Sidi Mohamed Ben Abdallah University - Polydisciplinary Faculty of Taza Department of Mathematics, Physics and Computer Science
Master: Intelligent and mobile system « SIM »

Extraction of Twitter Data

Presented by:

Zouitni Chaimae & El aboui Hafida

Proposed by:

Pr. Ismail Akharraz

Academic year: 2019/2020

Master :SIM /S1

Plan

1. Introduction
2. Data Extraction Steps
3. Used Tools

Plan

Introduction

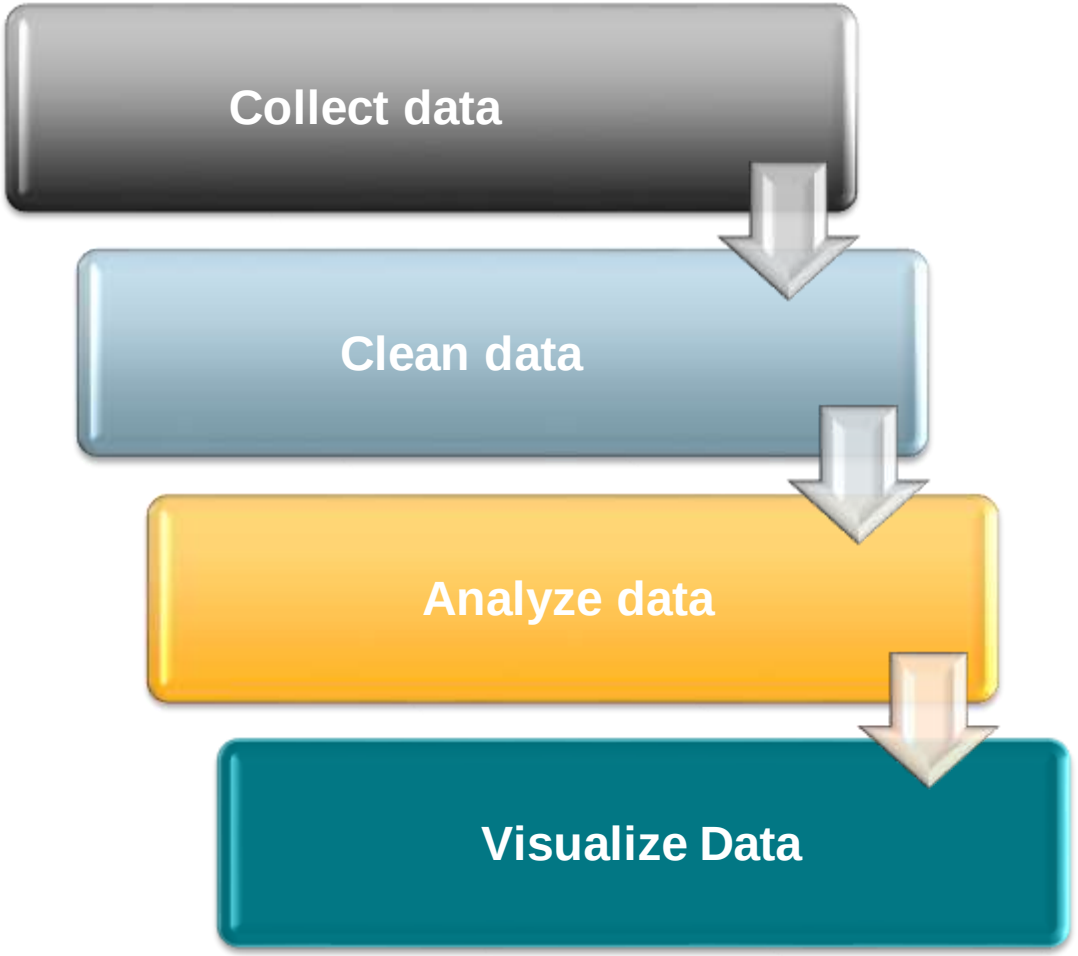
Definition

Twitter is a social network called "microblogging" which allows communication in the form of short messages not exceeding 140 characters called "tweets". They may contain shortened URLs, images, emoticons, animated gifs and videos.

Introduction

Objectif

The main objective of this project is to implement a data analysis based on machine learning to find the sentiment or polarity for a particular tweet.

