: it is a the inner tool offered by Twitter, which can be accessed simply by having an account and activate the functionality. It shows data about our tweets, followers and friends, statistical data about the use of the account, summary of activity and so forth.

For the sake of clarity, Table 1 shows a comparative summary of the aforementioned tools and our proposal based on 8 features that have been considered significant when it comes to analyze a Twitter profile, namely,

---

[2] https://foller.me

[3] https://accountanalysis.app

[4] https://www.twitonomy.com

[5] https://followerwonk.com

[6] https://analytics.twitter.com/about

- Friends: information about the users who are more socially close to target account.
- Usage: information about the time-based use of the OSN split in days of the week, hours of the day, months, years, etc.
- Interfaces: information about the devices or client applications (e.g. mobile phone, web browser and the like) used by the target account to interact in the OSN.
- Multimedia: information about publications (tweets) in which some kind of multimedia content (images, gifs or videos) has been included.
- Interests: information about the interests, hobbies or tastes of the target account.
- Location: information about spatial places where the user of the target account has been according to the geo-tags of the posted.
- Languages: information about the languages used in the publication of the tweets.

	WDTKAM	Foller.me	Accountanalysis	Twitonomy	Followerwonk	Twitter Analytics
Friends	✓	✓	✗	✗	✗	✗
Usage	✓	✓	✓	✓	✓	✓
Interfaces	✓	✓	✓	✓	✗	✗
Multimedia	✓	✓	✗	✗	✗	✗
Interests	✓	✗	✗	✗	✗	✗
Location	✓	✗	✗	✓	✓	✗
Languages	✗	✗	✓	✓	✓	✗

Table 1. Comparison between our solution and existing ones for Twitter profile analysis. A check or cross mark indicates that a feature is supported or not by the tool.

As we can see from Table 1, our tool WDTKAM fills in the gaps found in the rest of the existing tools in terms of raise awareness among Twitter users about their privacy and the information they contribute to social networks.

The key difference of WDTKAM with respect existing tools is that in our approach we analyze not only the account of a target user (TU) but also information about its related profiles (friends, followers and the like). This will be described in detail in sec. 3.3. Furthermore, the interface of the tool provides the information in a very verbose approach so that the outcome of the analysis is provided to users in a very clear format. In that sense, our work has certain similarities with the tool presented in [6]. That tool proposes a web-tool to analyze the privacy leaks of a given web page. Like WDTKAM, an easy-to-read interface is presented. However, it focuses on a completely different scenario.

3. Tool "What Do They Know About Me" (WDTKAM)

This section describes in detail WDTKAM tool. We should point out that the source code of the tool is freely available in a GitHub public repository[7]. In that sense, Fig. 1 shows the operational workflow of the application. This pipeline comprises the following steps.

To begin with, the TU introduces his Twitter account name (arrow 1 in Fig. 1). Next, WDTKAM makes several calls to the Twitter API different types of information of the provided account name such as the user's profile information (arrow 2), his followers (3), friends (4), tweets (5), favorites (6), lists (7,8,9), locations (10) and received retuits (11). The information obtained from the API is processed and analyzed by the application (12), showing the result of the analysis in a graphic interface (13).

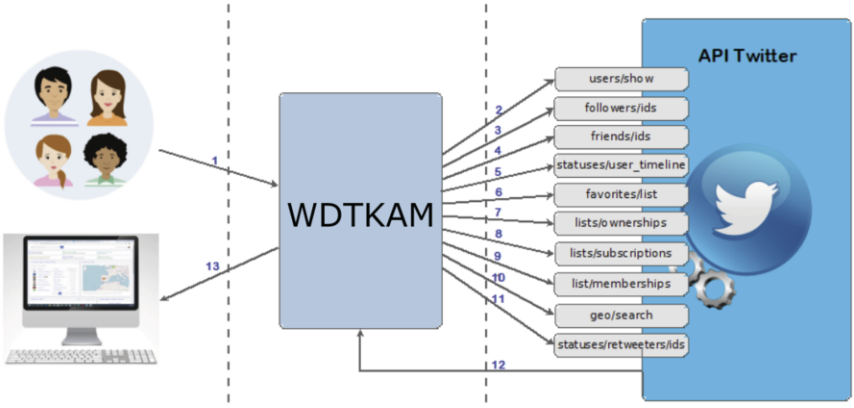

Figure 1. General flow of WDTKAM

3.1. Usage of the Twitter Rest API

Twiter provides a ready-to-go Application Programming Interface (API) that allows to extract a varied range of data for a particular account. This API is structured in different methods/calls following a REST approach. In that sense, the Twitter API methods used by WDTKAM to obtain the information about the TU are explained below, following the same order than in Fig. 1.

