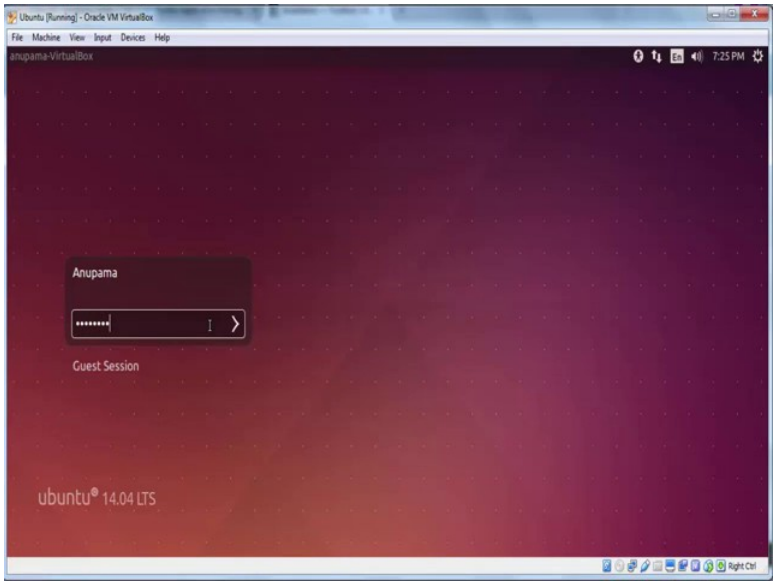
Privacy and Security in Online Social Networks
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Lecture – 12
Tutorial 3 Part 1 Twitter API

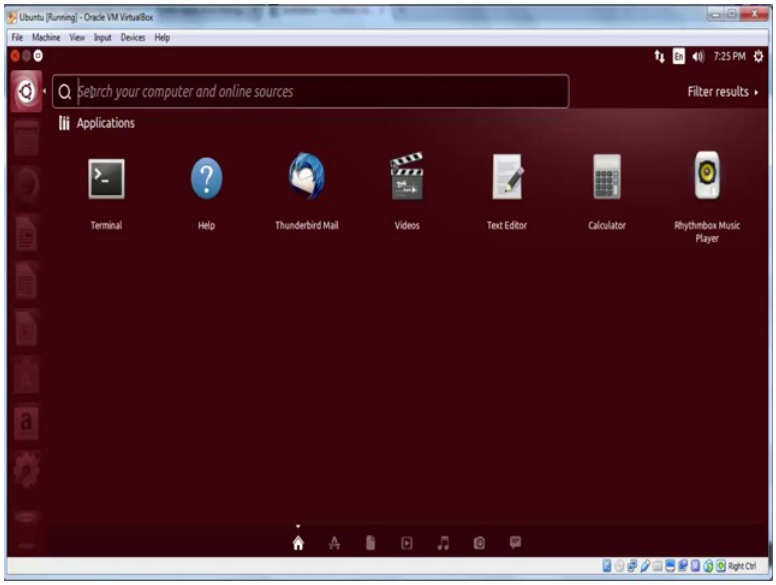
In this tutorial, we will learn how to collect data using twitter API.

(Refer Slide Time: 00:16)



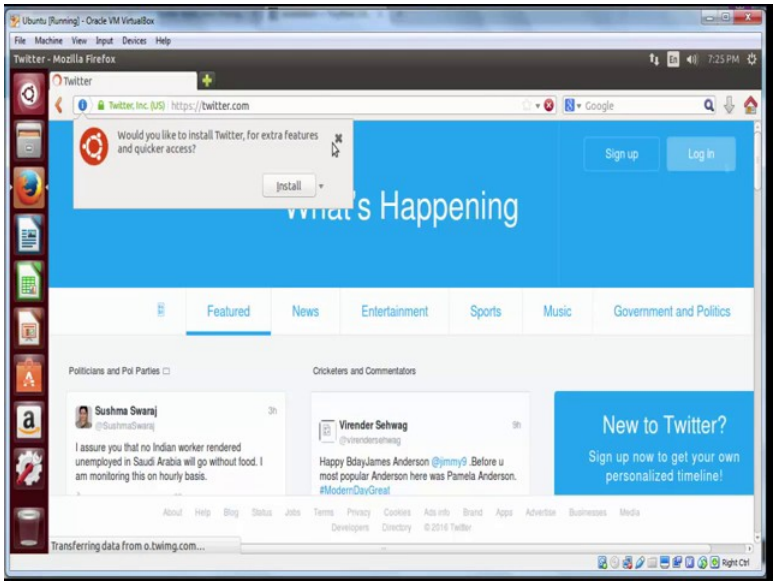
Let us get started and first create a twitter account.

(Refer Slide Time: 00:28)



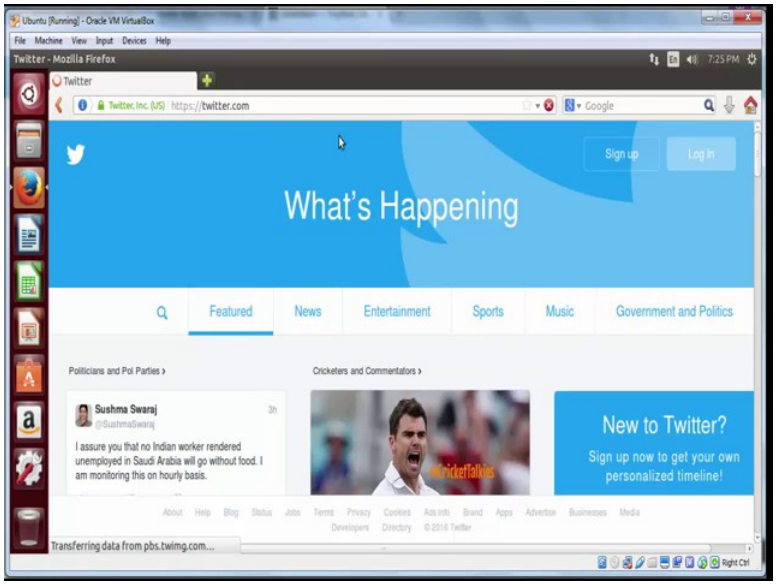
Now you may already have a twitter account, but please go through the entire video to understand how we successfully use the twitter API.

(Refer Slide Time: 00:33)



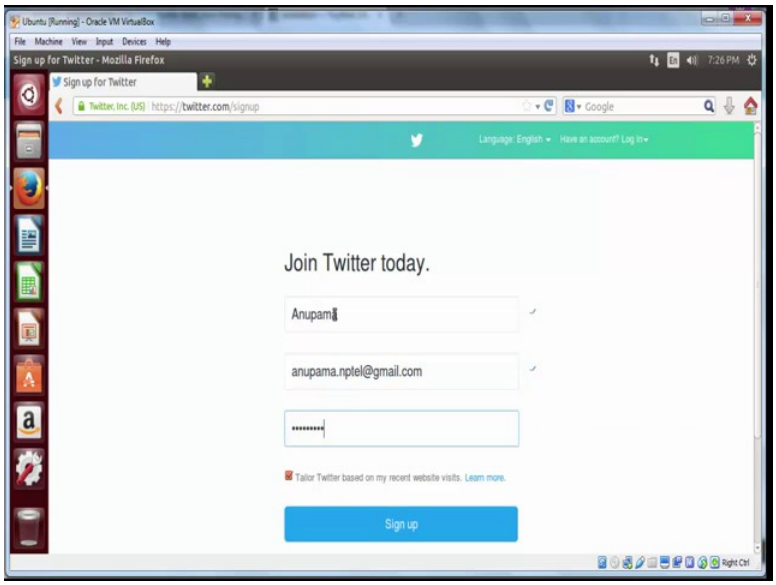
Open your browser and click on www.twitter.com.

(Refer Slide Time: 00:50)



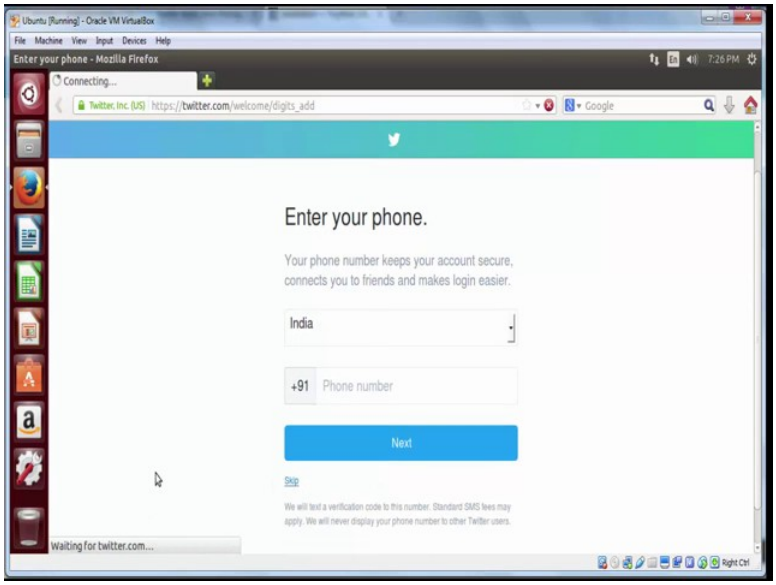
After you land up on the twitter's home page, on the top right click on sign up.

(Refer Slide Time: 00:56)



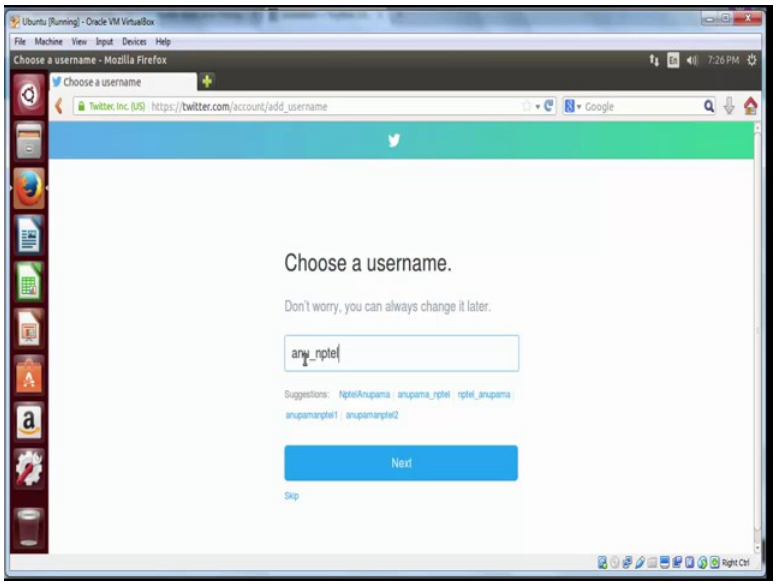
Creating an account is pretty straightforward. You only have to enter your full name, your email address and choose a password and then click on sign up.

(Refer Slide Time: 01:15)



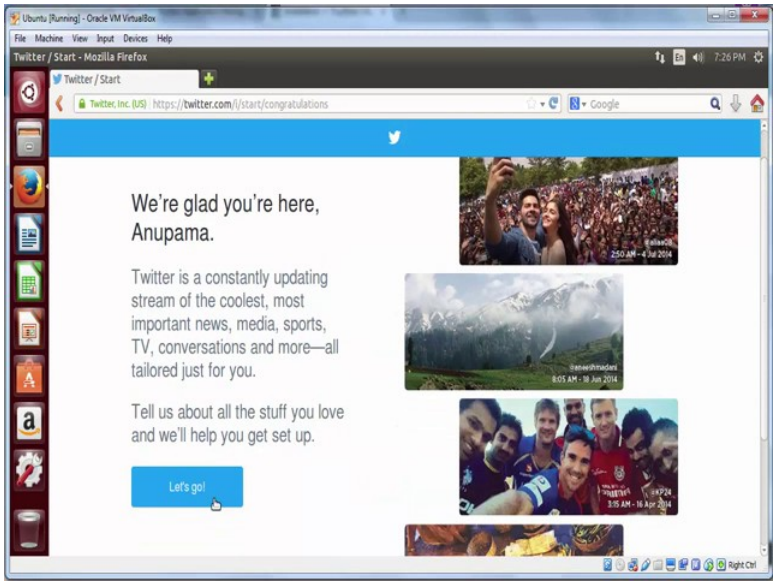
You may skip entering your phone number in the next step.

(Refer Slide Time: 01:19)



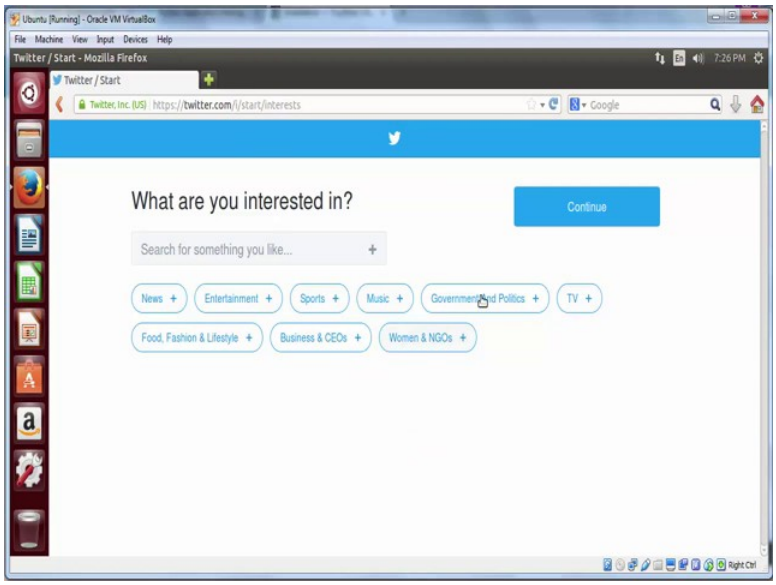
Choose a username, which will be a unique identity on twitter and then click next.

(Refer Slide Time: 01:32)

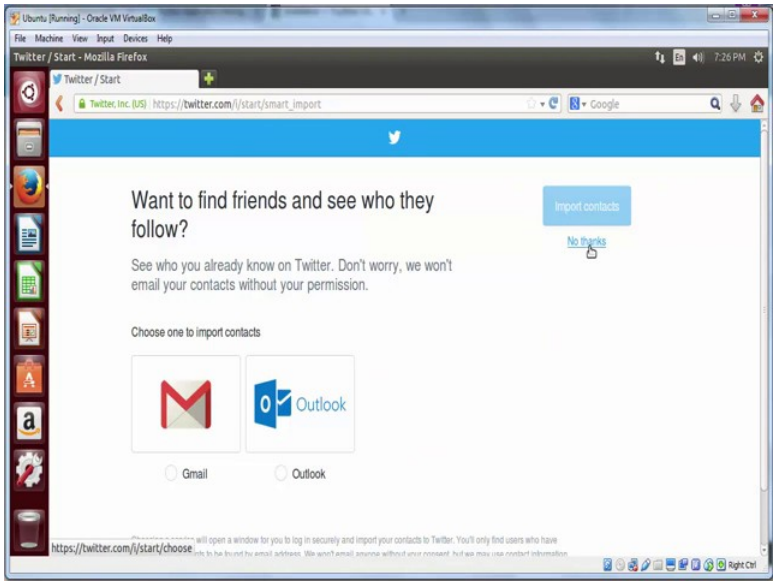


Now you can skip through rest of the steps to quickly create your twitter account.

(Refer Slide Time: 01:35)

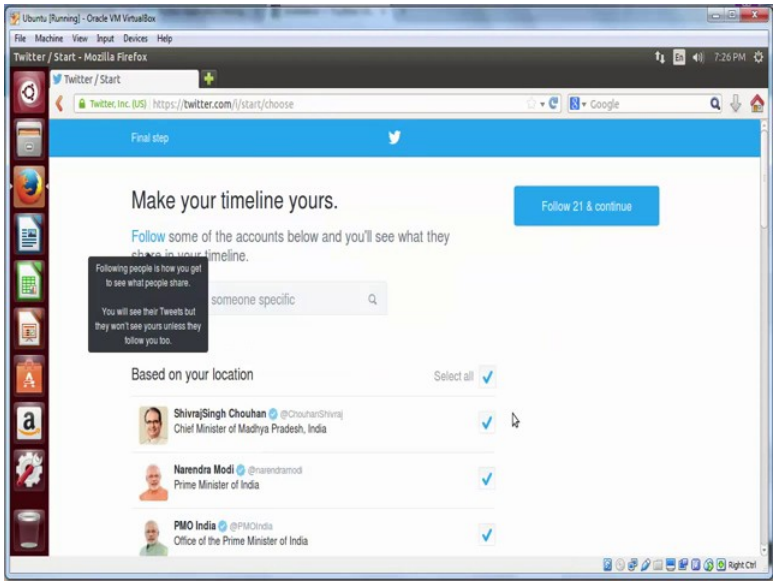


(Refer Slide Time: 01:38)



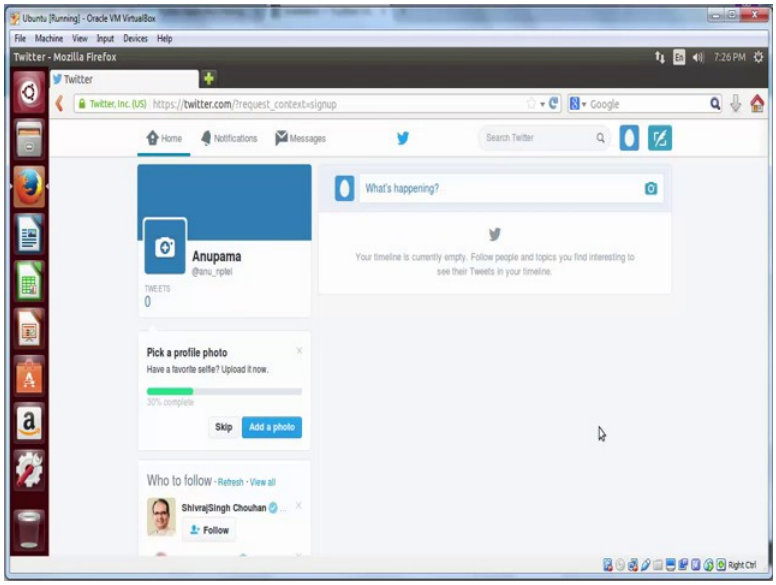
Do not export your contacts from your existing email address for now and click on no thanks.