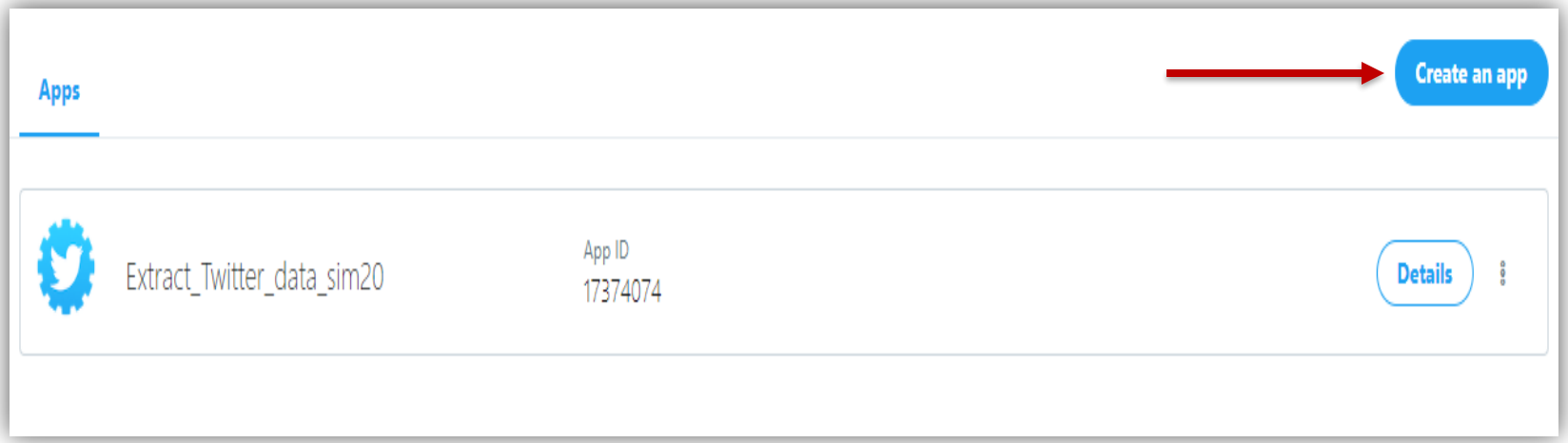


Data Extraction Steps

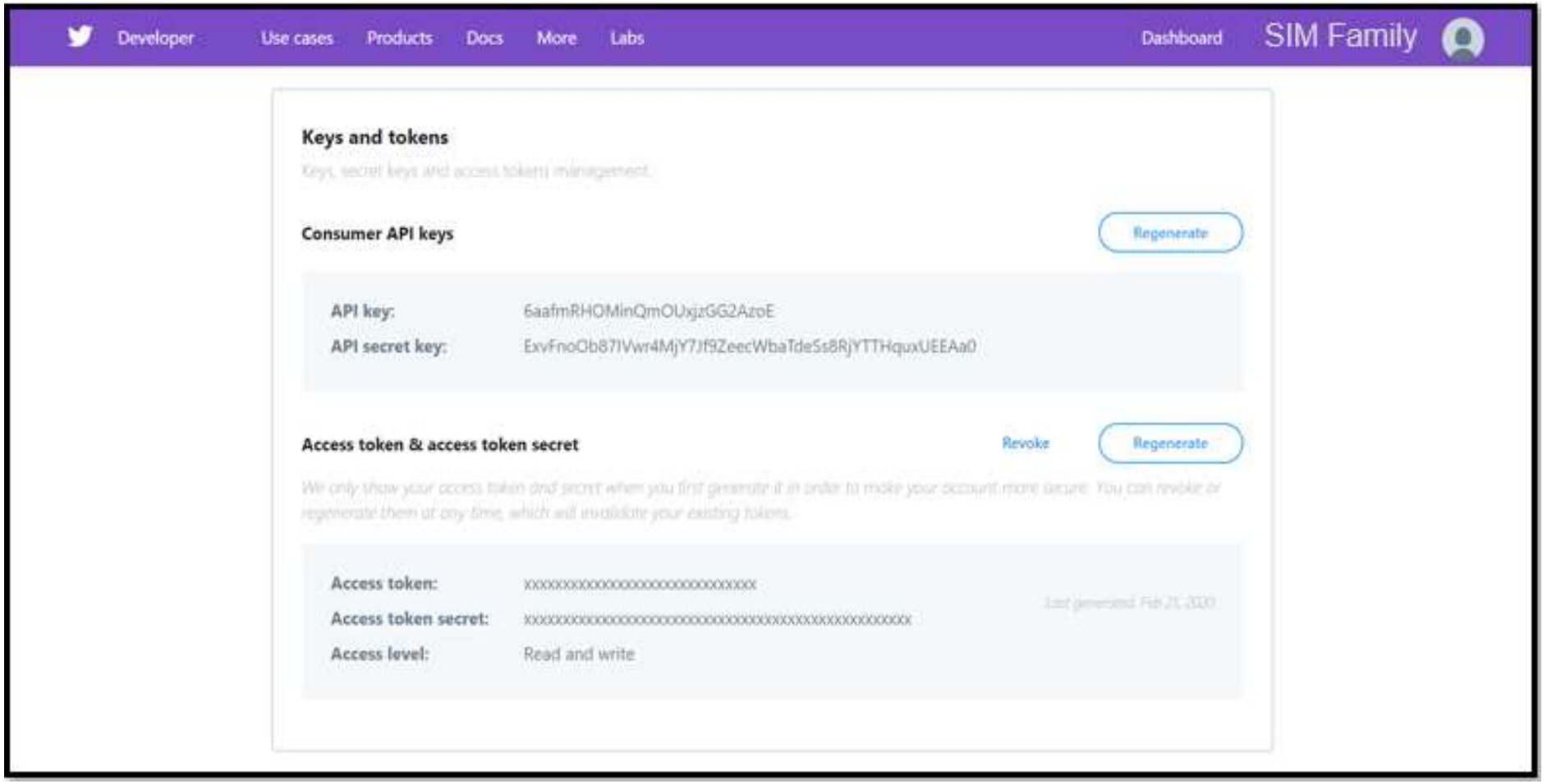
Step 1: Create Account

Create an application

- Click on create an app that interacts with the Twitter API.



Get Token ans Custmuer Keys



Data Extraction Steps

Step 2: Collect Data

1- Data Collection

Data collection is an approach guided by business scenarios which brings several advantages, but presents many challenges:

- Big data is quite fragmented and increasingly heterogeneous, by nature , its format its dispersion over within several internal or external systems.

- Understand the analysis and data capture techniques to be able to choose those that are best suited to the business scenarios to be treated.

- Collection requires the opinion of the expert who has the functional and technical knowledge to map the appropriate data

Data collection

1- Data Collection

```
consumer_key= 'yourkeyhere'
consumer_secret= 'yourkeyhere'
access_token= 'yourkeyhere'
access_token_secret= 'yourkeyhere'
```

```
auth = tw.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tw.API(auth, wait_on_rate_limit=True)
```

Data collection

1- Data Collection

Parameters:

- **id** – Specifies the ID or screen name of the user.
- **user_id** – Specifies the ID of the user. Helpful for disambiguating when a valid user ID is also a valid screen name.
- **screen_name** – Specifies the screen name of the user. Helpful for disambiguating when a valid screen name is also a user ID.