- GET (`users/show`)

This method allows to extract many different information associated with the profile such as name, email, location, description, date of creation, url of profile images, number of followers and followed, if the profile is protected and if it is a verified profile.

- GET (`followers/ids`)

This method returns the followers of the TU.

- GET (`friends/ids`)

[7] https://github.com/javicanin/QueSabenDeMi

The method is extracts the list of user identifiers that are followed by the TU.

- GET (`statuses/user_timeline`)

This method retrieves a list of the most recent tweets published by the TU. For each tweet, the API returns the date of creation, the text, location if it was active, hashtags and mentions, added multimedia content, to all kinds of information about retuits and responses.

- GET (`favorites/list`)

  This method returns a list of the last tweets marked as favorites by the TU.

- GET (`lists/ownerships`)

The invocation of this end point returns a structure containing information about the lists on which the TU is the owner or creator.

- GET (`lists/subscriptions`)

This method returns the lists to which the TU is subscribed.

- GET (`list/memberships`)

This call provides the lists of users of which the TU is a member.

- GET (`geo/search`)

The information returned by this method contains data about locations such as name, country, coordinates and a generic administrative classification such as country, city, neighborhood, etc.

- TSG (`statuses/retweeters/ids`)

  This method retrieves a list of user IDs that have retweeted any tweet created by the target account.

3.2. API limitations

It is worth noticing that Twitter limits the number of calls that an API end-client can made. This limitation which depend on the query made or endpoint invoked. Generally speaking, all queries have a 15-minute window and a call limitation for that window. For example, the `GET statuses/user_timeline` query is limited to 900 calls during that 15-minute time window and each call returns a maximum of 200 results. To overcome this situation, WDTKAM makes use of cursors that allow the incremental extraction of the data.

3.3 Analysis of the information

All the data obtained with the aforementioned methods, WDTKAM performs the some analytical tasks. As a result, some personal and private information items related to the target account are extracted. Here we list these items.

3.3.1. Mutual followers

Two users are said to be mutual followers when each of them follows the other in a reciprocal way, that is, each of them has the other in their "followers" and "following" lists. The list of mutual followers is of great importance for the tool because it is used to analyze the information about the profiles that are closest to the TU.

To do so, the tool intersects the sets of users extracted from the methods `followers/ids` and `friends/ids` described above.

3.3.2. Friends

Another information item that WDTKAM extracts is the list of Twitter users that may be friends to the TU in the real world. The procedure to obtain the user profiles that are the closest to the TU is as follows.

Firstly, the list of mutual followers described in the previous section is filtered by only keeping the users of that list the accomplish the two the following conditions,

a) some of their tweets have been *retweeted* or *replied-to* by the TU,

b) some of their tweets have been marked with a “like" by the TU,

Secondly, each account in the resulting list is evaluated from two different points of view to decide whether he/she is really a friend of the TU in real life,

- one quantitative that takes into account the number of interactions between the TU and the account, and
- another qualitative that also observes if the interactions are distributed homogeneously in the two criteria described above, taking into account the reciprocity in the interactions.

For example, if the TU has made a total of 100 interactions with another user, it would be reasonable to think that they might have a closer relationship based on the quantitative criterion. Then, the qualitative criterion also observes whether the interactions between the TU and the other user are more os less balanced.

3.3.3 Hobbies or interests

An innovative feature of WDTKAM with respect to the state of the art is the analysis of possible hobbies or interests of the TU.

In the current version of the tool, it only considers two pre-define topic, video-games and Spanish politics. Then, for each topic, a pre-defined list of Twiter profiles or entities considered influential or relevant to the any of the two subjects is defined. These profiles are normally marked as *verified* by the Twitter managers.