(Refer Slide Time: 01:43)



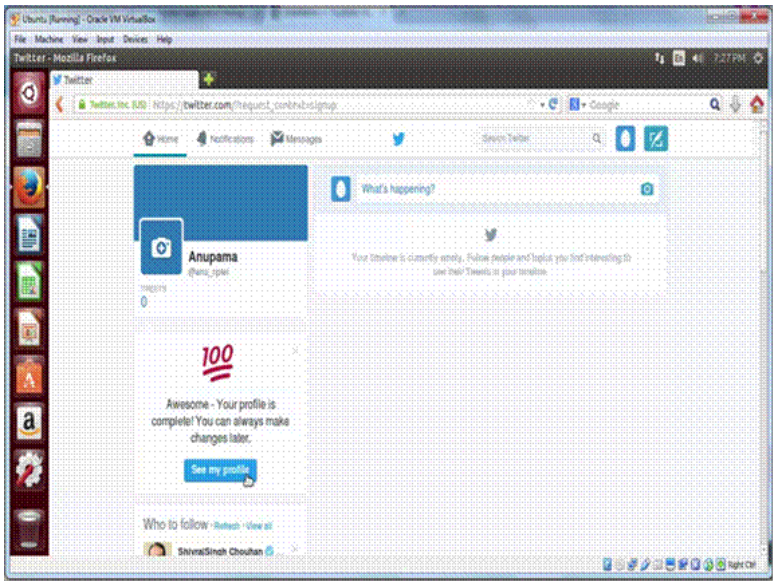
Do not follow any one and uncheck on select all.

(Refer Slide Time: 1:56)



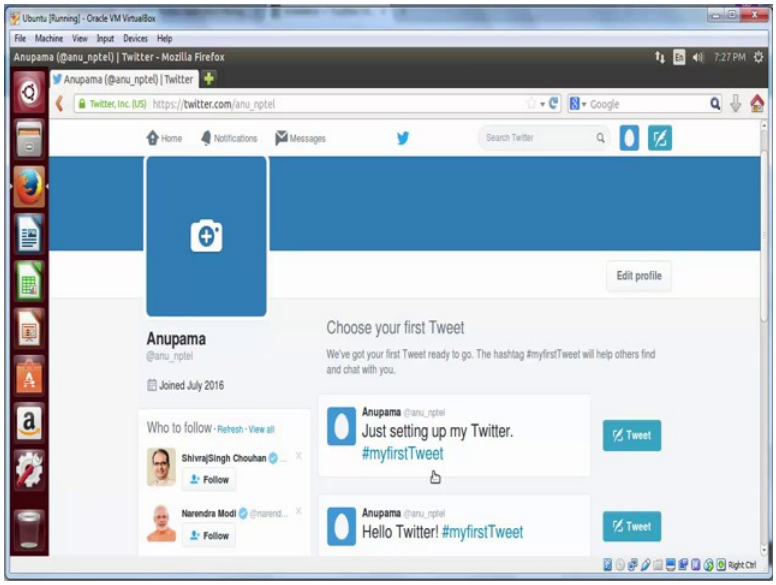
Then continue, skip through the rest of the steps on the left panel.

(Refer Slide Time: 02:06)



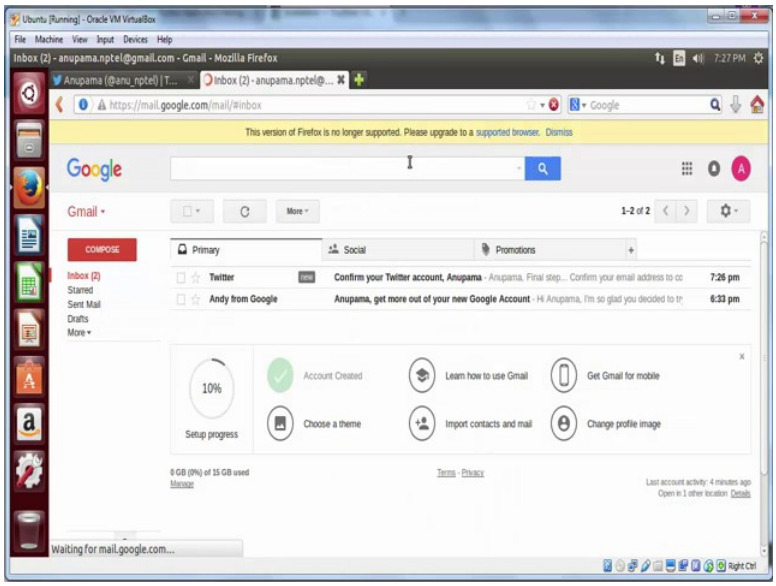
And in next 2 to 3 seconds your account will be ready.

(Refer Slide Time: 02:11)



Now let us try posting the tweet. You can either choose a tweet suggested by twitter or compose one yourself. In this case, we choose the one, which is suggested by twitter. Click on the tweet button and your tweet will be posted.

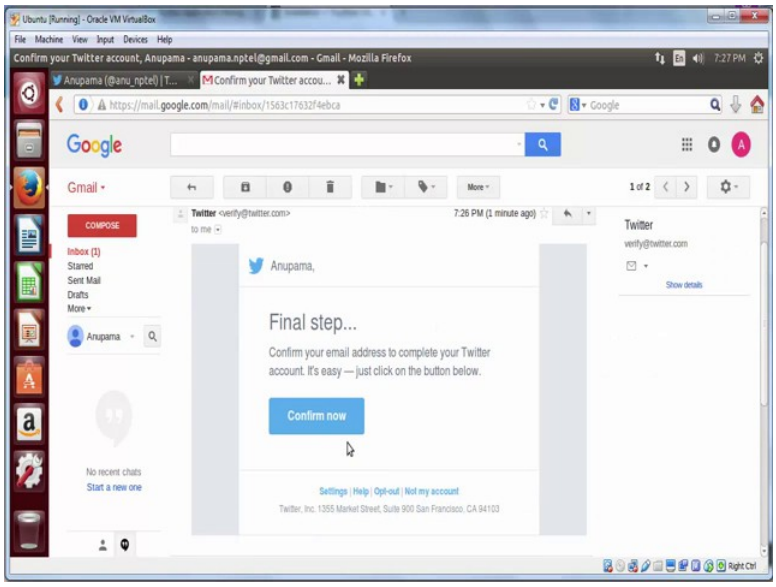
(Refer Slide Time: 02:42)



Before you proceed any further, it is important that you confirm your Twitter account by

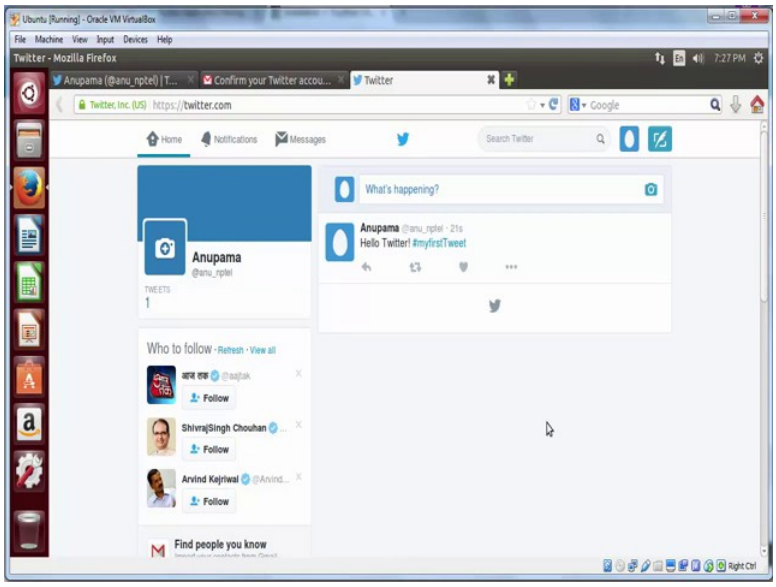
going to your email address using which you create created the twitter account. So, let us go to your email inbox to confirm the email, which has been sent by twitter dot com.

(Refer Slide Time: 02:48)



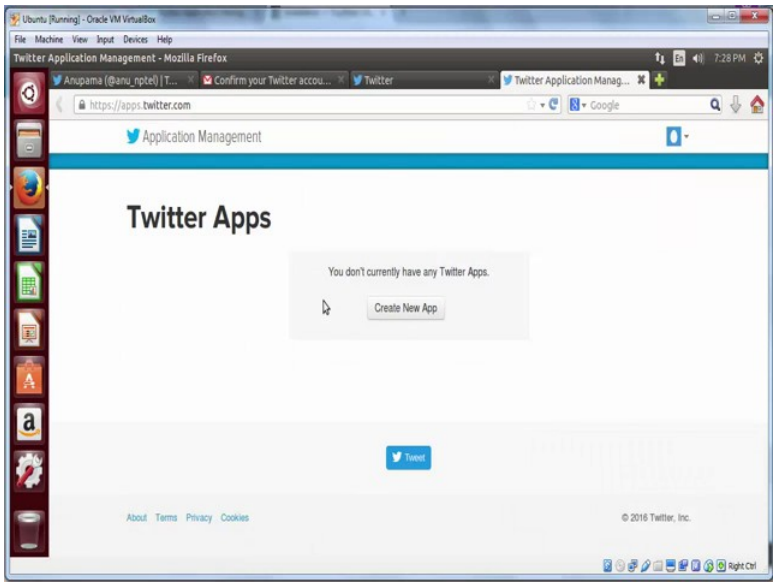
In your inbox, you will notice an email with confirm now button, click on it and you will be again redirected to your twitter's home page.

(Refer Slide Time: 03:03)



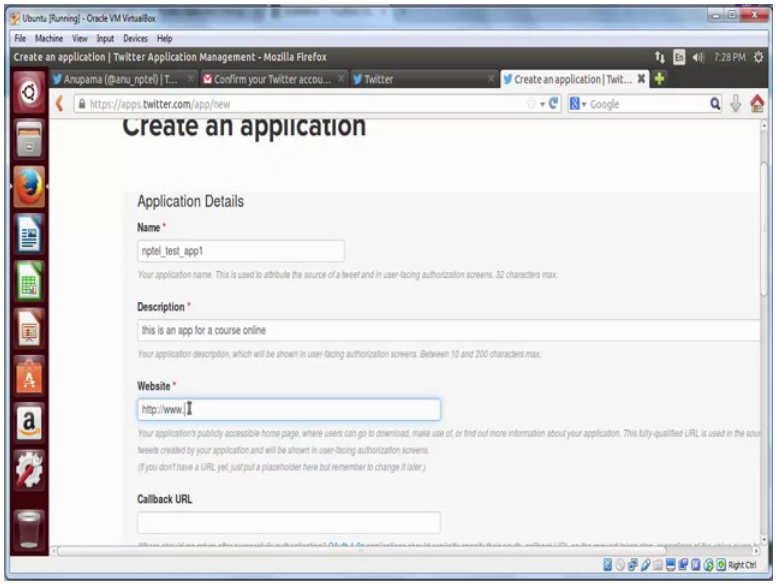
And there will be a message which says your account has been confirmed. Now you are good to go, and you will able to create a twitter API successfully.

(Refer Slide Time: 03:21)



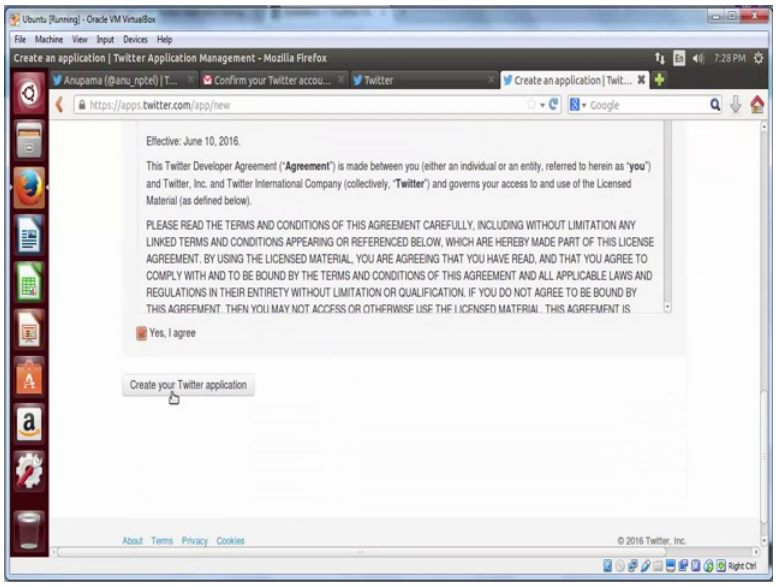
To create a twitter application go to apps.twitter.com and click on create new app.

(Refer Slide Time: 03:27)



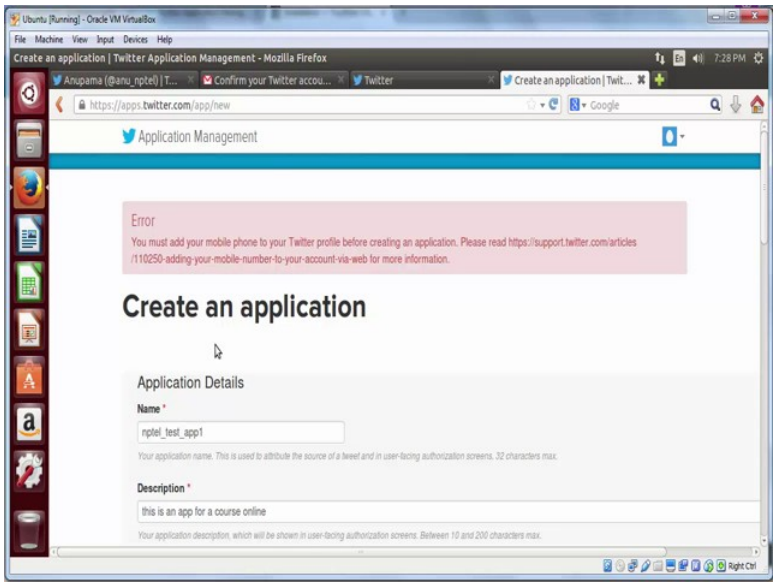
Enter a name for your application. And add a description. In the website add any URL with http in front. Here we use http www.google.com, and leave the callback URL blank.

(Refer Slide Time: 03:59)



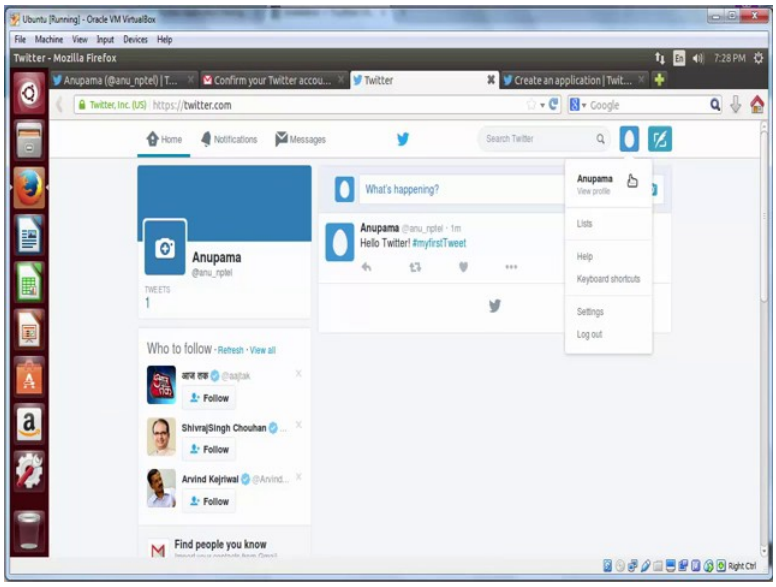
Now scroll down, and agree to the terms and condition; and then click on create your twitter application.

(Refer Slide Time: 04:06)



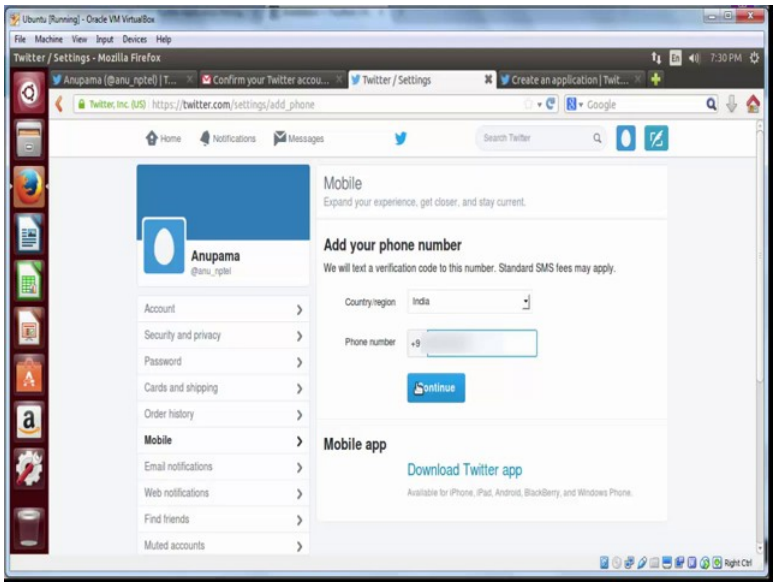
If you did not add your phone number to the twitter account, this step will throw an error. We need to add a valid phone number to a twitter account.

(Refer Slide Time: 04:21)



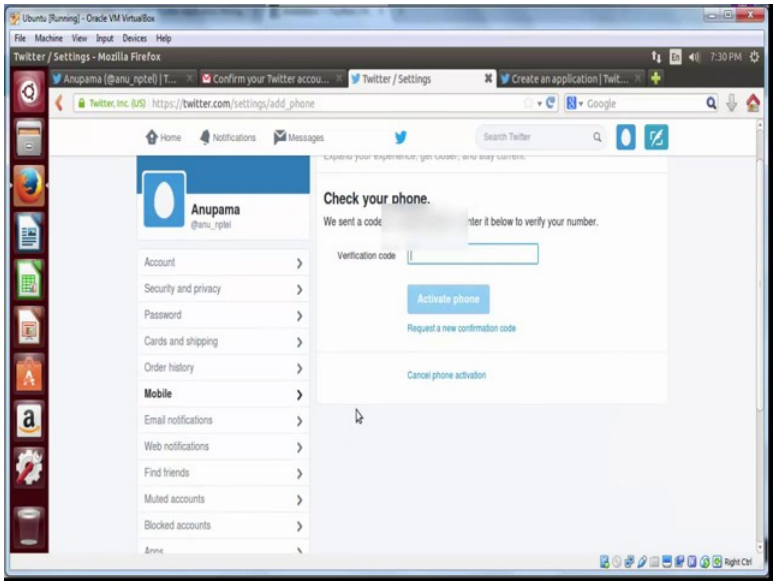
To do that go back to your twitter profile, on the top right click on your twitter profile picture or the egg icon and then select settings.

