MINI PROJECT REPORT

On

TWITTER HASHTAG ANALYSIS

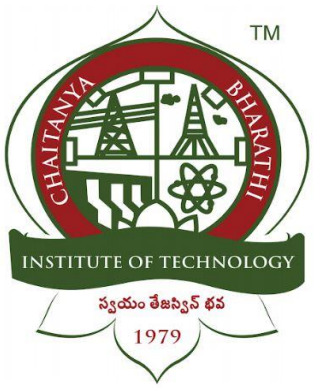
B.E (IT) – VI Sem

By

M YOGITHA NANDINI (160117737030)
P ARUN RAJ (160117737034)

Under the guidance of

KNVS Sridevi
Assistant Professor



DEPARTMENT OF INFORMATION TECHNOLOGY

CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY (A)

(Affiliated to Osmania University; Accredited by NBA(AICTE) and NAAC(UGC), ISO Certified 9001:2015)

GANDIPET, HYDERABAD – 500 075

Website: www.cbit.ac.in

2019-2020

CERTIFICATE

This is to certify that the project work entitled "**TWITTER HASHTAG ANALYSIS**" submitted to CHAITANYA BHARATHI INSTITUTE OF TECHNOLOGY,in partial fulfillment of the requirements for the award of the completion of VI semester of B.E in Information Technology, during the academic year 2019-2020, is a record of original work done by **M.Yogitha Nandini (160117737030),P.Arun Raj (160117737034** during the period of study in Department of IT, CBIT, HYDERABAD, under our guidance.

Project Guide
KNVS Sridevi
Asst. Professor, Dept. of IT,
CBIT, Hyderabad.

Head of the Department
Dr. SureshPabboju
Professor, Dept. of IT,
CBIT, Hyderabad.

ACKNOWLEGMENTS

We would like to express our heartfelt gratitude to KNVS Sri Devi, our project guide, for her invaluable guidance and constant support, along with her capable instruction and persistent encouragement.

We are grateful to our Head of Department, Dr.SureshPabboju, for his steady support and the provision of every resource required for the completion of this project.

We would like to take this opportunity to thank our Principal, Dr.P.Ravinder Reddy, as well as the management of the institute, for having designed an excellent learning atmosphere.

Our thanks are due to all members of the staff and our lab assistants for providing us with the help required to carry out the groundwork of this project.

ABSTRACT

Twitter Hashtag Analysis is used to know the tweets done by the particular username which is given as the input. This is easy way to know the tweets done by any user on twitter. Twitter Hashtag Analysis is to analyze the hashtag usage by the specified username.

The purpose of this project is to retrieve the tweets of the users and analyze the usage of hash tags in their tweets without logging into their individual account or without having any account on twitter.

The user will be allowed to search for the tweets by inputting the username. Based on user provided username, the system will search for username in twitter. The tweets with hash tag (#) will be extracted and displayed and the analysis on the hashtags is done and barplot is drawn.

Project is developed using the software Python and jupyter notebook tool. Project consists of modules named as tweepy, re and matplotlab.pyplot.

when we give the input of the particular username we get the tweets done by the particular username and the hashtags used in the retrieved tweets and with their count is displayed along with the barplot of the hashtag which are extracted from the retrieved tweets.

CONTENTS

LIST OF FIGURES:

1. INTRODUCTION

1.1 MOTIVATION

Twitter Hashtag Analysis is used to know the tweets done by the particular username which is given as the input. This project can be used by any individual.

Twitter Hashtag Analysis is to analyze the hashtag usage by the specified username. This is easy way to know the tweets done by any user on twitter.

The purpose of this project is to retrieve the tweets of the users and analyze the usage of hash tags in their tweets without logging into their individual account or without having any account on twitter.

1.2 PROBLEM STATEMENT

Twitter Hashtag Analysis is used to know the tweets done by the specified username which is given as the input. This project can be used by any individual.

When the username is given it should give the tweets done by the particular user and the no. of tweets is of user choice. This is easy way to know the tweets done by any user without logging into twitter account or without having any twitter account.

Twitter Hashtag Analysis is to analyze the hashtag usage by the specified username with the variation of the tweets.

2. EXISTING SYSTEM

Facebook

This is easily the largest social networking site in the world and one of the most widely used. And, Facebook was perhaps the first that surpassed the landmark of 1 billion user accounts. Apart from the ability to network with friends and relatives, you can also access different Facebook apps to sell online and you can even market or promote your business, brand and products by using paid Facebook ads.

Recently Facebook has lost the trust of millions of its users by allowing 3rd parties to access over 87 million users' personal data. This is a massive breech of trust and has created a feeling of unrest amongst the social media platform's audience. So much so that there is now a #deletefacebook campaign where people are completely removing themselves from Facebook and using other networks instead. If you're concerned about what Facebook is doing with your data, then why not check out my guide on alternatives to Facebook, and see if there's a better place for you to interact with family and friends.