- **since_id** – Returns only statuses with an ID greater than (that is, more recent than) the specified ID.
- **max_id** – Returns only statuses with an ID less than (that is, older than) or equal to the specified ID.
- **count** – Specifies the number of statuses to retrieve.
- **page** – Specifies the page of results to retrieve. Note: there are pagination limits.

Data Extraction Steps

Step 3: Data Cleaning

2- Data Cleaning

- Absence of measurement for a given instance.
- Data not entered or poorly understood.
- incomplete and inconsistent data.

```
def cleanTweet(self, tweet):
    # Remove Links, Special Characters etc from tweet
    return ' '.join(re.sub("(@[A-Za-z0-9]+)|([^0-9A-Za-z \t]) | (\w +:\ / \ / \S +)", " ", tweet) .split())
```

Data Extraction Steps

Step 4: Data Anaysis

3- Data Analysis

- Study the behavior
- Group according to criteria for targeted Emailing
- Identification of customer
- Calculation profitability
- Predict the evolution actions
- Allocation of automated loans, and support for credit decision

Data Extraction Steps

Step 5: Used Libraries

Install Libraries

Tweepy allows to use a Twitter API to stream tweets	**NLTK** allows to processes natural language.	**Pandas** providing fast, flexible and expressive data structures
Matplotlib Allows handling of Json files	**Json** Allows to create graphical interfaces	**PyQt** allows to create graphical interfaces