(Refer Slide Time: 04:28)



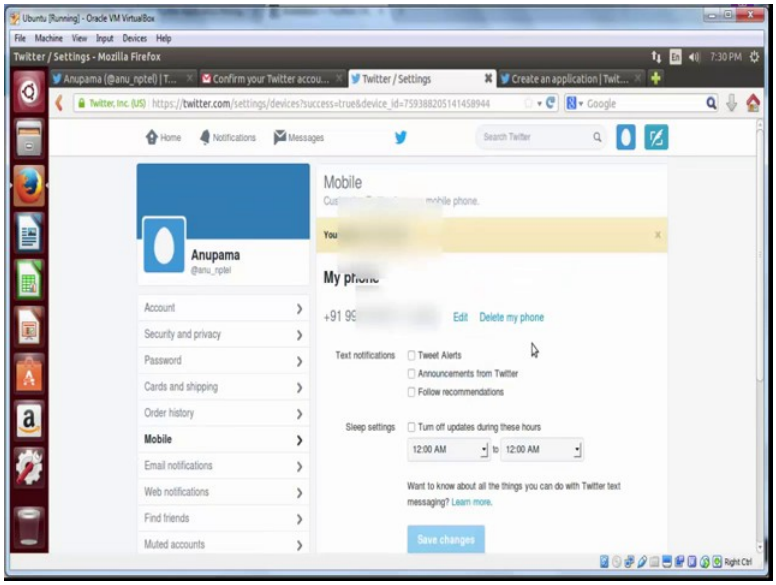
From the left panel, select mobile and enter your mobile number, click continue. And as soon as you do that, you will receive a verification code as a SMS on your registered mobile number.

(Refer Slide Time: 04:43)



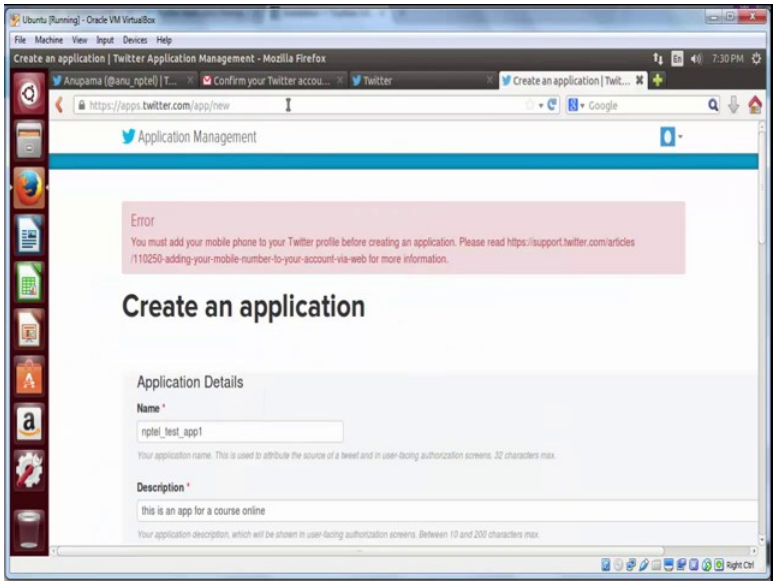
Enter the code which you receive which will be a 6 to 8 digit number, and click on activate phone.

(Refer Slide Time: 04:58)



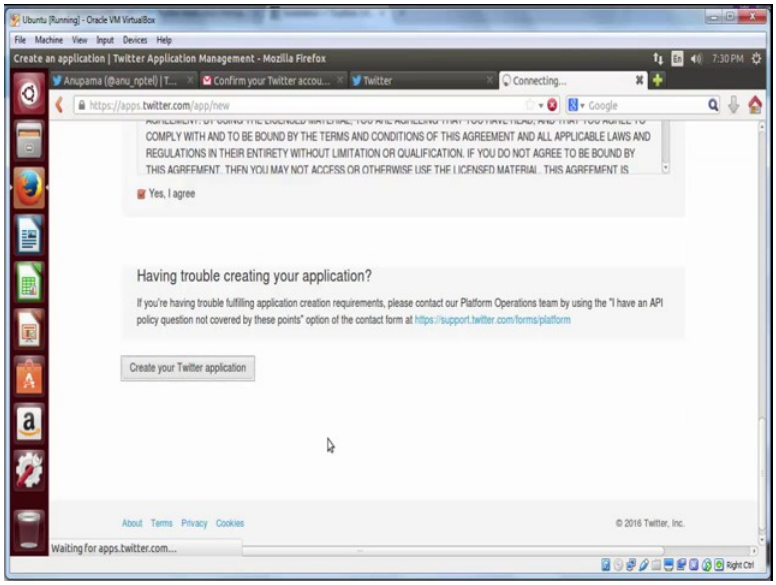
There will be a message which says your phone has been confirmed. Now let us again try creating the application.

(Refer Slide Time: 05:09)



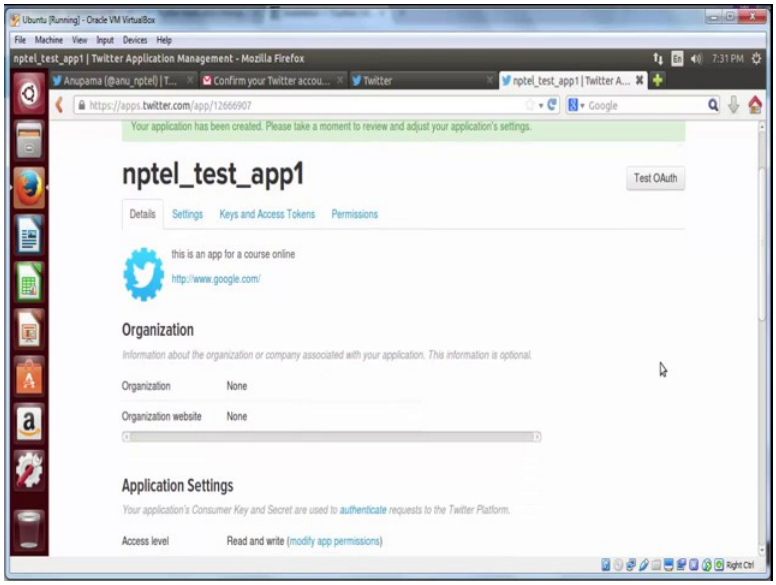
Go back to the apps.twitter.com tab.

(Refer Slide Time: 05:20)



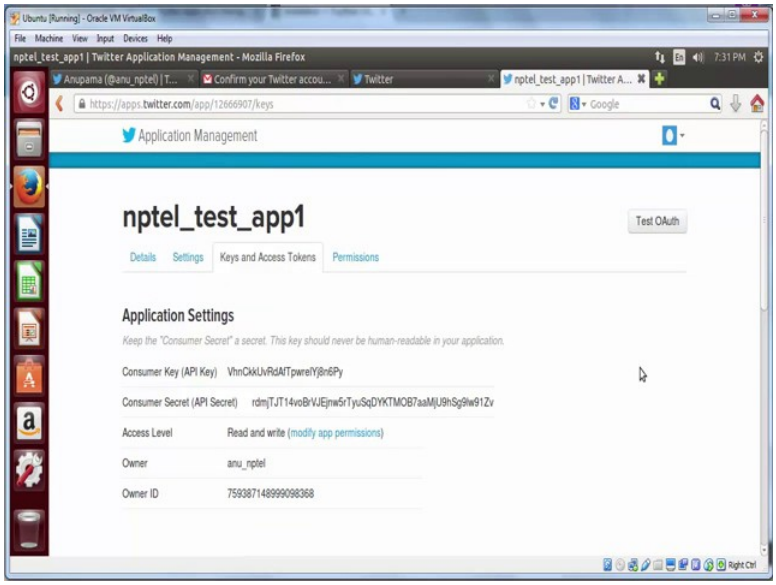
Scroll down and click on create your twitter application. This time it will successfully get created.

(Refer Slide Time: 05:24)



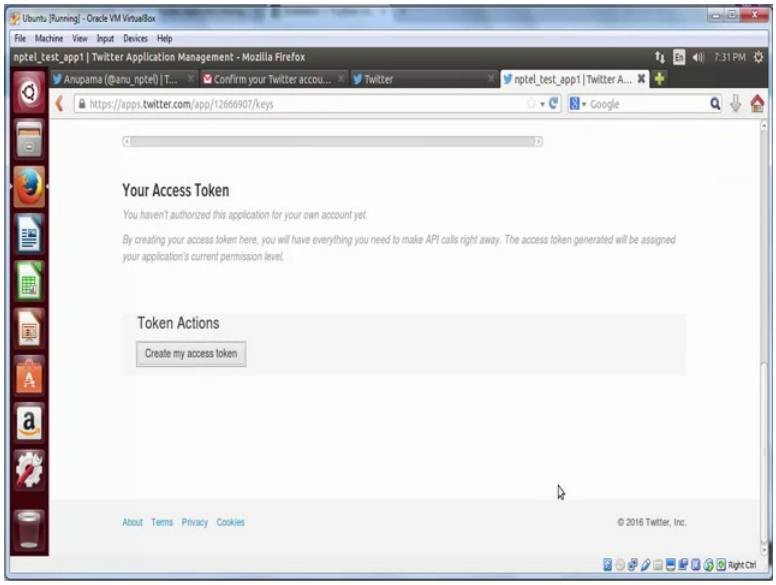
Now, you can get hold of the access keys and tokens to use this twitter application to gather data.

(Refer Slide Time: 05:35)



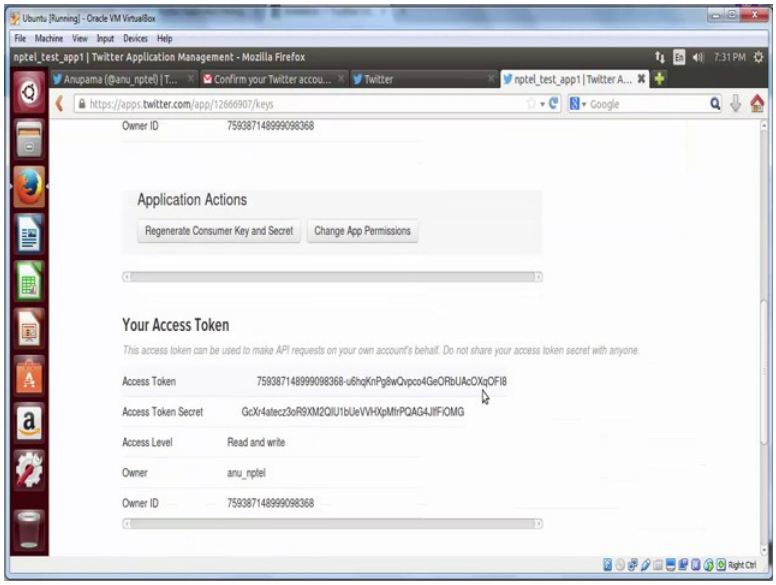
Click on the keys and access token tab. And you will be able to see your consumer key and your consumer secret.

(Refer Slide Time: 05:46)



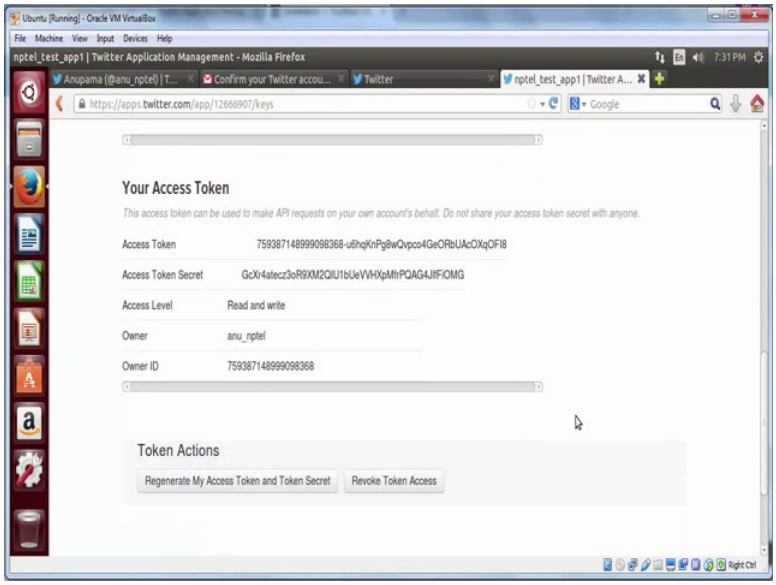
For a successful data collection, you need a set of two more keys which is called token actions. Click on create my access token.

(Refer Slide Time: 05:57)



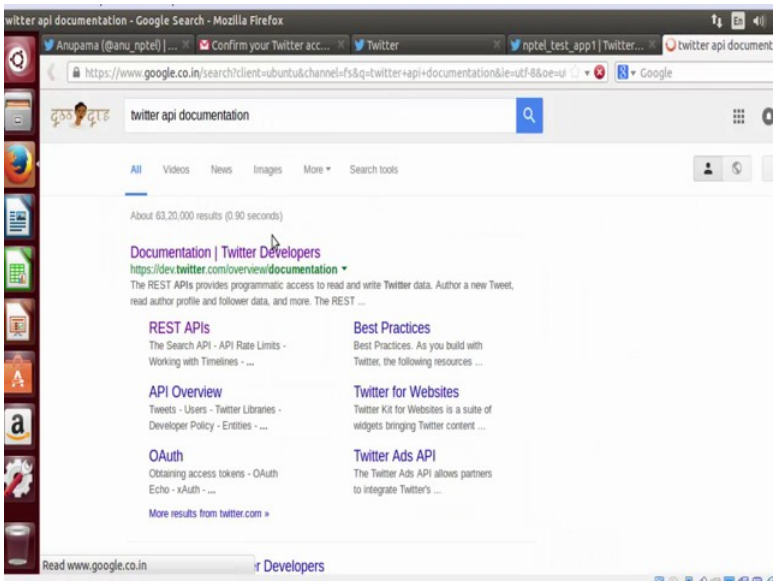
Which will result in generation of two more keys called access token and access token secret.

(Refer Slide Time: 06:03)



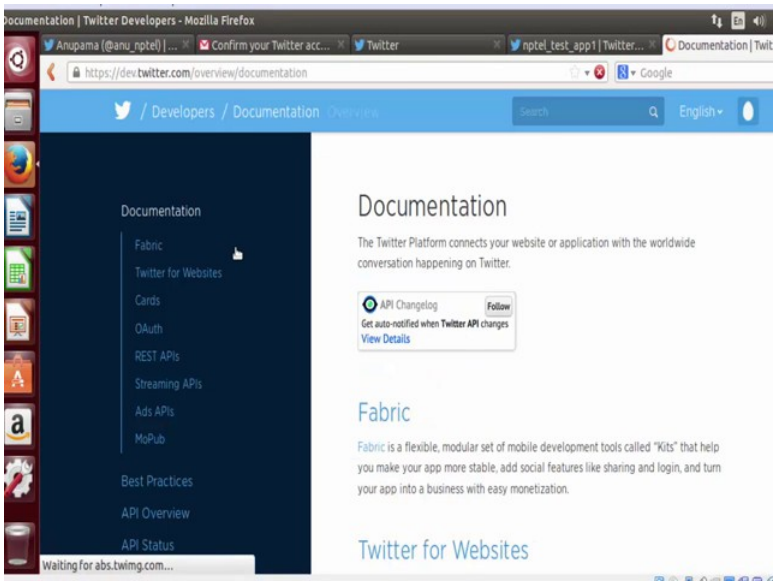
Now this set of four keys can be used to collect data, which we will see in the next step.

(Refer Slide Time: 06:23)

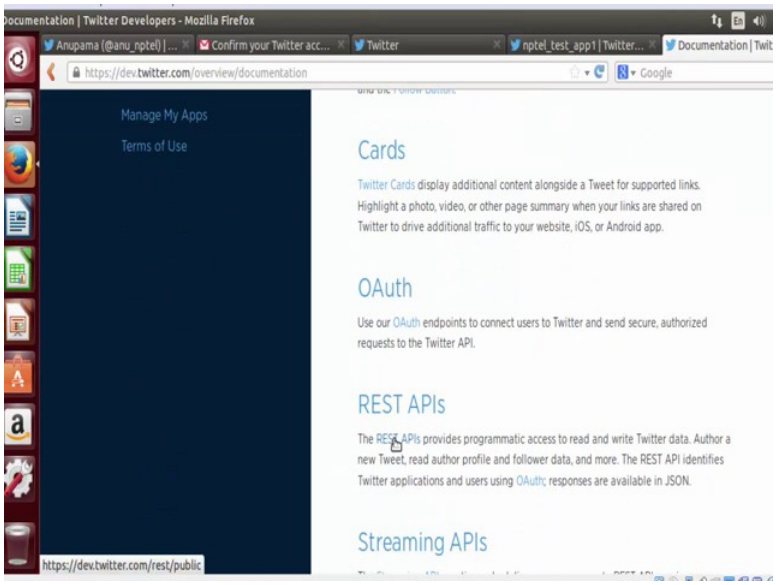


Let us first look at the various functionalities provided by twitter API. Go to twitter API documentation in a browser and navigate to the documentation.