Instagram

Instagram was launched as a unique social networking platform that was completely based on sharing photos and videos. This photo sharing social networking app thus enables you to capture the best moments of your life, with your phone's camera or any other camera, and convert them into works of art.

This is possible because Instagram allows you to apply multiple filters to your photos and you can easily post them to other popular social networking sites, such as Facebook and Twitter. It is now part of the Facebook empire.

3.PROPOSED SYSTEM

3.1 METHODOLOGY

Twitter is one of the most widely used social networks. For many organizations and people, having a great Twitter presence is a key factor to keeping their audience engaged. Part of having a great Twitter presence involves keeping your account active with new tweets and retweets, following interesting accounts, and quickly replying to your followers' messages.

Tweepy Module:

Tweepy is an open source Python package that gives you a very convenient way to access the Twitter API with Python. Tweepy includes a set of classes and methods that represent Twitter's models and API endpoints, and it transparently handles various implementation details, such as: Data encoding and decoding, HTTP requests, Results pagination, OAuth authentication, Rate limits, Streams.

If you weren't using Tweepy, then you would have to deal with low-level details having to do with HTTP requests, data serialization, authentication, and rate limits. This could be time consuming and prone to error. Tweepy, can focus on the functionality you want to build. Almost all the functionality provided by Twitter API can be used through Tweepy. The only current limitation, as of version 3.7.0, is that Direct Messages don't work properly due to some recent changes in the Twitter API.

Tweepy is a Python library for accessing the Twitter API. It is great for simple automation and creating twitter bots. Tweepy has many features. Tweepy is one of the library that should be installed using pip. Now in order to authorize our app to access Twitter on our behalf, we need to use the OAuth Interface.

Tweepy provides the convenient Cursor interface to iterate through different types of objects. Tweepy tries to make OAuth 1a as painless as possible for you. To begin the process we need to register our client application with Twitter. Create a new application and once it's done then we should have our consumer key and secret key.

Matplotlib.pyplot Module:

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK+.

Matplotlib.pyplot is a collection of command style functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g., creates a figure, creates a plotting area in a figure, plots some lines in a plotting area, decorates the plot with labels, etc.

In matplotlib.pyplot various states are preserved across function calls, so that it keeps track of things like the current figure and plotting area, and the plotting functions are directed to the current axes (please note that "axes" here and in most places in the documentation refers 8 to the *axes* part of a figure and not the strict mathematical term for more than one axis).

Re Module:

In Python, a regular expression is denoted as RE (REs, regexes or regex pattern) are imported through re module. Python supports regular expression through libraries. In Python regular expression supports various things like Modifiers, Identifiers, and White space characters.

In Python, a regular expression is denoted as RE (REs, regexes or regex pattern) are imported through re module. Python supports regular expression through libraries. In Python regular expression supports various things like Modifiers, Identifiers, and White space characters.

A regular expression in a programming language is a special text string used for describing a search pattern. It is extremely useful for extracting information from text such as code, files, log, spreadsheets or even documents.

While using the regular expression the first thing is to recognize is that everything is essentially a character, and we are writing patterns to match a specific sequence of characters also referred as string. Ascii or latin letters are those that are on your keyboards and Unicode is used to match the foreign text. It includes digits and punctuation and all special characters like $#@!%, etc.

