# Data Retrieval

## 1. Instagram API

Instagram Platform offers an Application Programming Interface (API), that can be used manage, retrieve and share the data on the Instagram. The Instagram API Endpoints can be fetched over the secure SSL protocol (https) and are located at the api.instagram.com. In order to be able to send https requests to the Instagram from the client side (in our case from the R script), following steps are necessary (cf. Instagram.com, 2016):

1. Register the client on the Instagram API homepage
2. Authorize the application by request the so-called Instagram access\_token by directing the user to the Instagram authentication
3. Start making the request to the Instagram API using the received access\_token

For the purpose of this paper the steps 2 and 3 could be automated using the capabilities of the R programming language for sending the https request and receiving and processing the Instagram response, that comes by default in the so-called JavaScript Object Notation (JSON) format (cf. Instagram.com, 2016). The created R script then processes the received JSON response of the Instagram API and stores the data into the simple R data frame.

For the purpose of this seminar paper the so-called Instagram Tag Endpoint was used. The corresponding https request requires apart from the valid access\_token (see above) the tag-name as the input parameter. Sending this https request it is possible to search the Instagram for the media units that contain the given hashtag (cf. Instagram.com, 2016a). In order to stay in the scope of this paper, no API details will be further discussed, further details on the authentication and requesting of the Instagram API can be found on the Instagram API homepage.

## 2. Data retrieval

The data retrieval was accomplished between the 11th and 13th of December 2015. The hashtag “#refugees” for the reasons of its direct relevance for the research topic has been used as a starting tag for the retrieval of the dataset from the Instagram. The media-request for this hashtag resulted in 112617 unique media units. The first exploratory data analyses have made it evident, that there are certain hashtags, that for the reason of its semantic meaning can be seen as once, that carry a typical, truly positive or negative sentiment towards the refugees-crisis. The following table shows the most frequent hashtagas for the both categories.

As can bee seen from the Table 1, the proportion of the clearly negative tags is very small. In order to improve the proportion of the negative tags, the data with the hashtags "refugeesnotwelcome”,“norefugees”, “nomorerefugees”, “wakeupeurope”, “überfremdung” has been retrieved from Instagram additionally. Furthermore, data for the following neutral, topic related tags has been additionally retrieved: “syrianrefugees”, “refugeecrisis”, “flüchtling”, “flüchtlinge”. The resulting dataset of 147454 data units was considered as big enough for this research.

|  |  |  |  |
| --- | --- | --- | --- |
| Positive | Frequency | Negative | Frequency |
| refugeeswelcome | 16154 | refugeesnotwelcome | 161 |
|  |  | norefugees | 89 |
|  |  | nomorerefugees | 63 |
|  |  | wakeupeurope | 54 |
|  |  | stopislam | 46 |
|  |  | fuckrefugees | 5 |
|  |  | überfremdung | 2 |

Table : Frequency of the tags with truly positiv&negativ semantics after the first data load

## 3. Structure of the dataset

Table 2 shows the media unit’s meta data returned when requesting the Instagram Tag Endpoint, that was considered to be relevant for the research.

|  |  |
| --- | --- |
| Meta Data | Description |
| “Type” | Media type: Image / Video |
| “Tags” | An array of HashTags used in the media unit description |
| “Caption: text” | Media unit description used by the author (containing some of the HashTags) |
| “Caption: created\_time” | Tie of the creation of the media in the Linux specific notation |
| “User: full\_name” | Full name of the author |
| “User: username” | Instagram username of the author |
| “User: userid” | Instagram specific userid of the author |
| “Id” | Unique media ID (primary key) containing the user id |
| “Likes: count” | Number of the times the corresponding media unit was liked by other users |
| “Link” | URL to the Instagram media unit |
| “Location: longitude” | Longitude of the media unit |
| “Location: latitude” | Latitude of the media unit |
| “Location: id” | Id of the specific location, can be used to retrieve further location specific data |

Table 2: Retrieved Instagram media-unit’s met-data that is considered to be relevant for this research (own representation based on Instagram.com, 2016a)

Sources:

Instagram.com, (2016). *Instagram Developer Documentation*. [online] Available at: https://www.instagram.com/developer/ [Accessed 3 Jan. 2016].

Instagram.com, (2016 a). *Tag Endpoints • Instagram Developer Documentation*. [online] Available at: https://www.instagram.com/developer/endpoints/tags/ [Accessed 3 Jan. 2016].