(Refer Slide Time: 06:28)

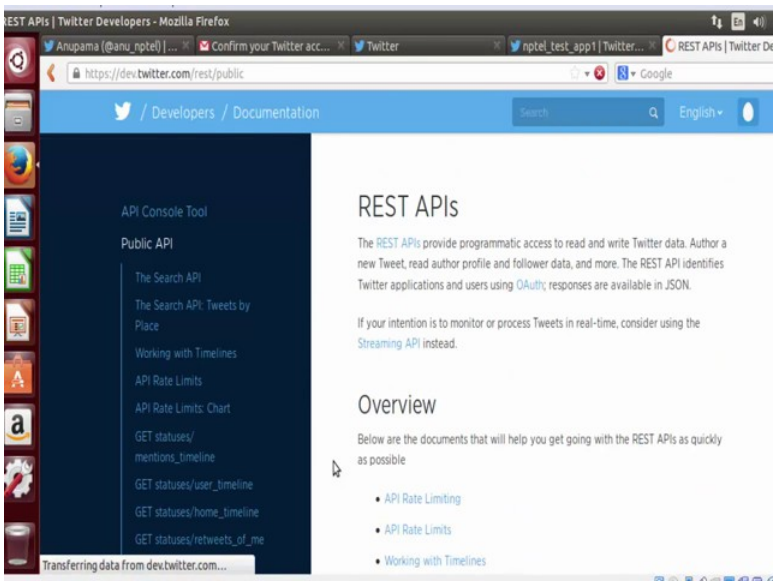


(Refer Slide Time: 06:33)



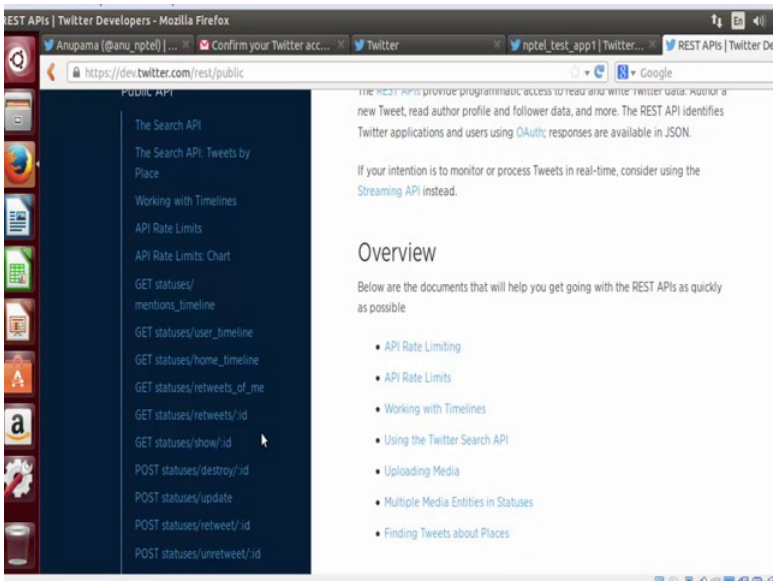
In this specific case, we will be using the rest API of twitter.

(Refer Slide Time: 06:35)



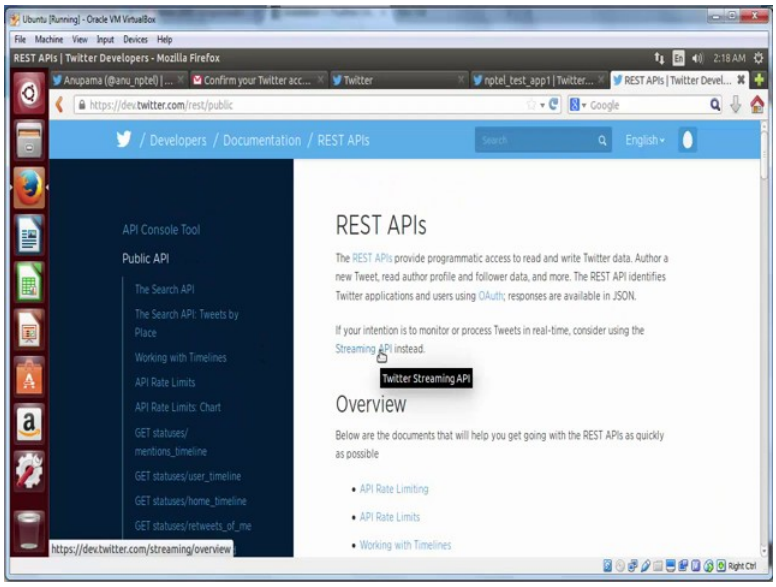
The rest API provides functionality to collect various kinds of data.

(Refer Slide Time: 06:45)



You can notice on the left panel that you can access data specific to a user, or public tweets or you can even get the follower and following information of users who have authenticated your app or of any particular user whose such data is public.

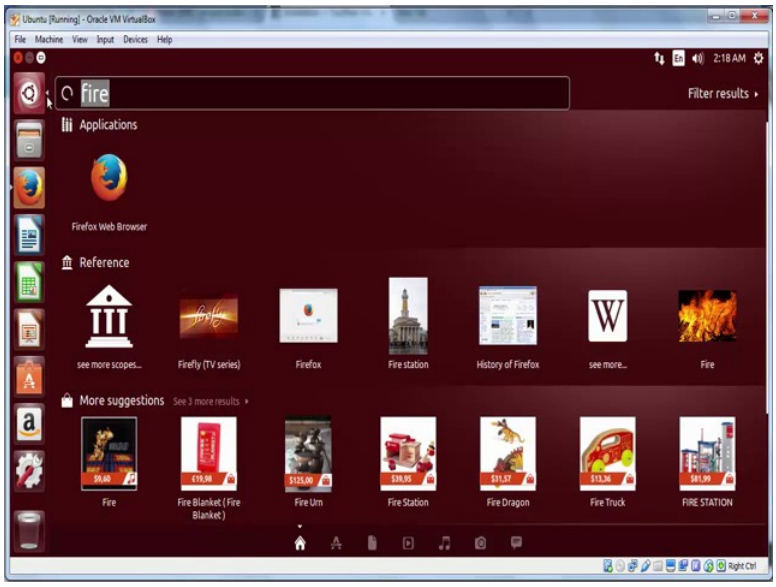
(Refer Slide Time: 07:07)



Apart from using search API, we will also be using the streaming API. The streaming

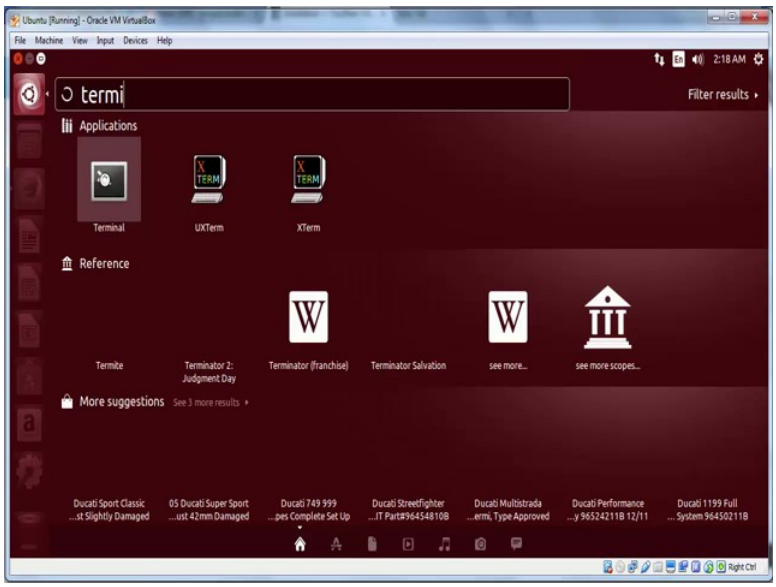
API gives you the functionality to monitor and process twitter data in real time. To use twitter API we will be using a python wrapper, which is called Twython.

(Refer Slide Time: 07:25)



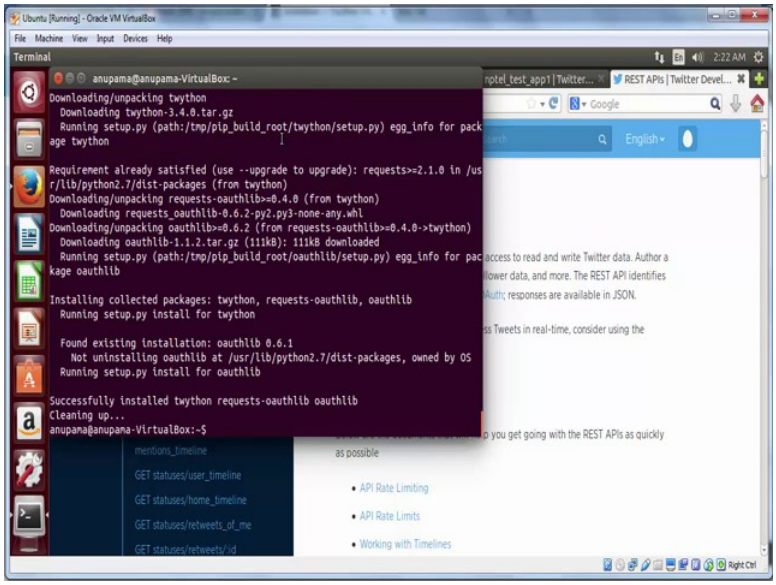
Now before we move on to how we use Twython, let us install Twython first.

(Refer Slide Time: 07:29)



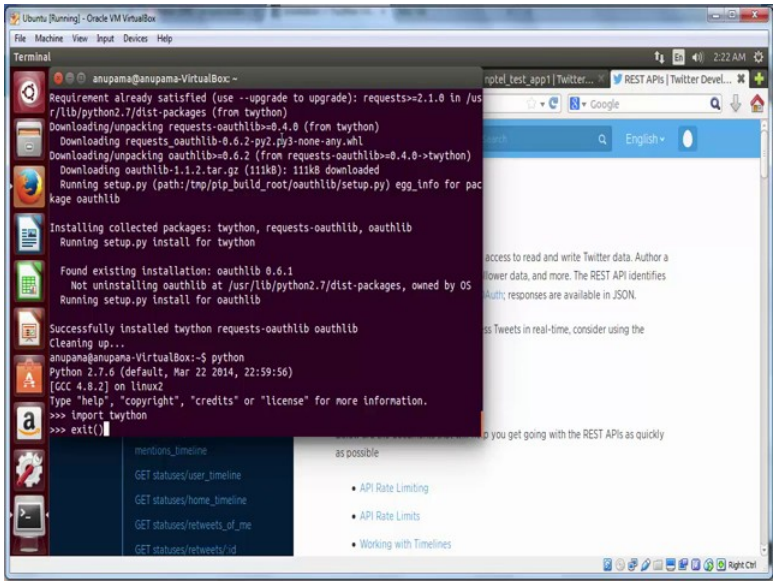
Go to your terminal.

(Refer Slide Time: 07:42)



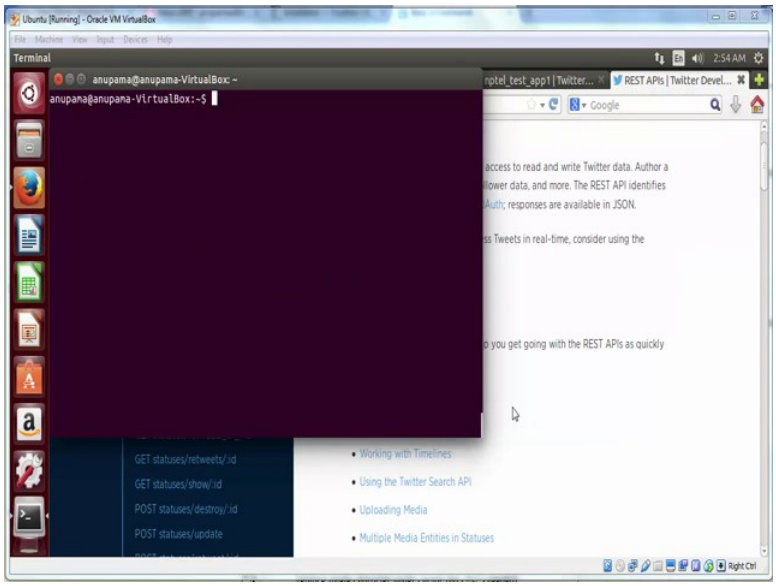
And type, the command sudo pip install Twython. Now you must already have the installed pip if you do not, then go to the previous week's video and look at the tutorial on how to install pip. When you press enter you will be asked for you sudo password and the installation of Twython will start.

(Refer Slide Time: 08:12)



You can check whether Twython has successfully installed or not by going to Python cli and typing import Twython. If it has been successfully installed, then you will be able import Twython. Now, let us exit the Python cli.

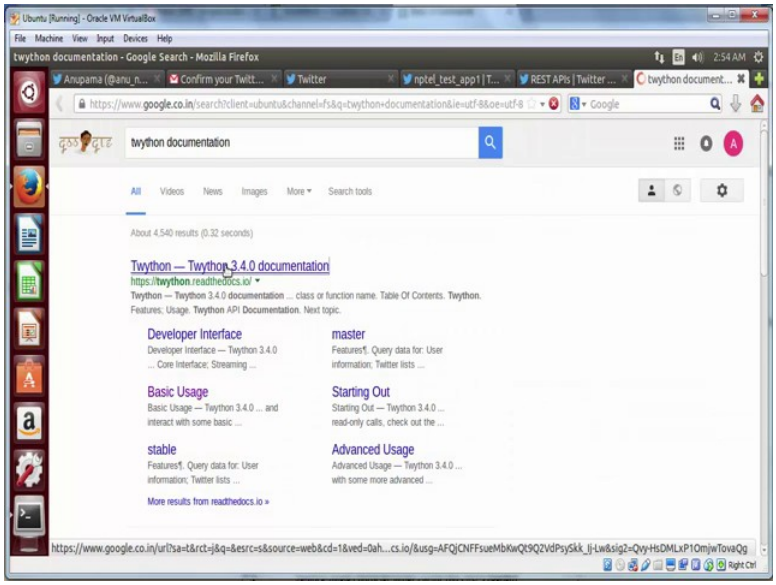
(Refer Slide Time: 08:32)



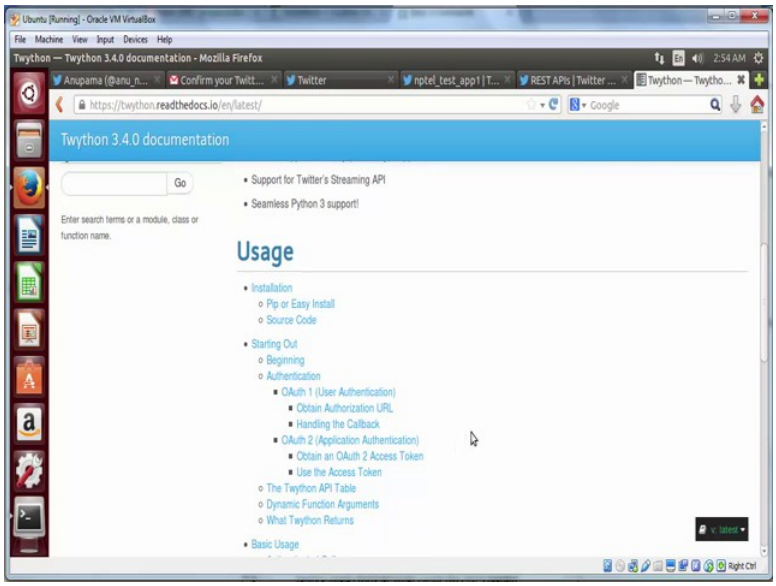
In the next step, we will see how we can use Twython and the twitter API credentials

which we have created together to be able to collect data from twitter. Now let us look at how Twython works. Let us go to Twython's documentation to understand the basic usage of Twython.

(Refer Slide Time: 08:57)

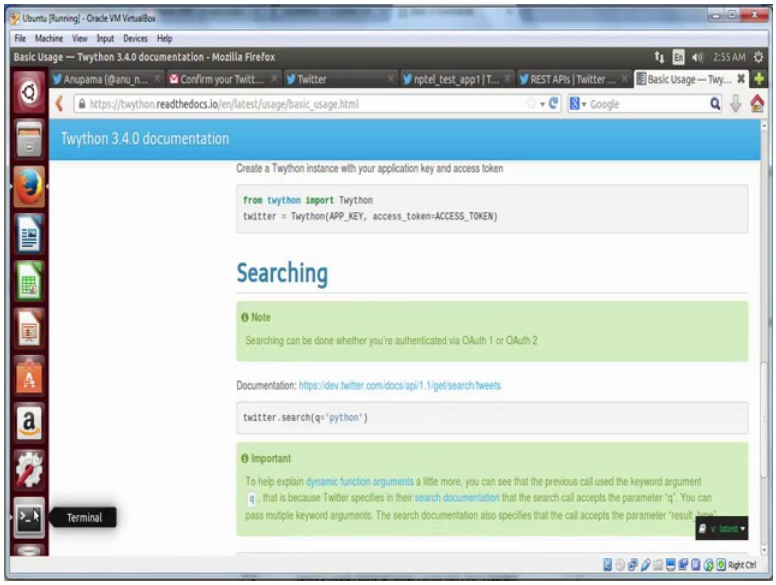


(Refer Slide Time: 09:04)



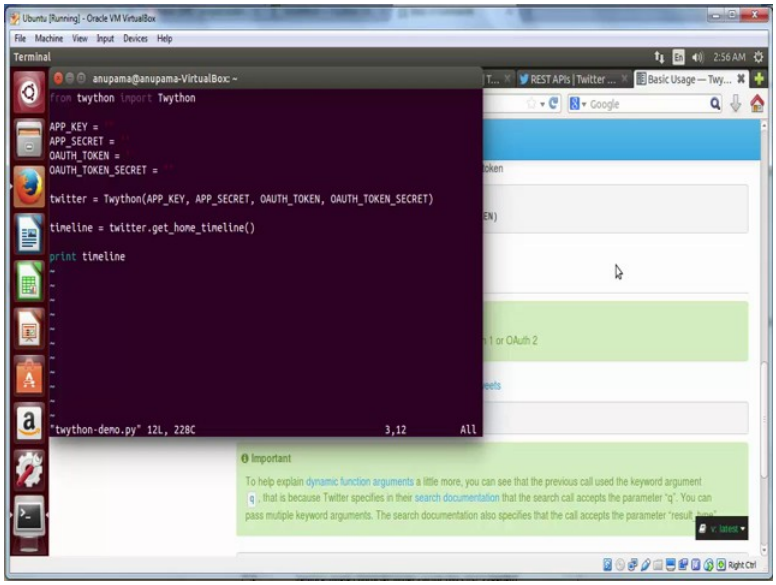
In the documentation, you will notice how to use Twython using several parameters.