3.2 ARCHITECTURE OF PROPOSED SYSTEM

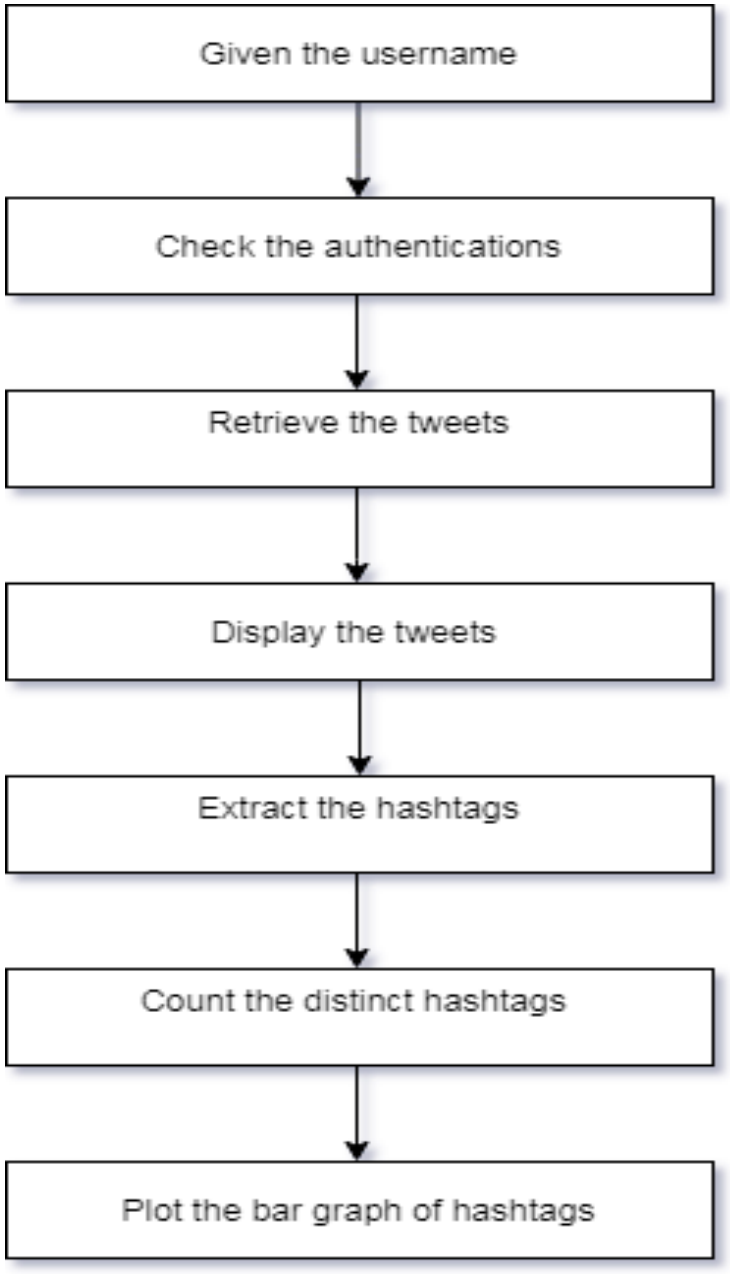


Fig 3.2.1 Flow chart of the twitter hashtag analysis

WORKFLOW

At the first give the username and then it checks the authentications. Tweets are retrieved from twitter using tweepy module with the help of API. Display the tweets of the user. Extract the hashtags using regular expressions. Counts the distinct hashtags and stores in dictionary. Plot the bar graph of hashtags to know the analysis of hashtags from the retrieved tweets.

4. SOFTWARE AND HARDWARE REQUIREMENTS

The requirements specification is a technical specification of requirements for the software products. It is the first step in the requirements analysis process it lists the requirements of a particular software system including functional, performance and security requirements. The requirements also provide usage scenarios from a user, an operational and an administrative perspective.

The purpose of software requirements specification is to provide a detail overview of the software project, its and goals. This describes the project target audience and its user interface, hardware and software requirements. It defines how the client, team and audience see the project and its functionality.

4.1 PYTHON

Introduction

Python is an interpreted, high-level, general-purpose programming language. Created by Guido van Rossum and first released in 1991, Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects.

Python is dynamically typed and garbage-collected. It supports multiple programming paradigms, including procedural, object-oriented, and functional programming. Python is often described as a "batteries included" language due to its comprehensive standard library.

Python was conceived in the late 1980s as a successor to the ABC language. Python 2.0, released 2000, introduced features like list comprehensions and a garbage collection system capable of collecting reference cycles. Python 3.0, released 2008,

was a major revision of the language that is not completely backward-compatible, and much Python 2 code does not run unmodified on Python 3. Due to concern about the amount of code written for Python 2, support for Python 2.7 was extended to 2020. Language developer Guido van Rossum shouldered sole responsibility for the project until July 2018 but now shares his leadership as a member of a five-person steering council.

Python is a fully-functional programming language that can do *anything* almost any other language can do, at comparable speeds.

Python is capable of threading and GPU processing just like any other language. Most of the data processing modules are actually just Python wrappers around C/C++ code.

"Modules" are pre-written Python code that you "import" in your Python program. Since there are many tasks that people commonly do, we have modules that people have written that do these tasks for you, and they usually do them in the cleanest and most efficient method possible. Sometimes you will see people refer to "DRY." This stands for Don't Repeat Yourself, which often also translates into "Don't Repeat Someone Else."

The phrase "wrapper" means that someone has placed, like a wrapper, Python code over another language. So, when you have a Python wrapper around C++ code, what someone has done is written some Python code that interacts with the C++ language. This allows you to make use of various aspects of the language being wrapped, in this case C++, without actually needing to know or understand that language.

Thus, Python can be used to make games, do data analysis, control robot and hardware, create GUIs, or even to create websites.

"GUI" stands for Graphical User Interface, and is used to describe a program that incorporates graphics to make the program more interactive for the user.

Features

Python provides lots of features that are listed below.

1) Easy to Learn and Use

Python is easy to learn and use. It is developer-friendly and high level programming language.

2)Expressive Language

Python language is more expressive means that it is more understandable and readable.

3)Interpreted Language

Python is an interpreted language i.e. interpreter executes the code line by line at a time. This makes debugging easy and thus suitable for beginners.