Used Tools

Used Tools

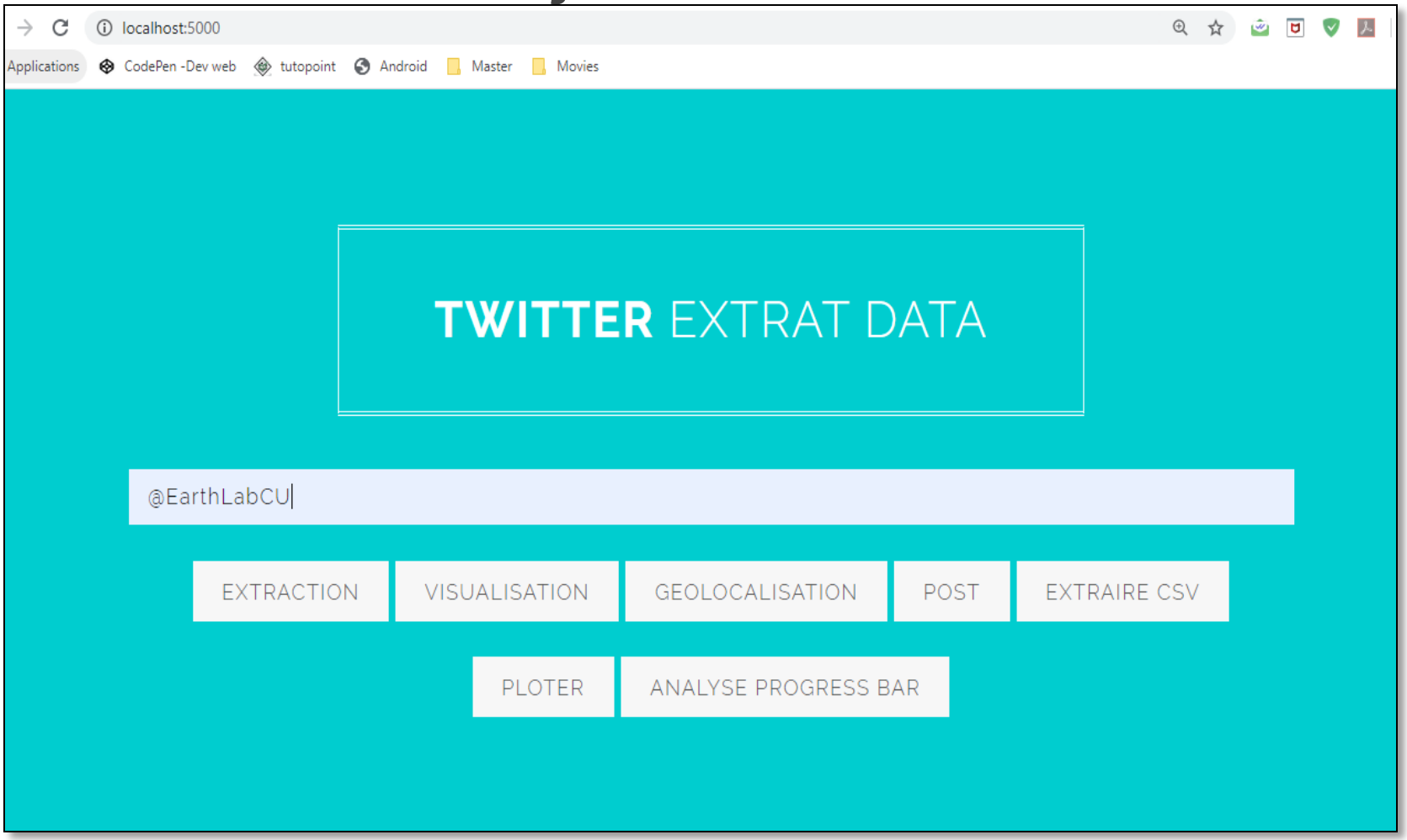
Web framework. provides (tools, libraries and technologies)

open source , development environment (IDE) for scientific programming in the Python language.

Display, layout and animated on a web page

Implementation

Project Interface



Conclusion

Conclusion

- Twitter's API is immensely useful in data mining applications, and can provide vast insights into the public opinion.

- Twitter's API can be leveraged in very complex big data problems, involving people, trends, and social graphs too complicated for the human mind to grasp alone.

Thank you for attention