(Refer Slide Time: 09:32)



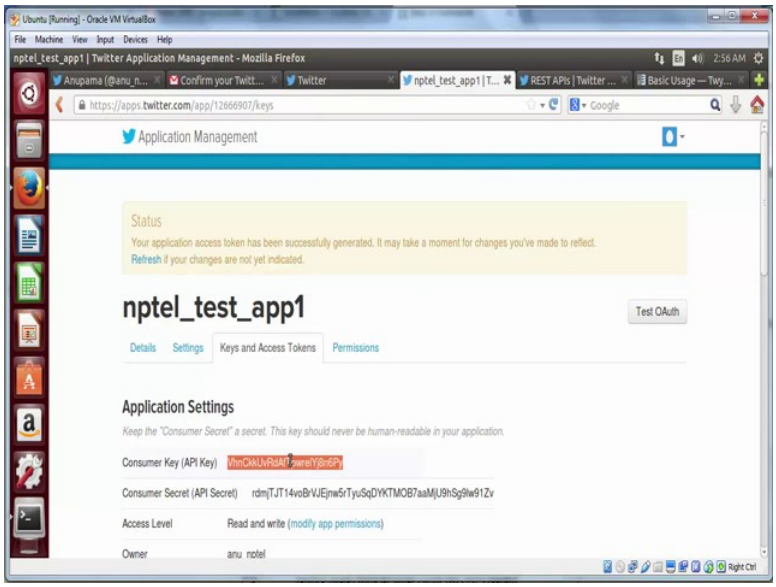
Go to the basic usage, and you will see that you need the four keys which we already created in the previous step when we created our twitter app. We will be using these four keys to access data. There are several end points which Twython provides which are very similar to the twitter API itself.

(Refer Slide Time: 09:36)



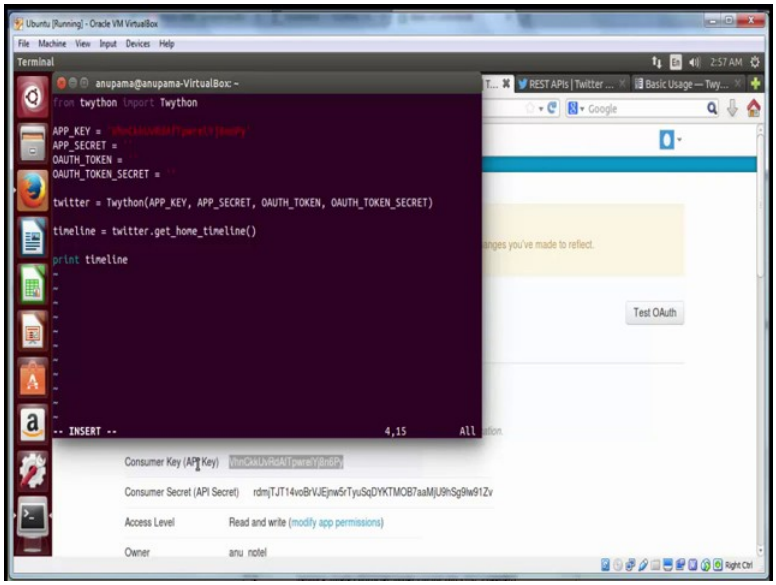
Let us go back to the terminal, and create a Python file. You will notice that this file imports Twython using the command form Twython import Twython. It also has a list of four variables which are blank right now let us fill these parameters using the app which we created earlier.

(Refer Slide Time: 10:06)



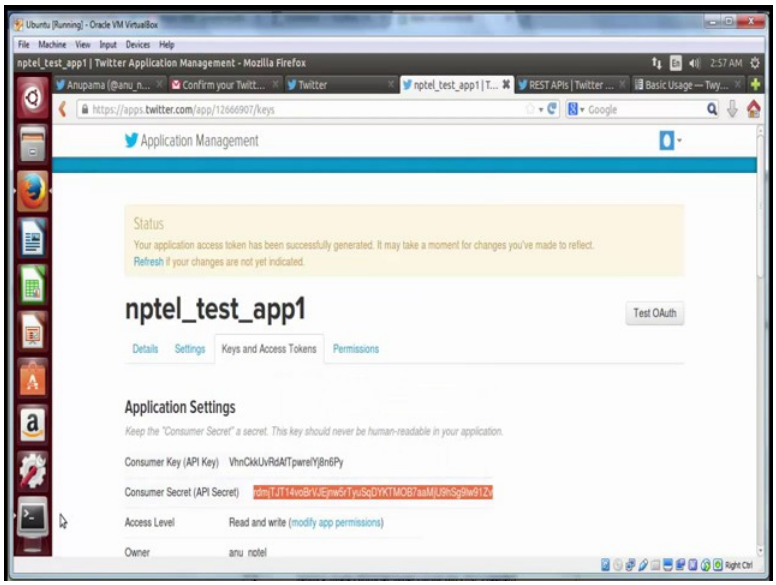
Get hold of the consumer key.

(Refer Slide Time: 10:12)

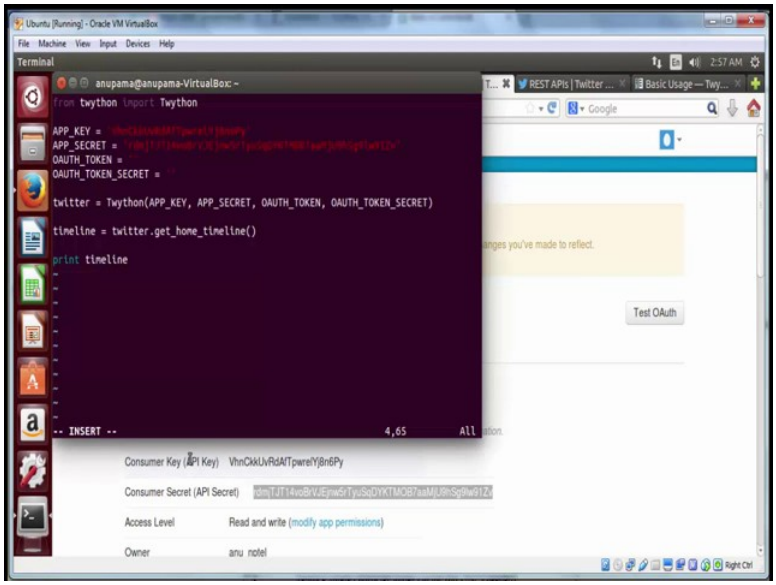


And put it in place of app key.

(Refer Slide Time: 10:23)

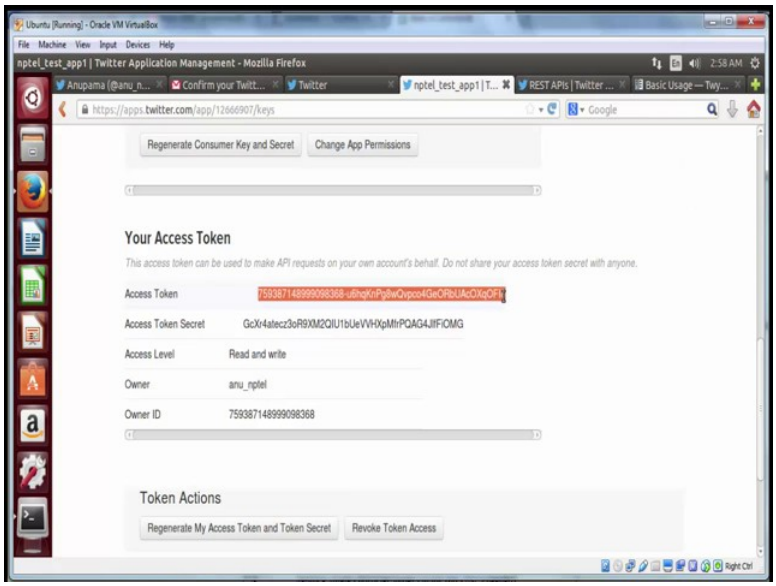


(Refer Slide Time: 10:25)

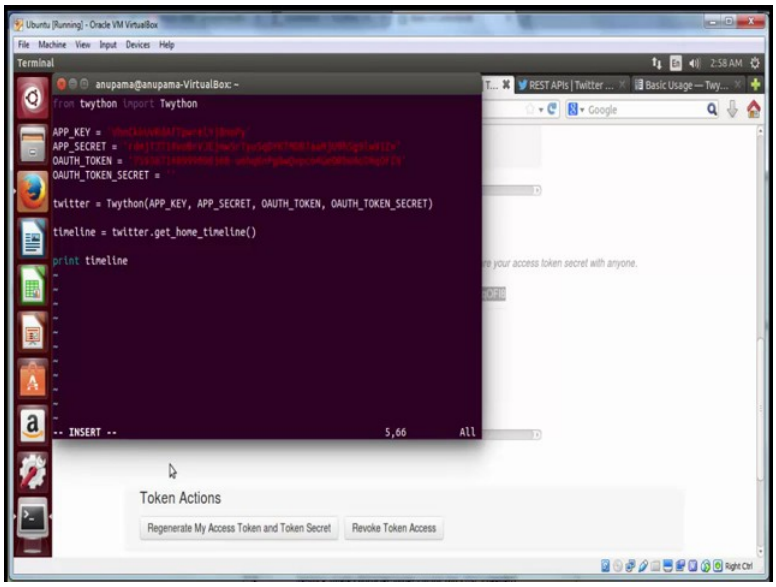


Similarly, you will be using the consumer secret, which is also called the API secret. Now remember we had created two more set of keys, which is your access token, and access token secret.

(Refer Slide Time: 10:32)

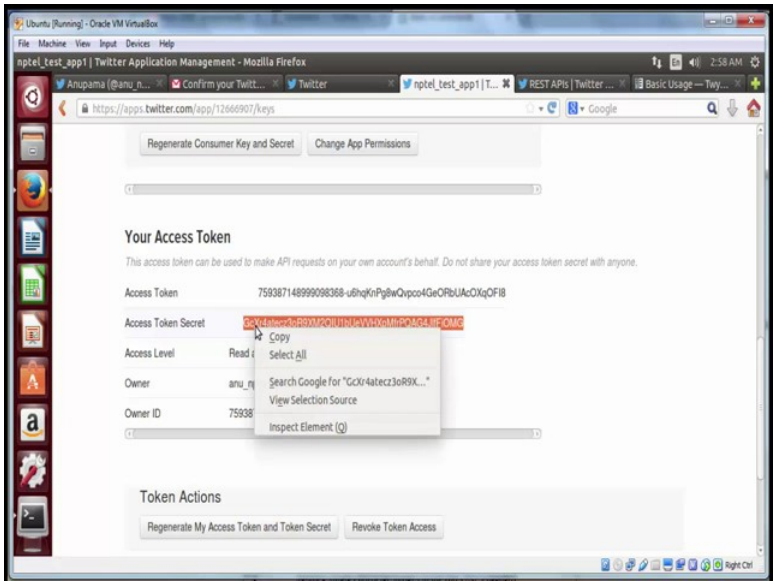


(Refer Slide Time: 10:39)



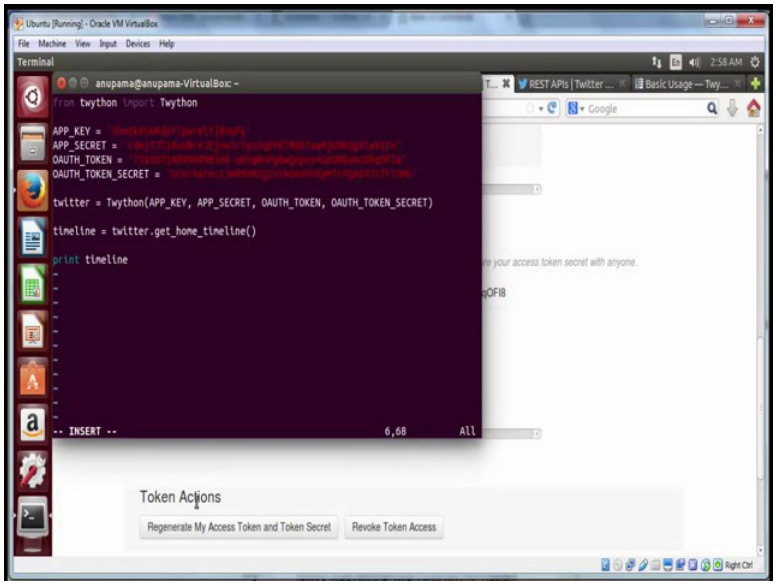
Copy paste your access token in the oauth token variable.

(Refer Slide Time: 10:45)



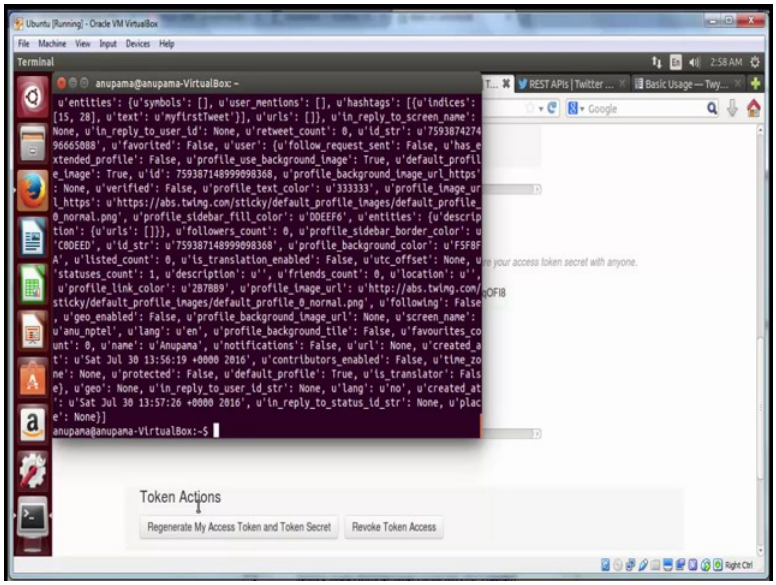
And copy the access secret in place of the oauth token secret.

(Refer Slide Time: 10:49)



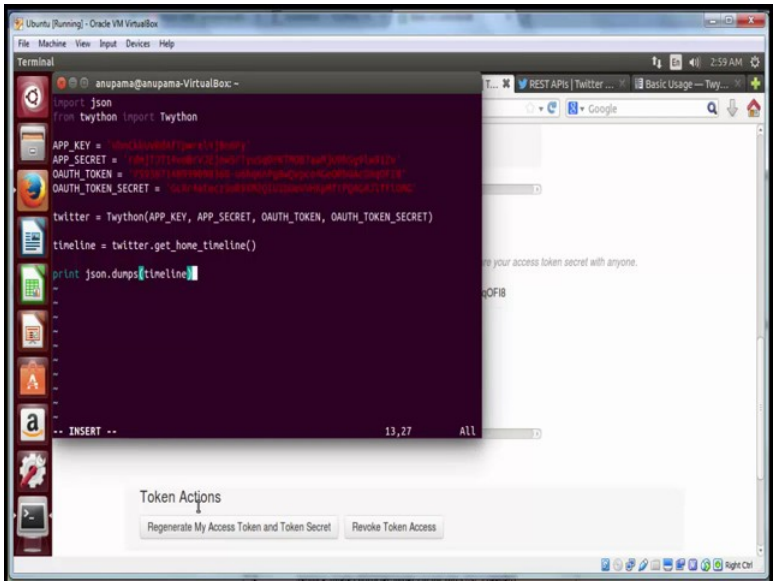
Now next line is just an initiator, which connects you to the twitter API. In this particular case, we are going to fit the twitter timeline of the authenticated user that is your own account. Timeline equal to twitter dot get home timeline will get you all the tweets, which appear in your timeline. And we are going to print this.

(Refer Slide Time: 11:14)



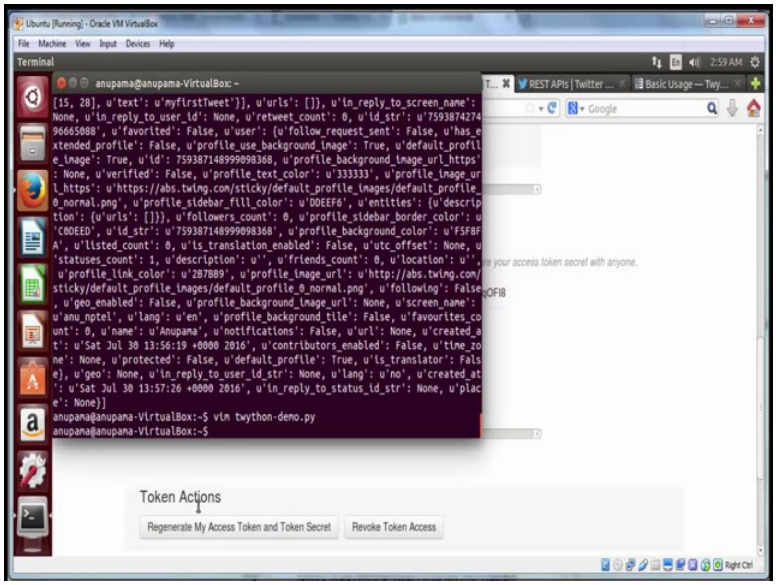
Let us see what happens when we execute this program. Run python, python demo dot py, and there is a bunch of texts this text is pretty unreadable. And it is also not encoded in Unicode.

(Refer Slide Time: 11:28)



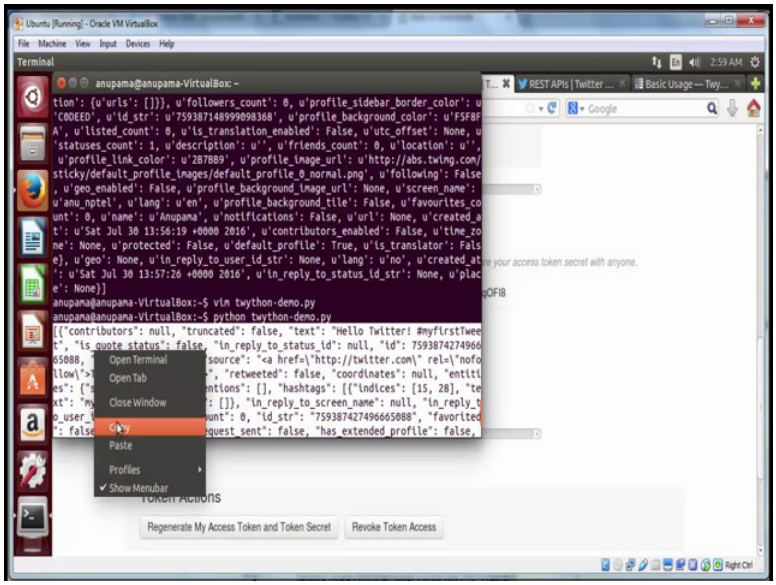
Therefore, we are going to make few changes. We will import json library which lets you format the json data; and we are going to print this data in string format by using the command json dot dumps. Now let us save this file again.