4)Cross-platform Language

Python can run equally on different platforms such as Windows, Linux, Unix and Macintosh etc. So, we can say that Python is a portable language.

5)Free and Open Source

Python language is freely available at offical web address.The source-code is also available. Therefore it is open source.

6)Object-Oriented Language

Python supports object oriented language and concepts of classes and objects come into existence.

7)Extensible

It implies that other languages such as C/C++ can be used to compile the code and thus it can be used further in our python code.

8)Large Standard Library

Python has a large and broad library and prvides rich set of module and functions for rapid application development.

9)GUI Programming Support

Graphical user interfaces can be developed using Python.

10)Integrated

It can be easily integrated with languages like C, C++, JAVA etc.

4.2 JUPYTER NOTEBOOK

Introduction

Jupyter Notebook is an open source web application that you can use to create and share documents that contain live code, equations, visualizations, and text. Jupyter Notebook is maintained by the people at Project Jupyter.

Jupyter Notebooks are a spin-off project from the IPython project, which used to have an IPython Notebook project itself. The name, Jupyter, comes from the core supported programming languages that it supports: Julia, Python, and R. Jupyter ships with the IPython kernel, which allows you to write your programs in Python, but there are currently over 100 other kernels that you can also use.

It is very flexible tool to create readable analyses, because one can keep code, images, comments, formula and plots together

5. IMPLEMENTATION OF PROJECT

First of all you have to open the command prompt and then enter jupyter notebook. Then we get the url to open files in jupyter notebook tool and select the program. Run the program.

The modules used are

1. tweepy
2. re
3. matplotlab.pyplot

Functionalities of the modules are

tweepy module is used to get the tweets from the twitter through twitter api authentication using the keys and tokens which are consumer keys and consumer secret codes.

re module is used to extract the hashtag regular expression from the retrieved tweets of the particular username.

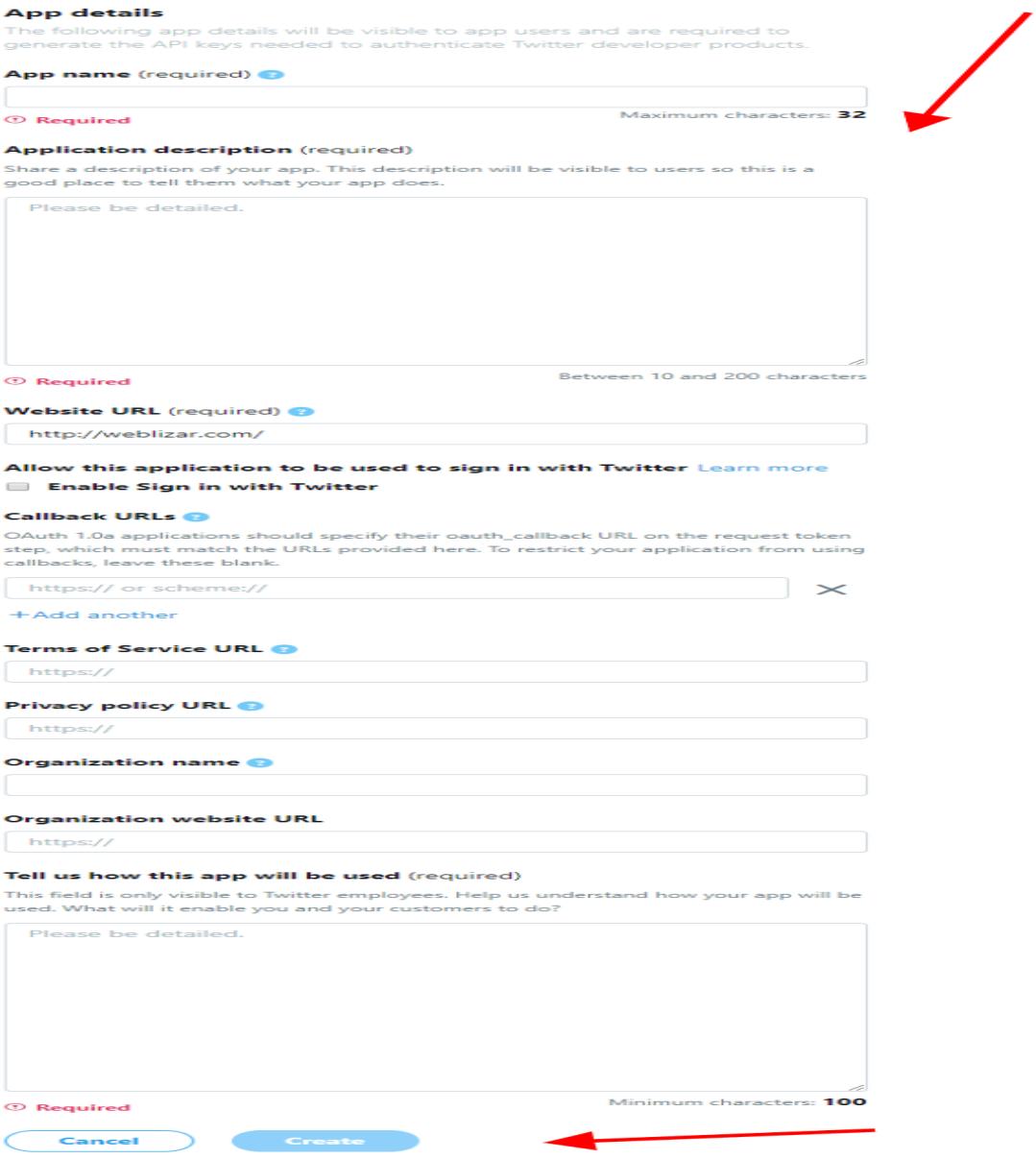
matplotlab.pyplot is used for plotting the barplot of the hashtags used from the retrieved tweets.