Next, we obtain the number of profiles of each topic that are followed by the TU. Depending on the resulting number, WDTKAM infers how relevant the topic is for the TU based on pre-defined ranges.

3.3.4 Locations

Finally, WDTKAM tries fo infer the TU's visited locations through the location field that is freely filled in on his profile. Furthermore, it also checks the meta-data of his tweets so as to detect if any of them includes a geo-reference field.

In that sense, it is well-known that only a very small percentage of Twitter users (around 1%) have activated the geo-tagging of their tweets [7]. Besides, the location feature of a Twitter user profile is just a text-free field. Therefore, many users enter textual data that has nothing to do with locations, resulting in unpredictable responses.

Consequently, in most of the cases, the tool will not be able to infer reliable location data from a given TU.

4. Case of Use

In this section, we describe the full functionality of WDTKAM from a user point of view.

The application has an initial view that allows a TU to introduce his Twitter username as Fig. 2 depicts. Then, the tool performs the API calls and analyses the data as described in sec. 3.3. Finally, the tool transitions to a second view displaying the results of such an analysis (see Fig. 3).

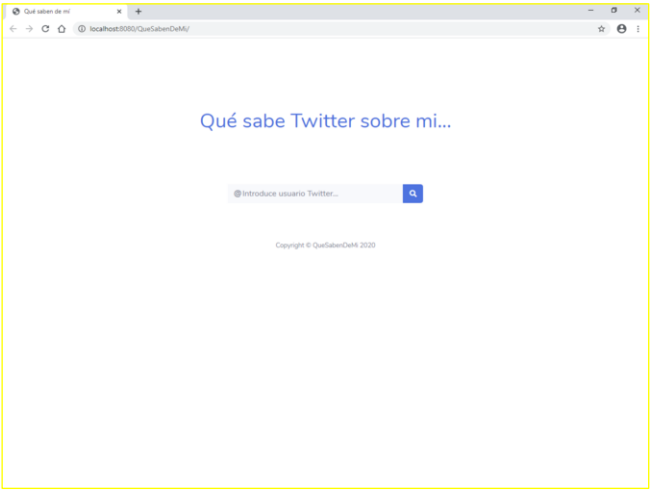

Figure 2. Initial view of WDTKAM (Spanish version).

This second view is divided into two areas. The top area contains a search field to introduce a new target account. Besides, it also contains the profile image of the TU and his user name. The second area corresponds to the rest of the view, which displays the key results of the analysis. Each type of result is shown in a particular panel as listed below,

- Target-account general information: This area shows a palette of small panels containing user-profile information.
- Best Friends or preferred profiles (*"Tus mejores amigos o perfiles preferidos"* in Fig. 3): This panel comprises a table with the user profiles that have been considered to have a closer relationship with the TU according to the procedure described in section 3.
- Map of Locations (*"Dónde usas Twitter"* in Fig. 3): This panel contains the map of locations extracted from the TU´s tweets, if at any time this option was active. This panel also shows the locations obtained from the TU profiling information.
- Twitter usage (*"Cuándo usas Twitter"* in Fig. 3): This panel comprises a plot that informs about the frequency of active use (publications) of Twitter in the format of hours of the day, days of the week, months and years.

- Interface or use devices (*"Desde el teléfono, ordenador, aplicación"* in Fig. 3) : It shows a graph with the number and type of devices that the TU uses when connects to Twitter. A detailed view of this panel is depicted in Fig. 4.
- Multimedia: this panel shows the tweet with multimedia content (images, animated gifs or videos) with the highest impact among the TU's followers.
- Interests: this last section shows phrases related to the tastes, hobbies or interests of the target account.