(Refer Slide Time: 11:51)



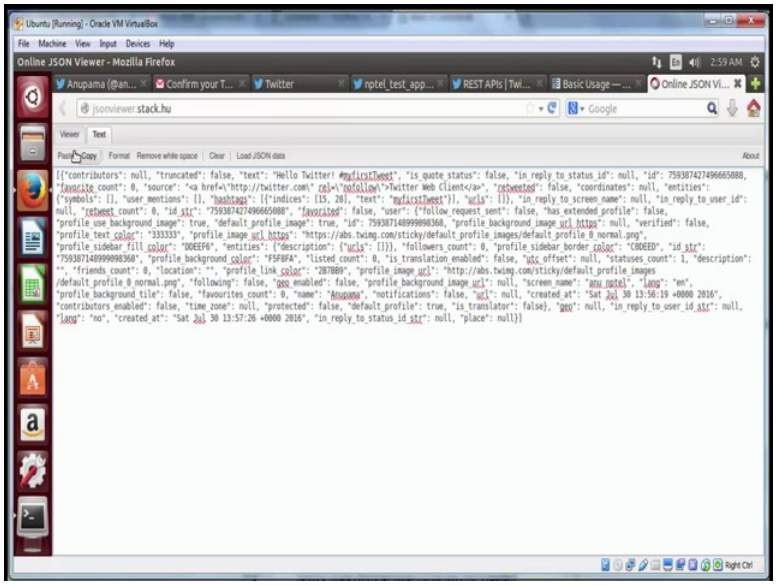
And see what happens then we run it. This time we have a much cleaner looking code, but it is still not readable.

(Refer Slide Time: 12:07)



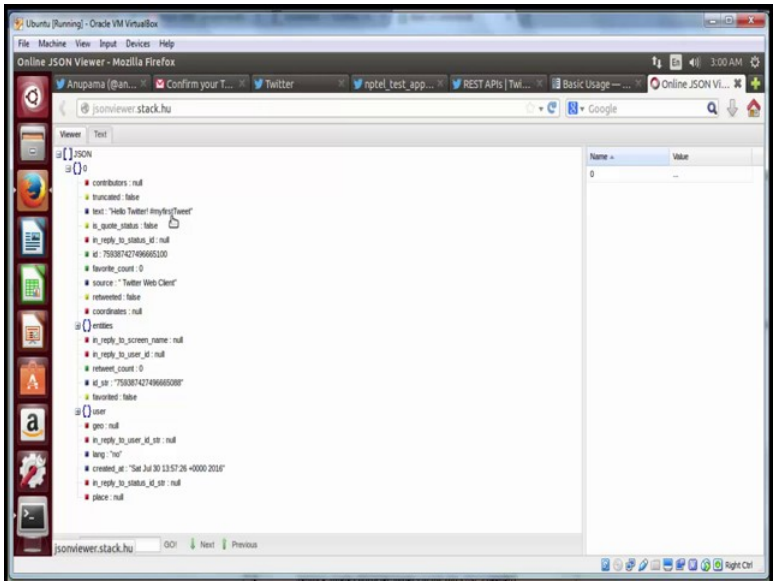
Let us use a web service to make this data look more pretty. Copy this text.

(Refer Slide Time: 12:23)



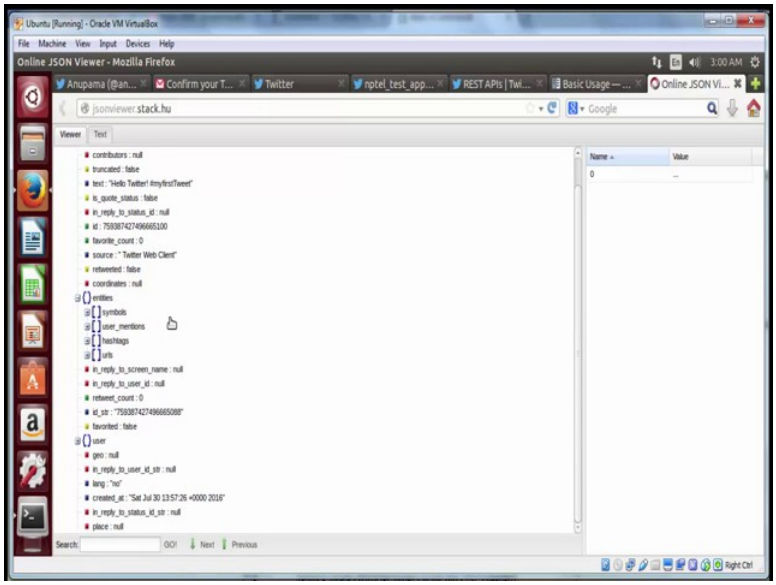
And in your browser go to json viewer dot stack dot hu. This is a service, which converts json format into a readable format, paste the text and click on viewer.

(Refer Slide Time: 12:30)



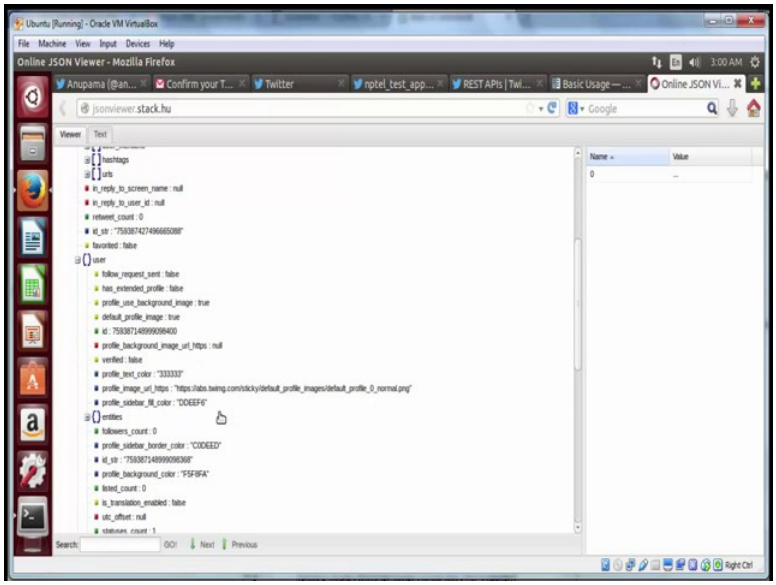
Now you can see that this is a list, which has one object indexed by 0. If you expand it you will be able to see various parameters like the text of the tweet.

(Refer Slide Time: 12:48)



Remember that this is the first tweet, which you posted. You can also click on entities and see the other parameters of the tweet.

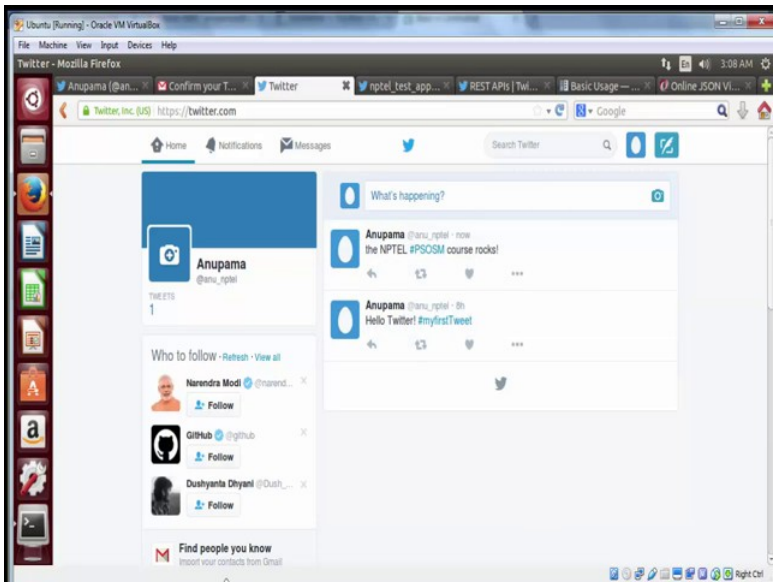
(Refer Slide Time: 12:59)



You can expand the user information and find out about the user, which has posted this particular tweet. You will notice that in our case the number of followers and the number

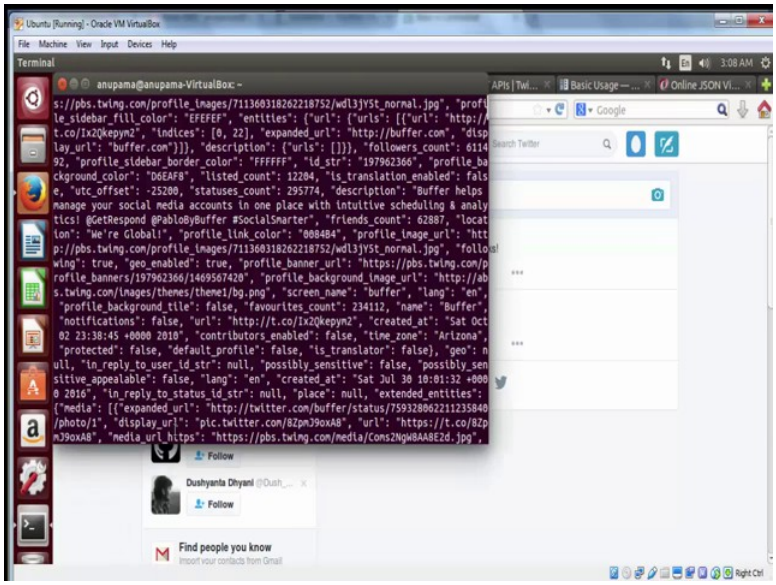
of friends is equal to 0, because we are not following anyone and we do not have any followers yet in our freshly created account.

(Refer Slide Time: 13:18)



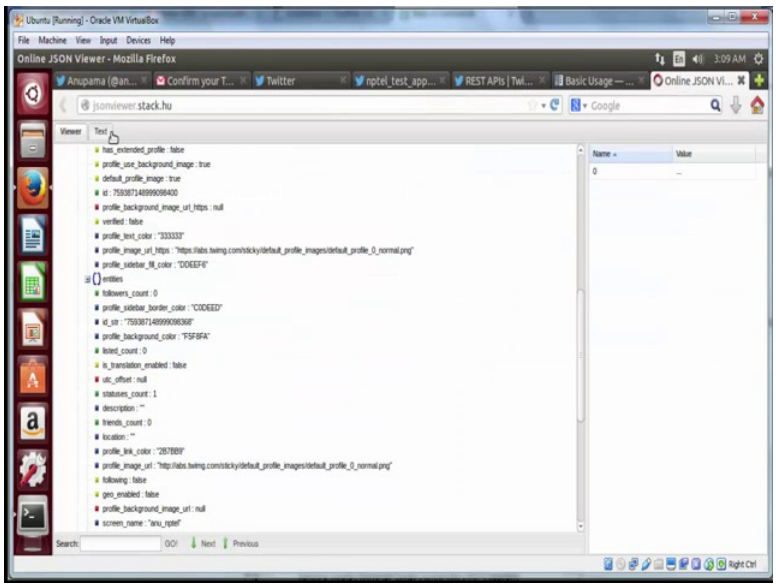
Let us go back to twitter to check the same. Let us post a newer tweet and see how it gets reflected by the program. You can type anything. Compose a tweet and simply click on the tweet button.

(Refer Slide Time: 13:43)



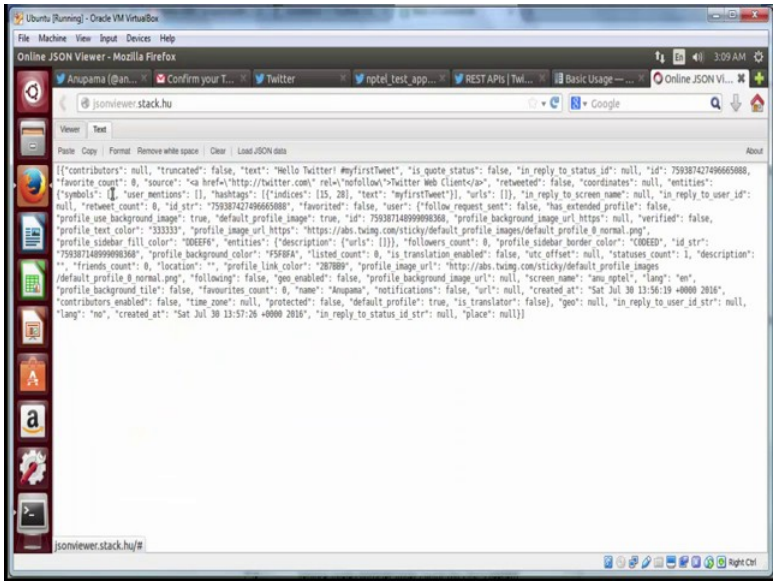
Now let us run the program again and see if this gets captured by the program. Go back to your terminal and type python space the same file name which you created in the earliest step. Now we again have a bunch of text. Let us copy that again into the web service, which we earlier used.

(Refer Slide Time: 14:08)



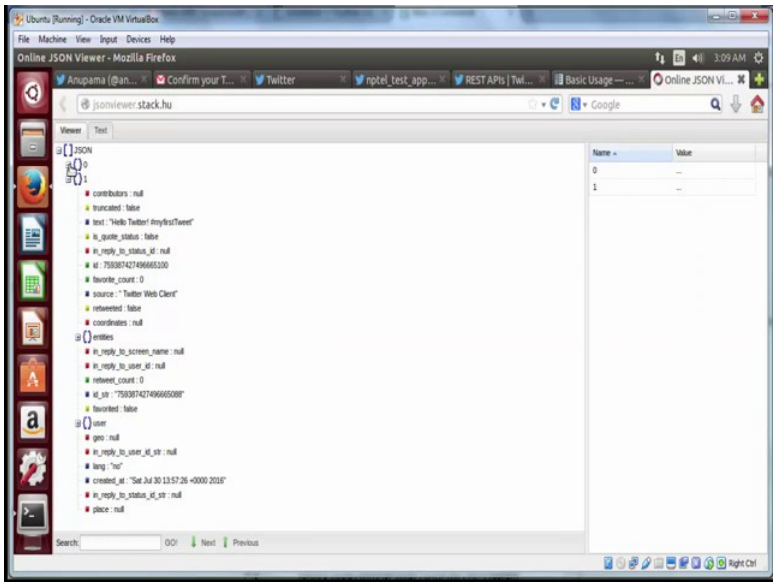
This time the text does look bigger.

(Refer Slide Time: 14:11)



And if you click on viewer you will be able to see two objects in the list.

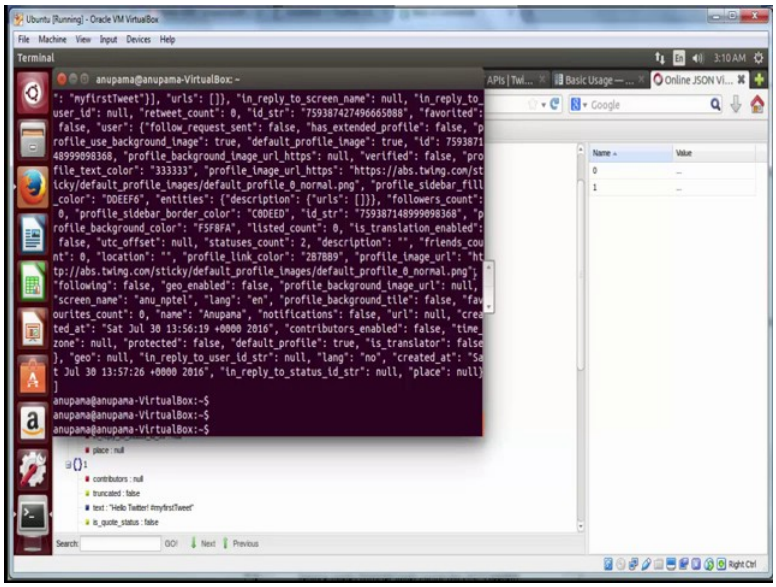
(Refer Slide Time: 14:17)



Now if you expand the first one, you can see the previous tweet; and if you see the 0th

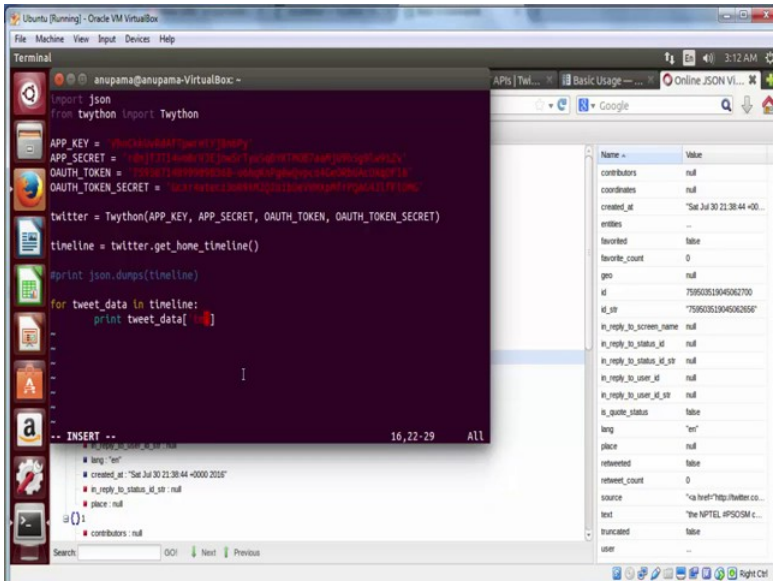
object that is the latest tweet. So, now we are ready with the simple program which can capture a user timeline. Using this program, you cannot just capture your own user timeline, but also fetch the public data of any Twitter user. You can check how to do that by going through the Twython documentation.