To extract tweets from twitter with python using jupyter notebook, you need to follow these basic steps:

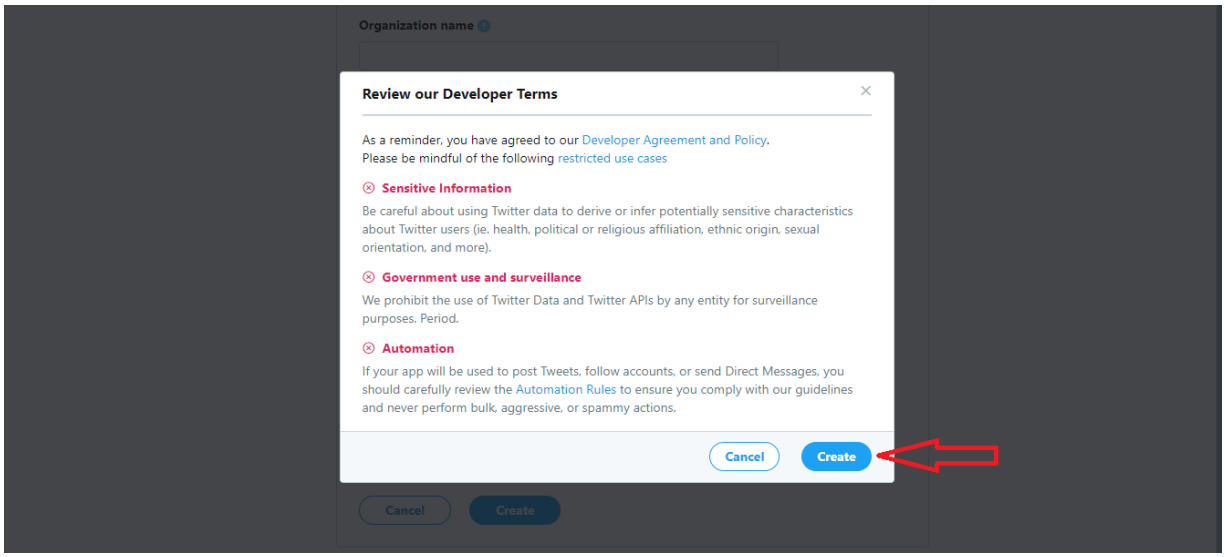
1. Given the username
2. Check the authentications
3. Retrieve the tweets
4. Display the tweets
5. Extract the hashtags
6. Count the distinct hashtags
7. Plot the bar graph of hashtags

Follow these steps to generate twitter API key

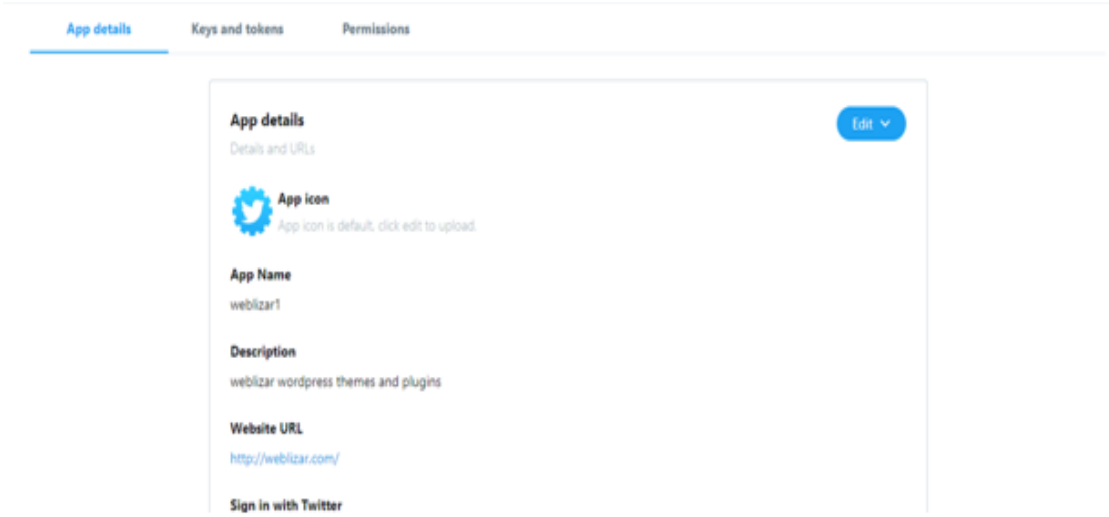
1. First you need to Login in Twitter and visit this link-> https://developer.twitter.com/en/apps
2. Enter the application name, description, website address, term & policy URL. After filling all the information, submit the form by clicking on the create button.