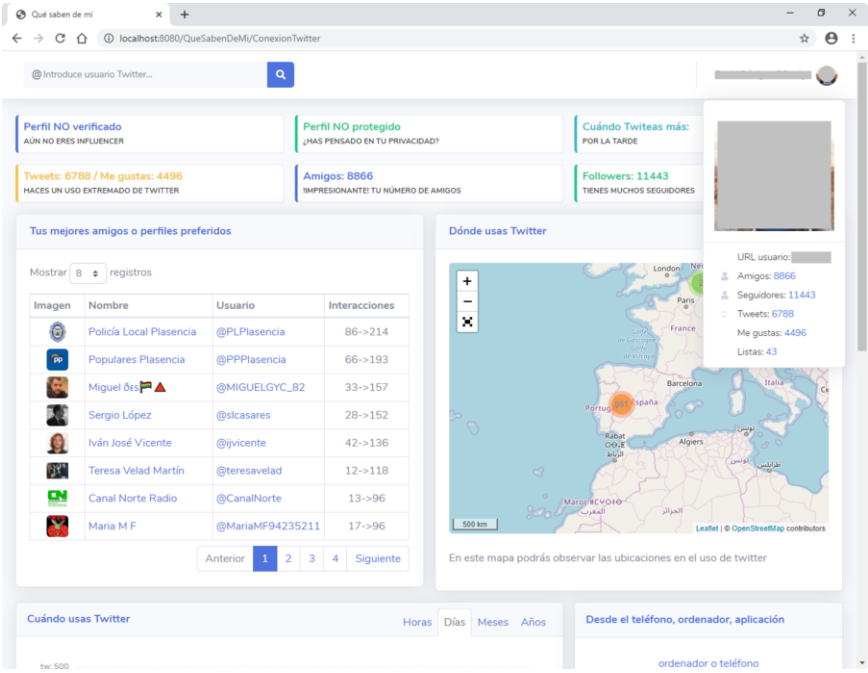

Figure 3. Example of tool view with the analysis results (Spanish version).

Figure 4. Example of the tool panel with the devices that the TU has used to interact with Twitter (Spanish version).

As we can see, the interface of the tool intends to present the results of the analysis in a quite clear manner. This way, a user can realize the simple but meaningful private information that can be easily extracted from the posts that they share on Twitter.

5. Conclusions and Future Work

Nowadays, the ubiquity of OSNs makes users to not be fully aware that using these platforms make them to share personal information that can be accessed anywhere in the world, with the consequent privacy issues.

In this scope, the present work introduces WDTKAM, a web tool that tries to offer a different perspective on OSNs, without diminishing the importance of the great social work they perform. Its objective is to raise awareness among OSN users about the privacy issues that the information that they share in OSNs might cause.

To do so, WDTKAM makes use of the functionality provided by Twitter so as to collect all the possible information about a particular account. Then, a simple but effective analysis of the crawled data is done. Finally, the key results are shown in a clear manner by means of a web-based view.

Future work will focus on improving the analysis made by the tool. In that sense, the usage of techniques from the Natural Language Processing (NPL) field will be deeply studied. This way, the tool will also able to analyze the content of the tweet so as to, for example, uncover the potential hobbies of the target account in a more accurate manner.

References

[1] Mejía Llano J. C. Estadísticas de redes sociales 2020: usuarios de Facebook, Instagram, Youtube, Linkedin, Twitter, Tiktok y otros. From https://www.juancmejia.com/marketing-digital/estadisticas-de-redes-sociales-usuarios-de-facebook-instagram-linkedin-twitter-whatsapp-y-otros-infografia, 2020.

[2] UNICEF Comité Español. Los niños y niñas de la brecha digital en España. From https://www.unicef.es/publicacion/estado-mundial-de-la-infancia-2018-los-ninos-y-ninas-de-la-brecha-digital, 2018

[3] Vermeren I. The Best Twitter Analytics Tools. From https://www.brandwatch.com/blog/twitter-analytics-tools, 2019.

[4] Ángeles M. Las mejores herramientas de pago para análisis de Redes Sociales. From https://velogig.com/las-mejores-herramientas-de-pago-para-analisis-de-redes-sociales. 2019

[5] Gómez Henao D. Top 5 de herramientas de Social Listening para tu negocio. Recuperado de https://www.smdigital.com.co/articulo/top-5-de-herramientas-de-social-listening-para-tu-negocio/, 2020.

[6] Massardier-Meca M., Ruiz-Martínez A., Towards a Privacy Web Scanner for End-Users, Communications in Computer and Information Science, Volume 1208, 2020.

[7] Huang B., & Carley K. M. A large-scale empirical study of geotagging behavior on Twitter. In Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (pp. 365-373), 2019.