(Refer Slide Time: 14:46)



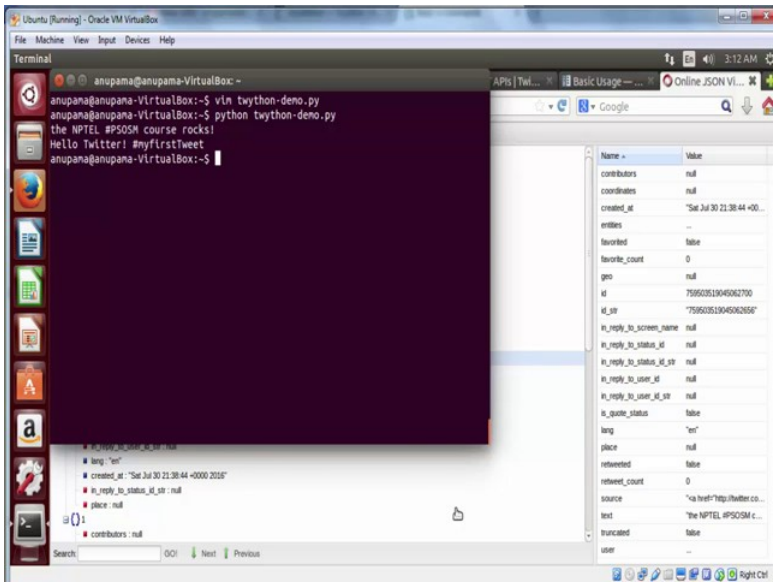
Let us make few more modifications in the program, so that we have a bunch of more readable output.

(Refer Slide Time: 14:52)



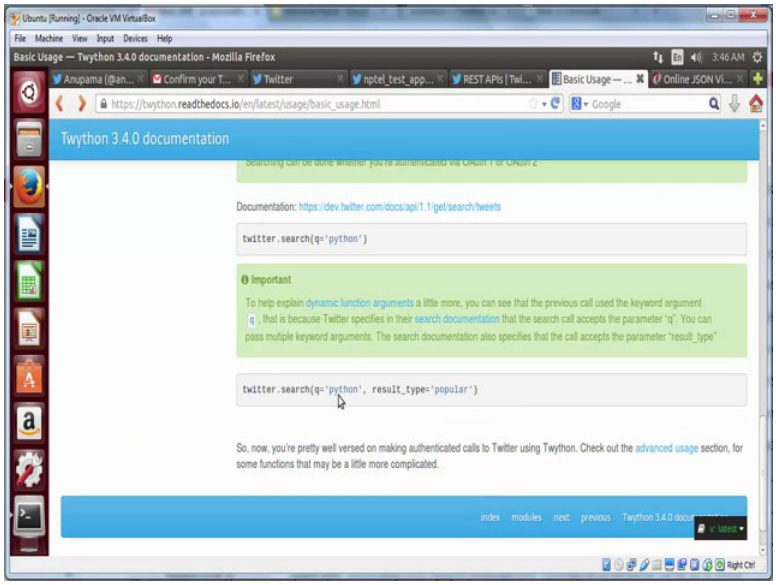
What if we had to print only the text of the tweets which the user has posted so far, to do that, we are going to iterate over timeline data one by one and print the text available in each tweet data. We will start a loop and then access the json object which is returned by the twitter API. Note that each tweet is basically a dictionary with a set of keys, in this case we want to access the text.

(Refer Slide Time: 15:27)



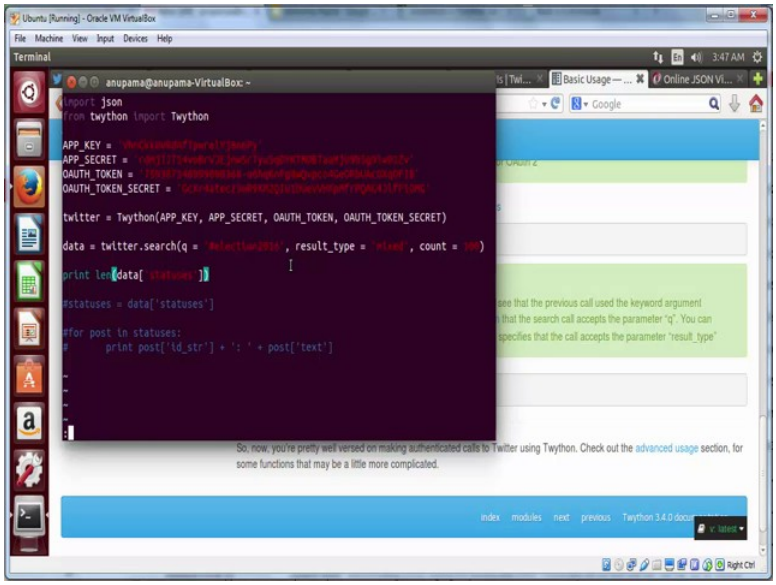
Now let us run the program again and see what happens. Now you will see the output as two lines, which is the two tweet's text. Now, we are ready with the simple program to fetch your own Twitter timeline.

(Refer Slide Time: 15:51)



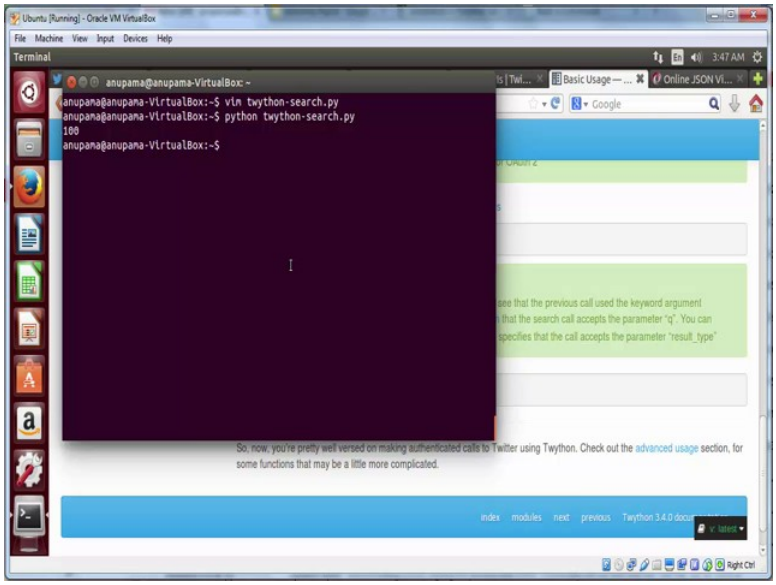
Now let us use Twython to do more things. You will notice that Twython documentation says that you can also search tweets based on a specific keyword.

(Refer Slide Time: 15:59)



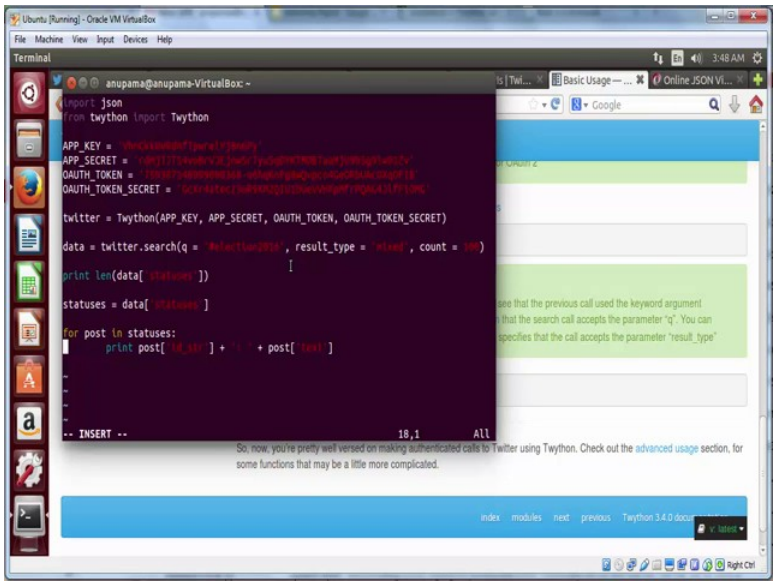
Let us see how we can do that. Go back to the terminal and create a python file. Now I have already created one, which looks very similar to the previous one. It contains the same list of app key, app secret, access token and access secret. The only difference being that now we are going to collect data from twitter search end point based on a specific key word, in this case election 2016. And the result type is mixed which means that it will be a mix of popular and recent tweets. We also have count equal to 100, which means that this will be the number of tweets, which will be returned. In this case, we will print the amount of tweets, which are returned to us by the twitter API. Let us save this program and run it.

(Refer Slide Time: 16:51)



You can see that the output is 100.

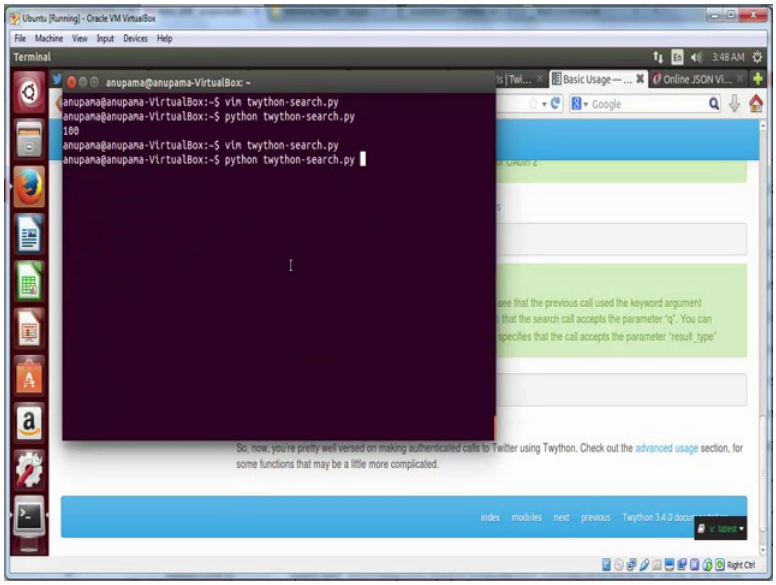
(Refer Slide Time: 17:05)



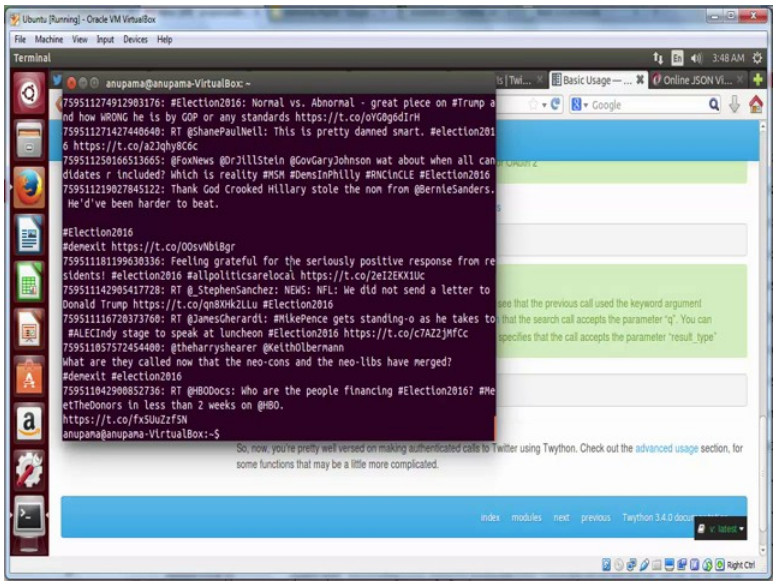
Let us go back to the program and see what is the data which actually exists returned by the twitter API. Here we can use statuses equal to data statuses to access the list of tweets which Twitter API has returned through this method. And let us try to print the text of

each tweet along with the tweet id by using the following code.

(Refer Slide Time: 17:27)

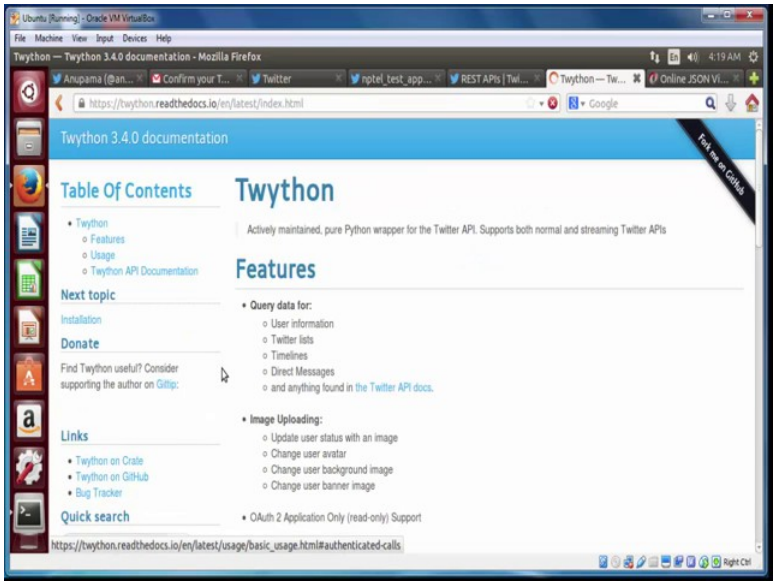


(Refer Slide Time: 17:32)



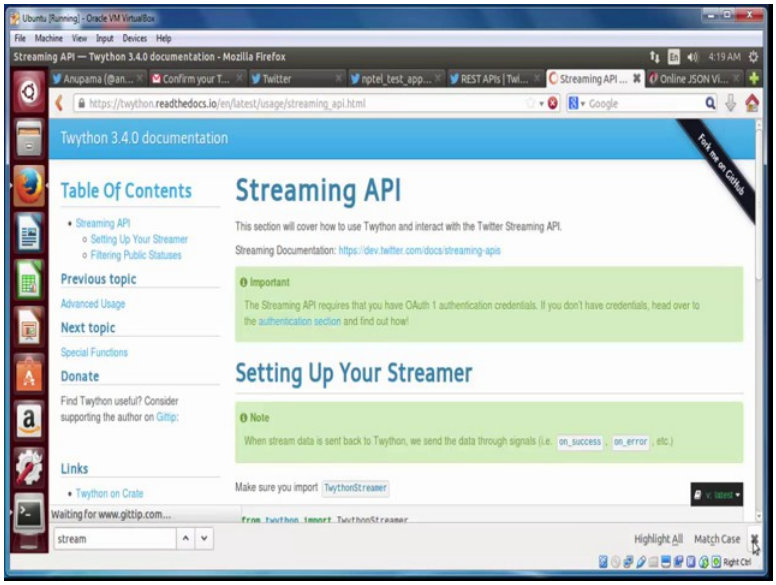
Now when you save the file and run it. You will notice a bunch of text. Here you will notice that first string is the tweet id, and second part of the text after the colon symbol is the tweet text.

(Refer Slide Time: 17:50)



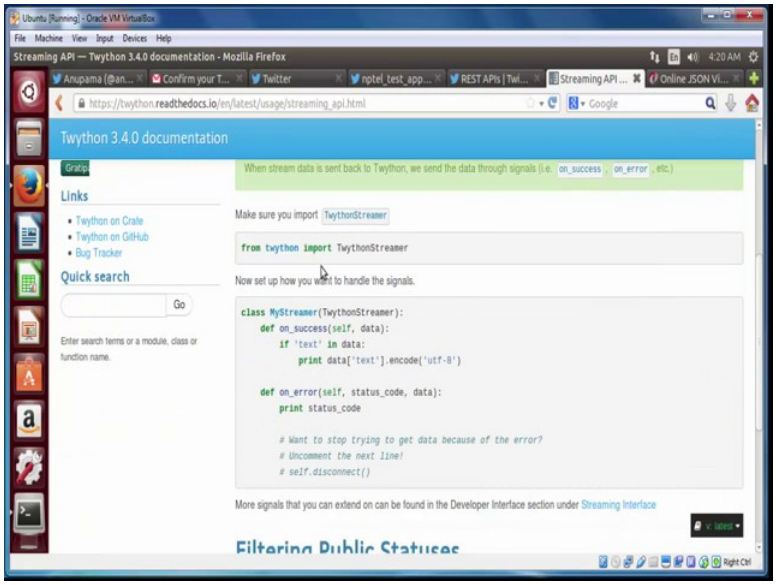
Now, we will see how to use the streaming endpoint. Twython provides the functionality to access the streaming end point which you can find in Twython's documentation.

(Refer Slide Time: 18:12)

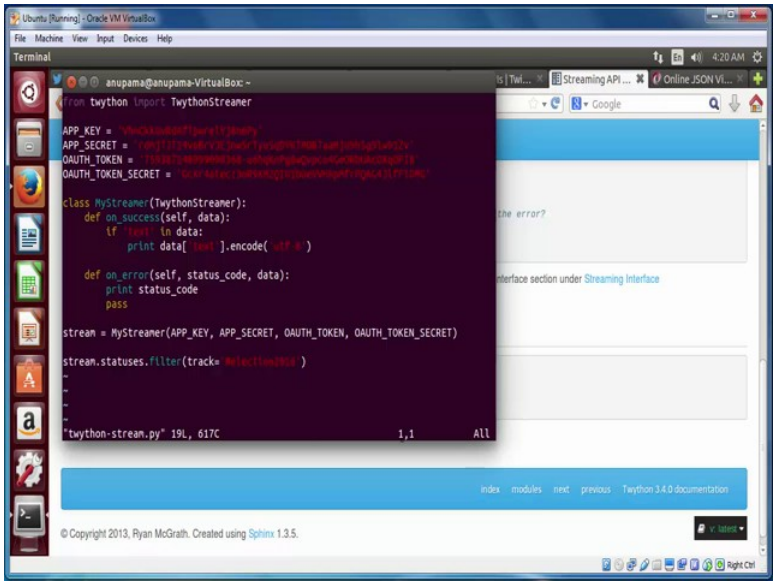


The documentation already has a starter code, which we will use to get data from twitter in real time.