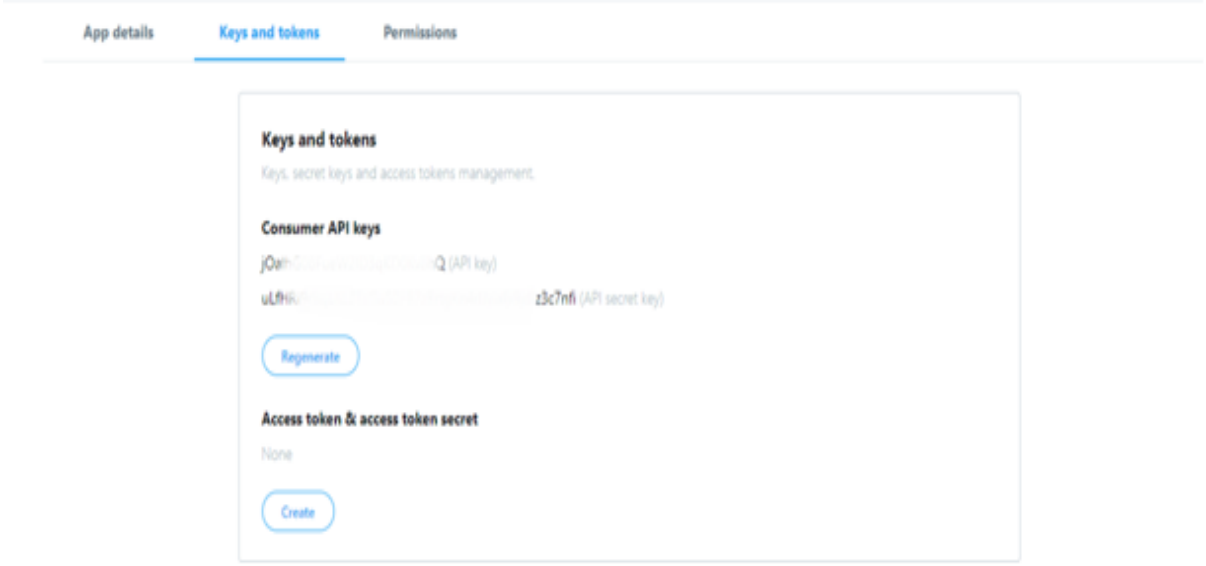
3. A pop up window will open, click on the create button to create the Twitter app.



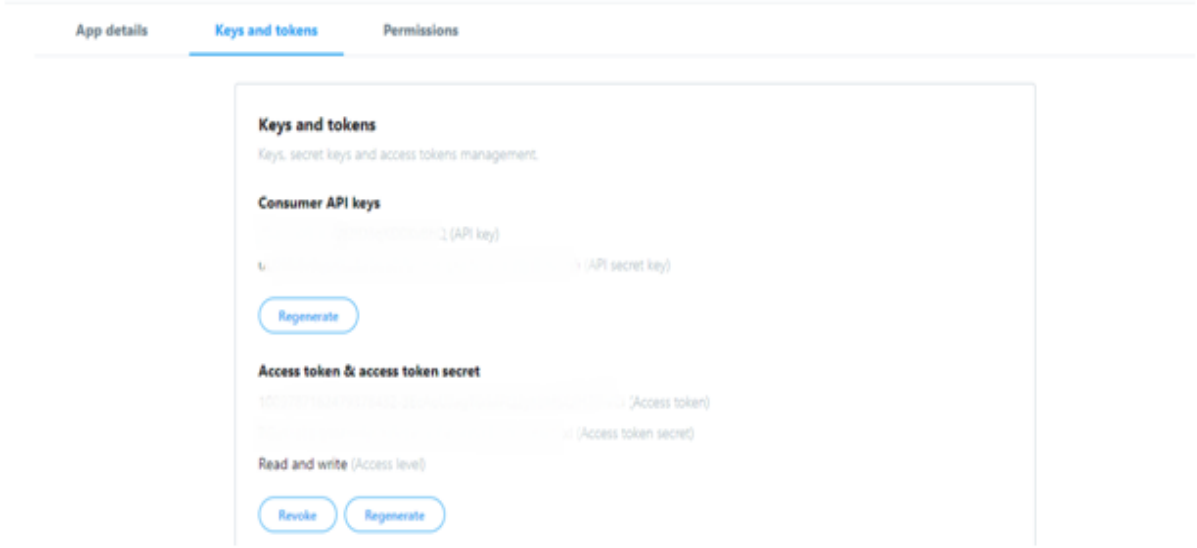
4. Two codes will be generated API key and API Secret key.



5. To generate the Access Token and Access Token Secret, click on the create button.



6. Now all four API key, API Secret key, Access Token, Access Token Secret, codes are generated.



Step 1:

After execution of the program the user must provide the input of the particular username for retrieving the tweets.

Step 2:

The function definition get_tweets is called by passing the username and then the tweepy helps to retrieve the tweets after checking the authentication of the twitter account and stores the tweets of the specific username given by the user and are displayed.

Step 3:

From the retrieved tweets the hashtags are extracted using re module with regular expressions. The distinct number of hashtags are counted using the function definition getDuplicatesWithCount.

Step 4:

The distinctly counted hashtags are stored in the dictionary and displayed. The dictionary which contain the hastags as the keys and count of the hashtags as values are plotted on the barplot.

Step 5:

The barplot contains x-axis as hashtags and y-axis as count of the hashtags and displayed.

5.1 RESULTS

From the Fig-5.1 when we give the input of the particular username we get the tweets done by the particular username and the hashtags used in the retrieved tweets and with their count which is resulted as in the output screen.

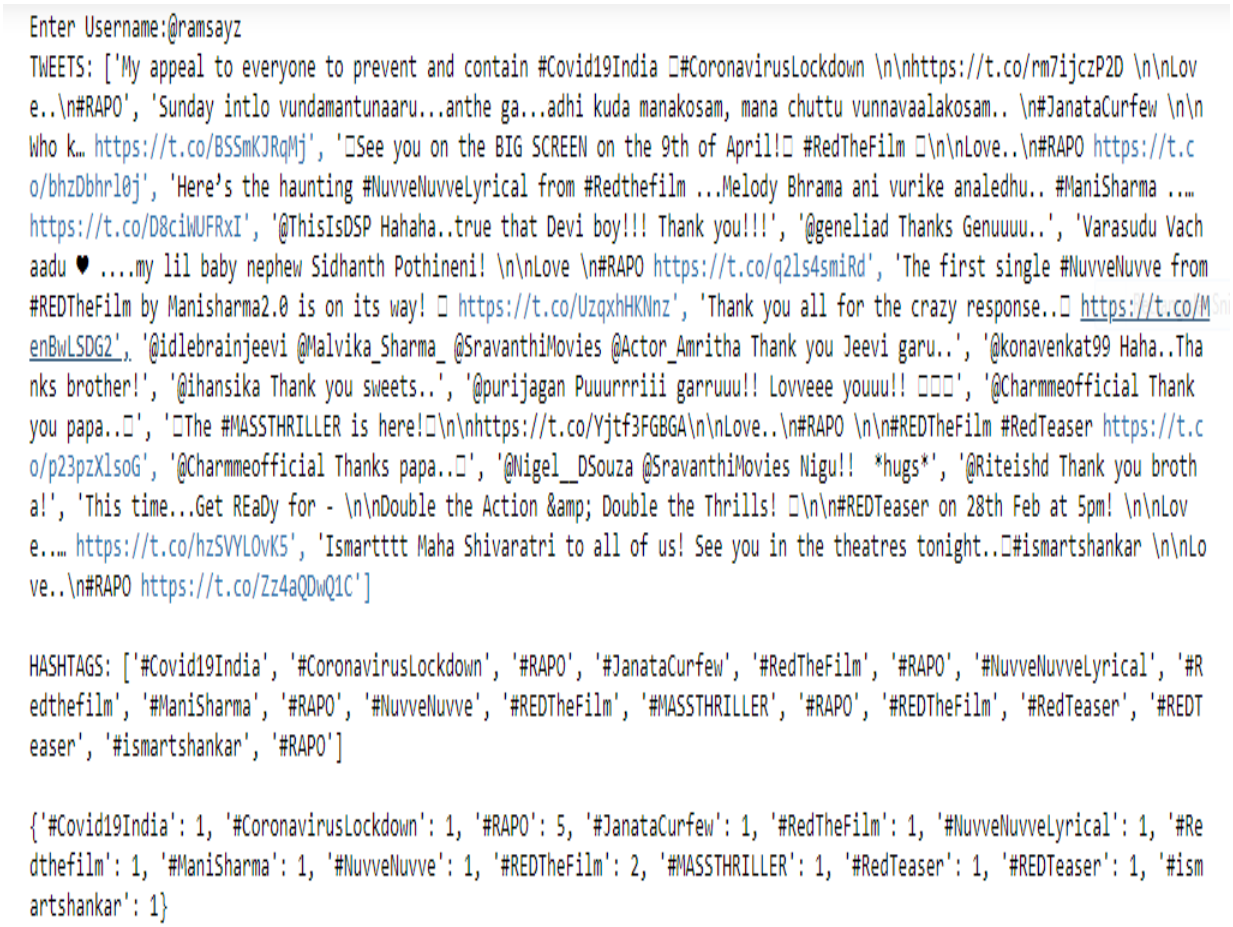


Fig 5.1.1 Displaying the tweets and extracted hashtags

From the Fig-5.2 Here is the barplot of the hashtag which are extracted from the retrieved tweets is resulted as in the output screen.

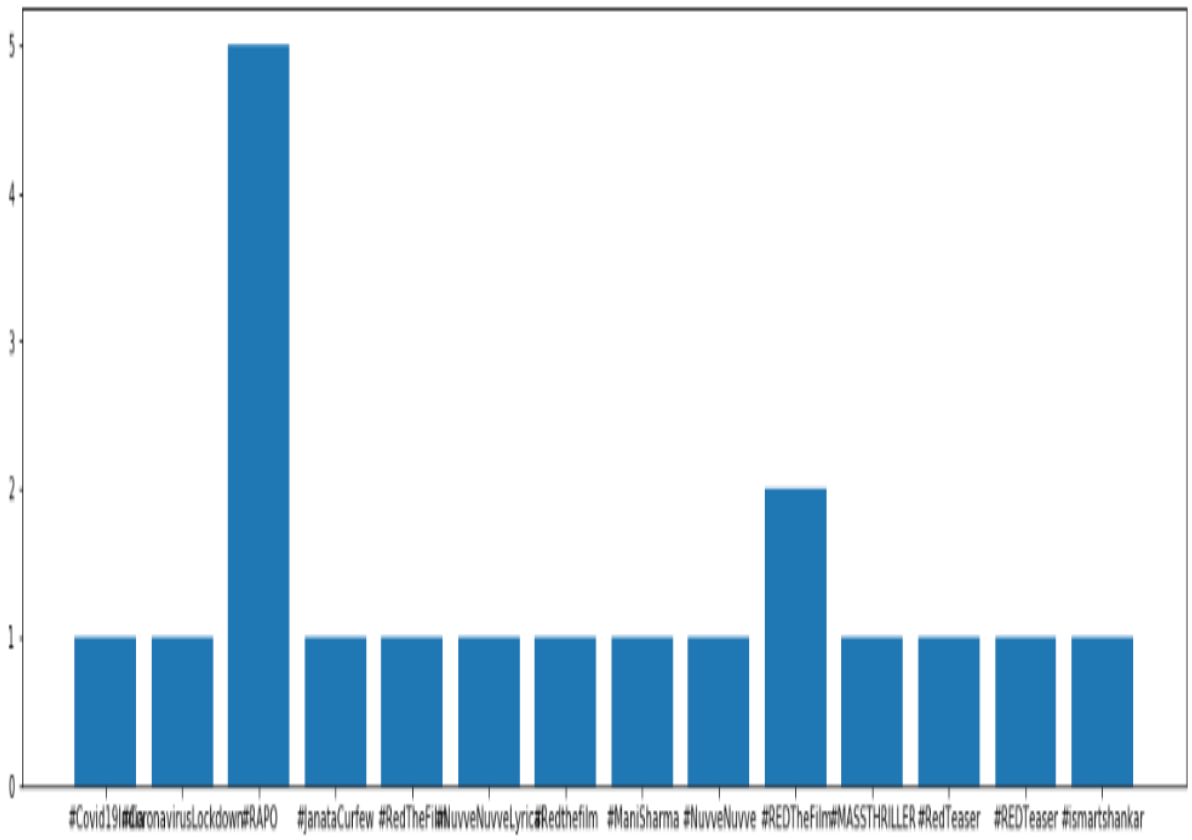


Fig 5.1.2 Displaying the barplot of hashtags

6. CONCLUSION & FUTURE SCOPE

Enormous amounts of messages get published each day on social media sites. For example, Twitter processes 230 million tweets (messages that are 140 characters long) a day (twitterstats). The explosion of textual messages can cause information overload. Our goal is to design systems that can analyze and summarize some social media content.

This project is designed such a way that an individual can easily know tweets of any particular username and analyze the hashtag usage without accessing the authorized website(twitter).

Making use of much more functionalities.

Display the results more effectively using a website.

BIBILIOGRAPHY

https://www.google.co.in/

https://docs.tweepy.org/en/latest/code_snippet.html

https://matplotlib.org/3.1.1/tutorials/introductory/pyplot.html

https://stackoverflow.com/

https://www.quora.com/

https://github.com/