(Refer Slide Time: 18:19)



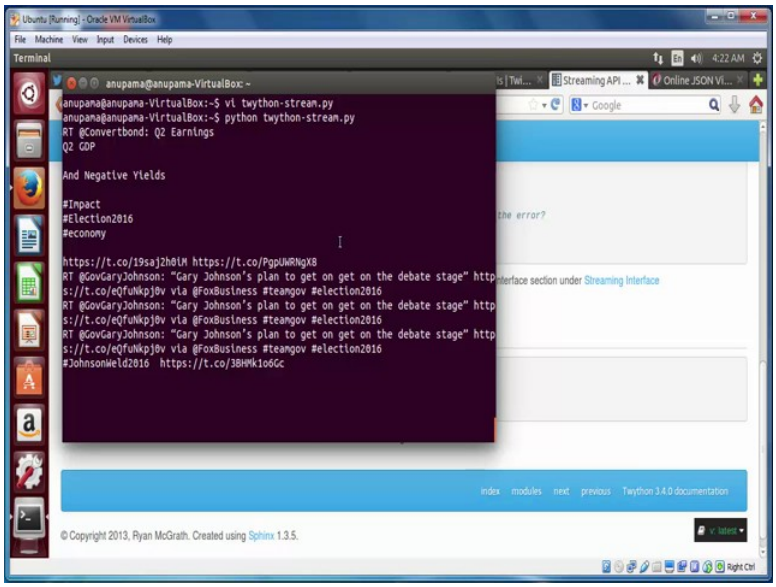
(Refer Slide Time: 18:25)



Let us go back to the terminal and create a python file. Now I have already created a file, which is very similar to the previous ones with the list of all the access tokens and secrets which we want. However, in this case, we have a class MyStreamer which has basically two functions which define what we will do, when we get the data successfully, and when we fail. In case we get the data successfully, we are going to print the text of it

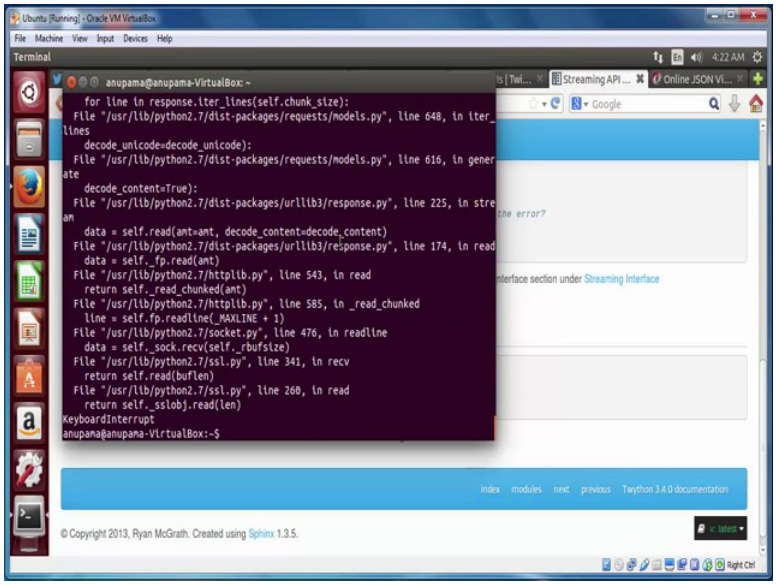
after encoding it into utf-8. And in case, it shows an error we are going to print this status code and pass. We will use the same keyword, which we used before in the previous example, where we were searching for past tweets made using the keyword #elections2016. In this case we are going to do the same in real time.

(Refer Slide Time: 19:19)



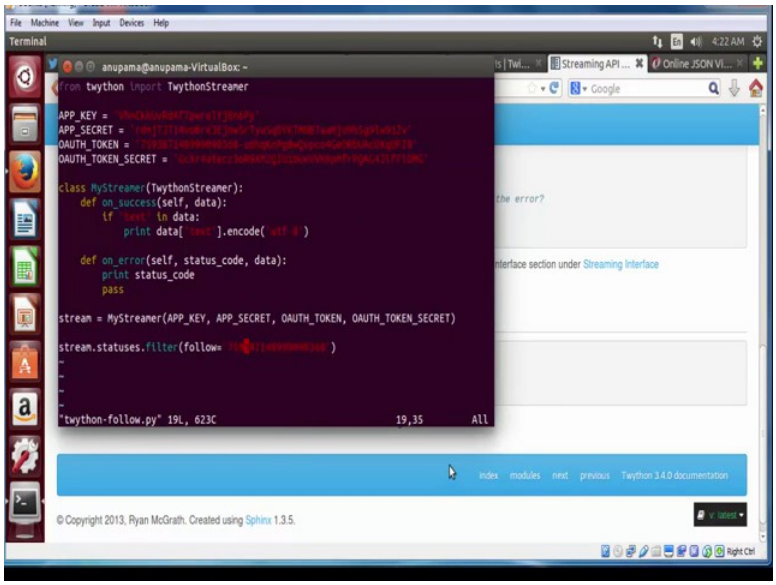
Let us save the file and run it. Now remember since is this data is being collected in real time, you will not see the results immediately, depending on when the tweets are being posted. You will notice that tweets will start coming if the particular keyword is still actively being used. So, the tweets, which you see now, are being generated in real time. Now, what if we wanted to track a user instead of tracking a particular keyword?

(Refer Slide Time: 20:03)



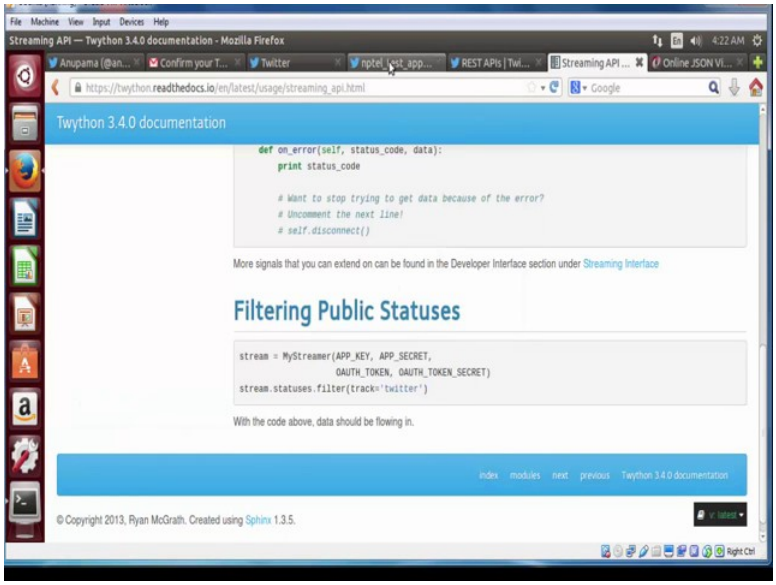
That can also be done using the Twitter API.

(Refer Slide Time: 20:08)



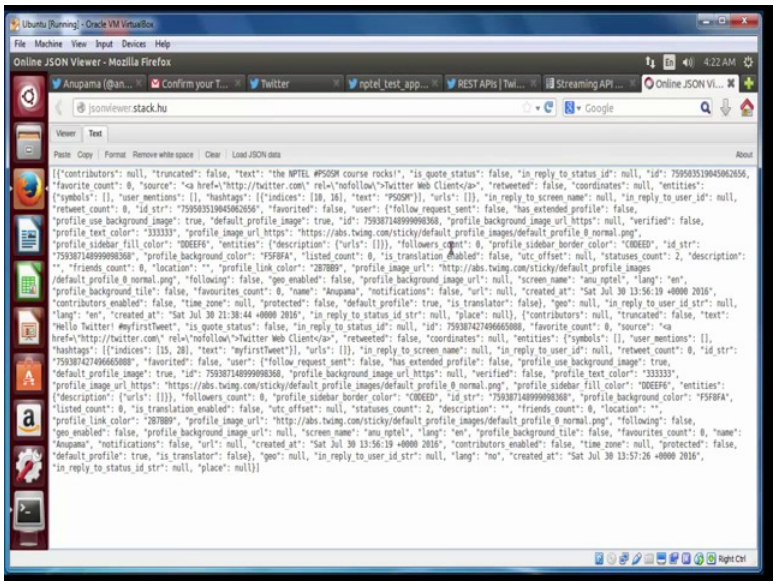
Let us see another example to see how that is possible. We have a slightly different file this time with only one change instead of tracking a keyword, we are following an id; this is your own account's Twitter id which we got from the previous step.

(Refer Slide Time: 20:28)

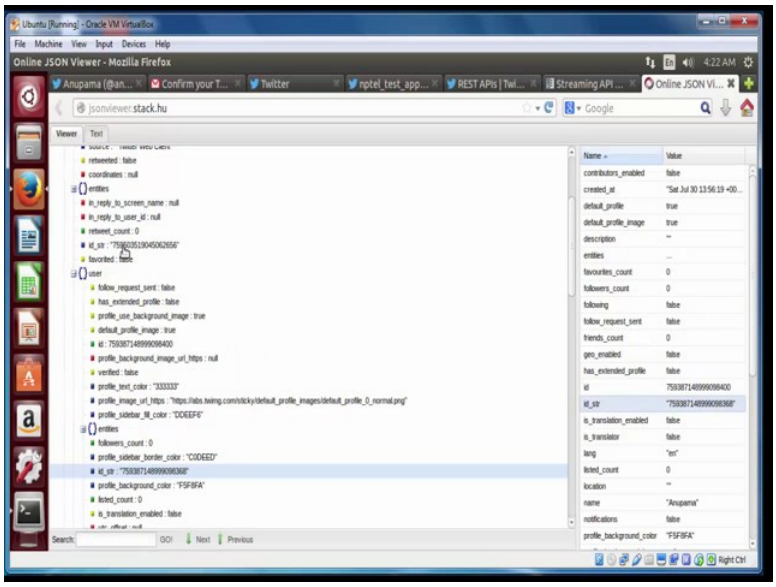


You can check this by going to the text, which you dumped in the json viewer.

(Refer Slide Time: 20:30)

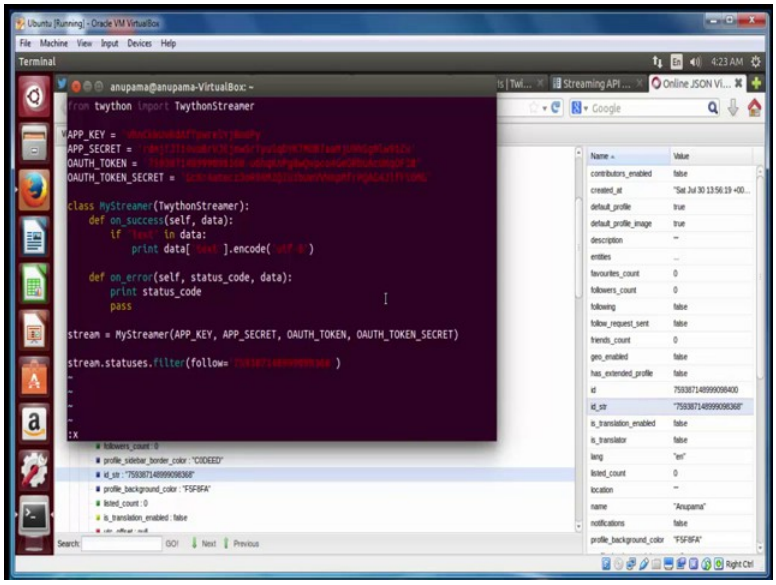


(Refer Slide Time: 20:36)



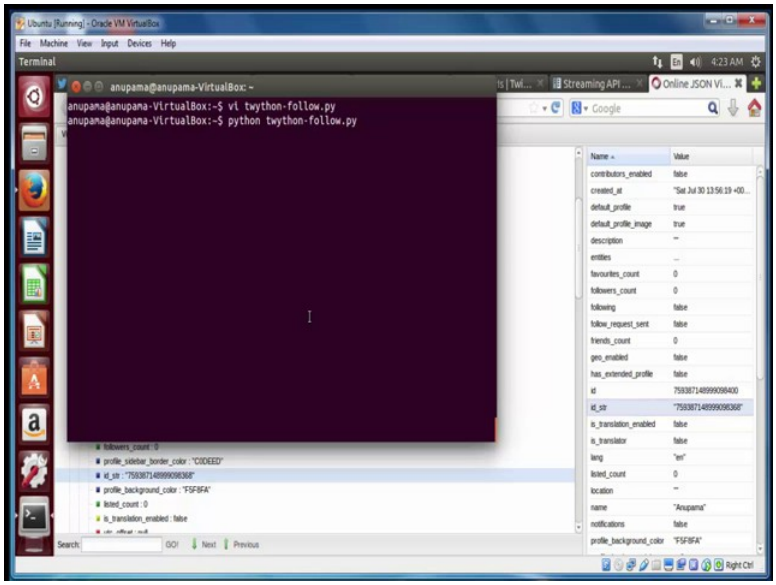
And check the id_string in the user part of the data. Copy; paste that user id into the program.

(Refer Slide Time: 20:57)



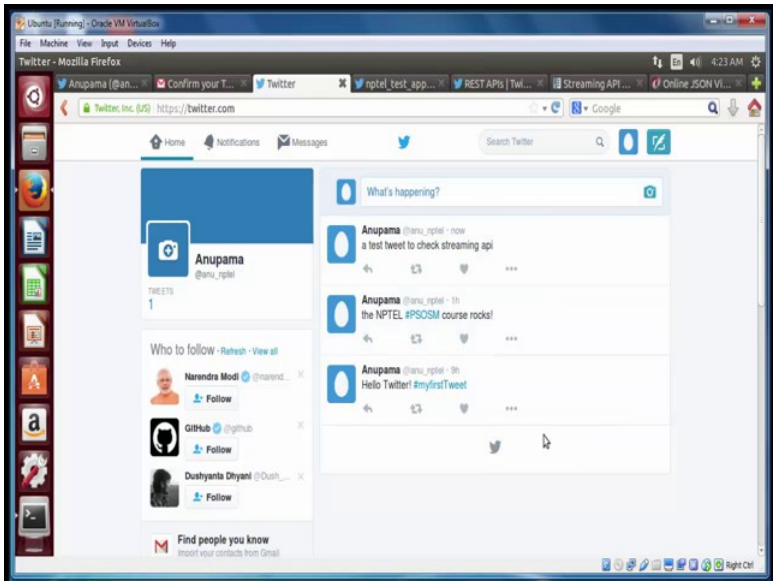
And you will able to track that particular user. Now, let us save this file run it and see what happens.

(Refer Slide Time: 21:05)



So, we are not able to see any text that is because nothing is happening in real time in our account.

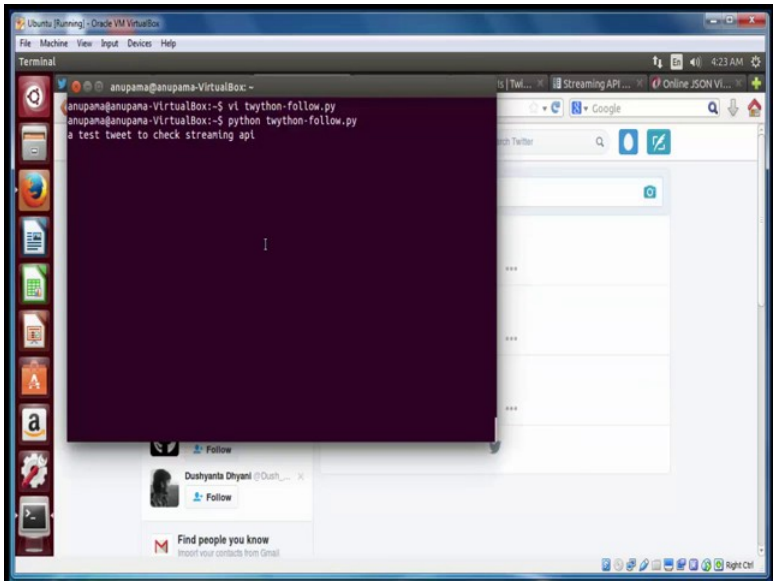
(Refer Slide Time: 21:19)



Let us try posting the tweet and see how that reflects in the code. Go to your Twitter account and post anything. As soon as you click on the tweet button, you will be able to

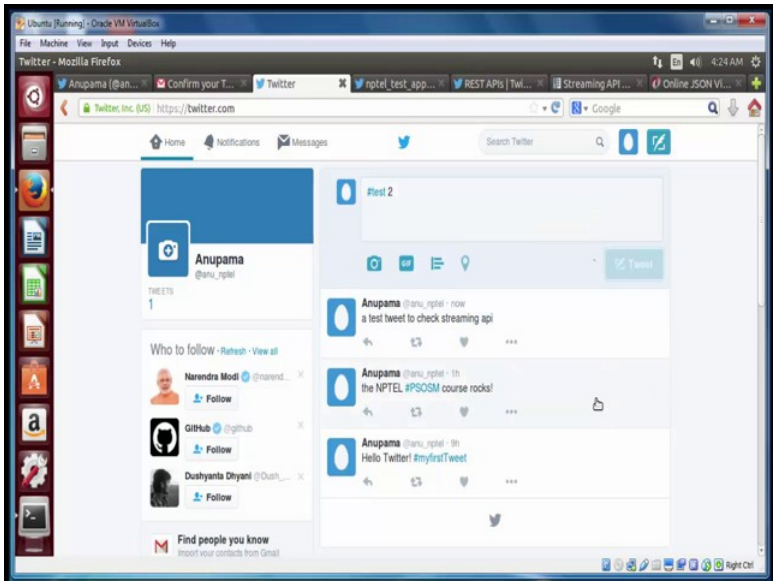
see the same change in your terminal.

(Refer Slide Time: 21:44)



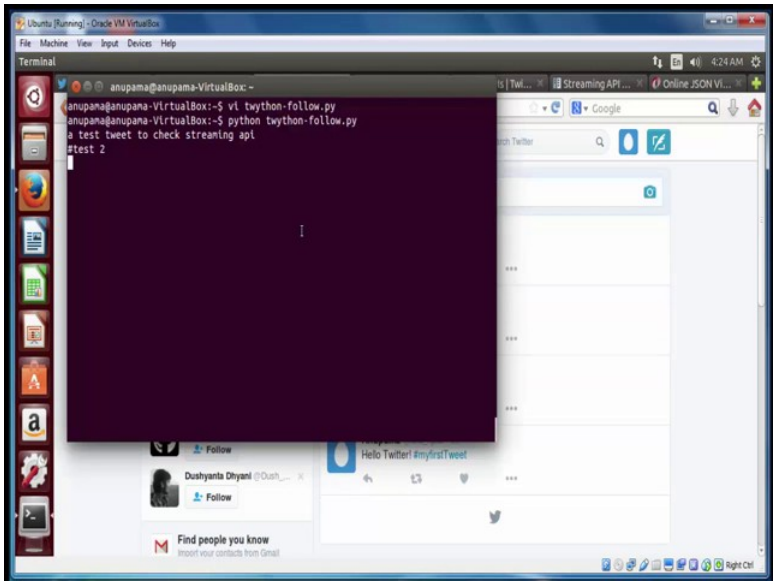
Now your program has been successfully able to capture the post, which you posted in real time.

(Refer Slide Time: 21:53)



Let us make another tweet.

(Refer Slide Time: 22:03)



And go back to the terminal. You will notice that the new tweet has appeared. In the next tutorial, we will see how we can save this data in a database.