Downloading Tweet Streams and Parsing

Ayan Bandyopadhyay

IR Lab.
CVPR Unit
Indian Statistical Institute (Kolkata)

To download this slide go to: https://goo.gl/AYWI1S

Downloading Tweet Streams and Parsing

- Downloading Tweet Streams

  It is imagined that Tweets are flowing in the microblogging site twitter.com like water in river.

- Tweet Parsing (JSON format)

Downloading Tweet Streams

- Downloading Tweet Streams
  - Basic requirement
  - Necessary steps
  - Write Java code

Downloading Tweet Streams

- Basic Requirement
  - OS: Windows or any flavor of Unix that supports Java. ✔
  - JVM: Java 5 or later ✔
    - Assuming JAVA_HOME has been set properly
    - Assuming CLASSPATH has been set properly

Downloading Tweet Streams

- Downloading Tweet Streams
  - Basic requirement ✔
  - Necessary steps

Downloading Tweet Streams and Parsing

- We will use Twitter4j library
- About Twitter4j
  - Go to http://twitter4j.org/en/index.html
  - Twitter4J is an unofficial Java library for the Twitter API.
  - With Twitter4J, you can easily integrate your Java application with the Twitter service.

Downloading Tweet Streams

- Necessary steps
  - Download twitter4j-core-4.0.4.jar and twitter4j-stream-4.0.4.jar ✔

    (http://mvnrepository.com/artifact/org.twitter4j/twitter4j-core/4.0.4 and http://mvnrepository.com/artifact/org.twitter4j/twitter4j-stream/4.0.4)
  - Put these jars in the CLASSPATH or add the jars in the CLASSPATH variable ✔
  - Create new Twitter App and get Access tokens
  - Save these Access tokens in the init.properties file

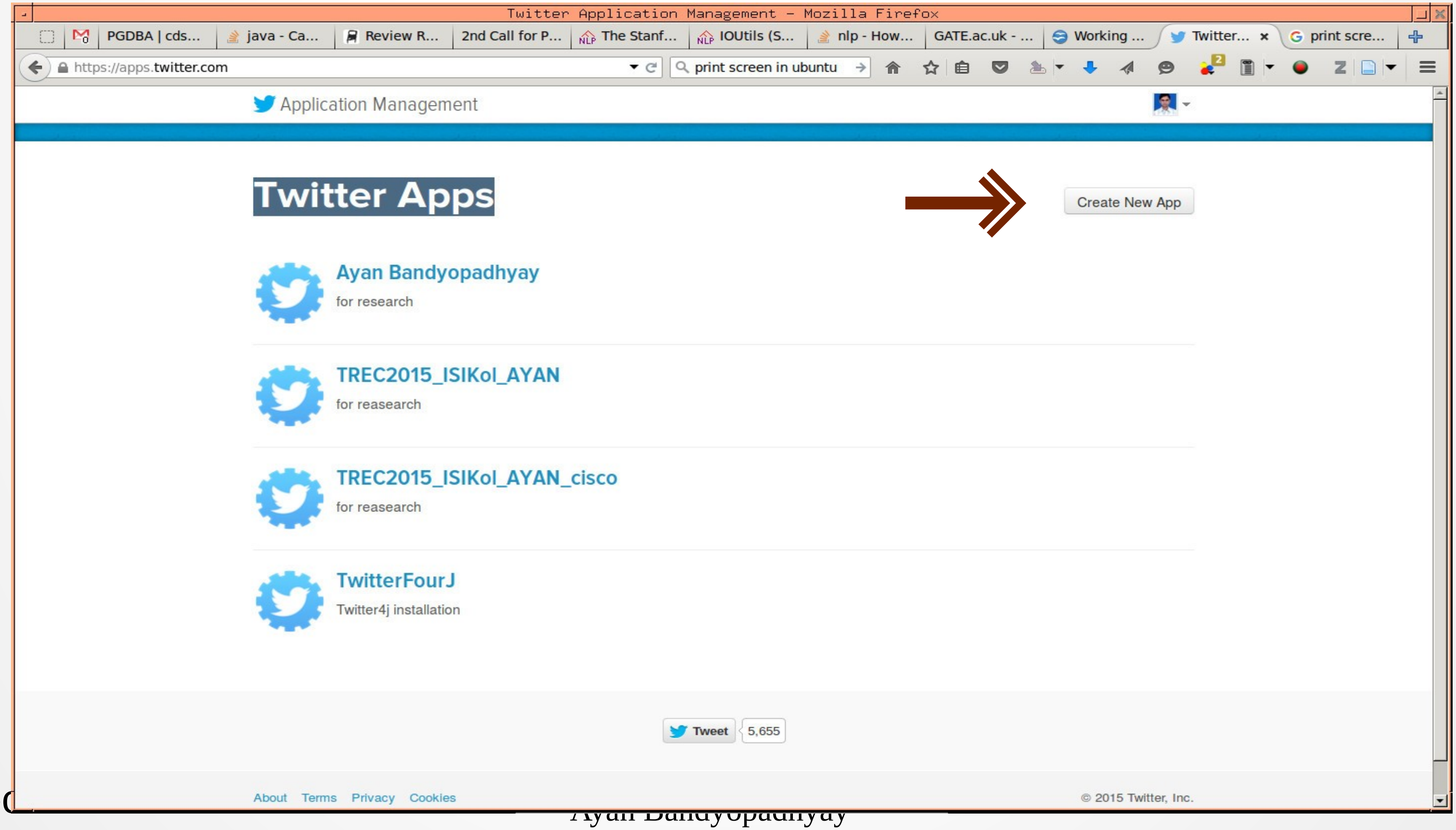
Downloading Tweet Streams

- Necessary steps
  - Download twitter4j-core-4.0.4.jar and twitter4j-stream-4.0.4.jar ✔

    (http://mvnrepository.com/artifact/org.twitter4j/twitter4j-core/4.0.4 and http://mvnrepository.com/artifact/org.twitter4j/twitter4j-stream/4.0.4)
  - Put these jars in the CLASSPATH or add the jars in the CLASSPATH variable ✔
- Create New Twitter App and get Access tokens

Downloading Tweet Streams

- Go to https://apps.twitter.com/ and click the “Create New App” button



Application Management

Twitter Apps

Create New App

Ayan Bandyopadhyay
for research

TREC2015_ISIKol_AYAN
for reasearch

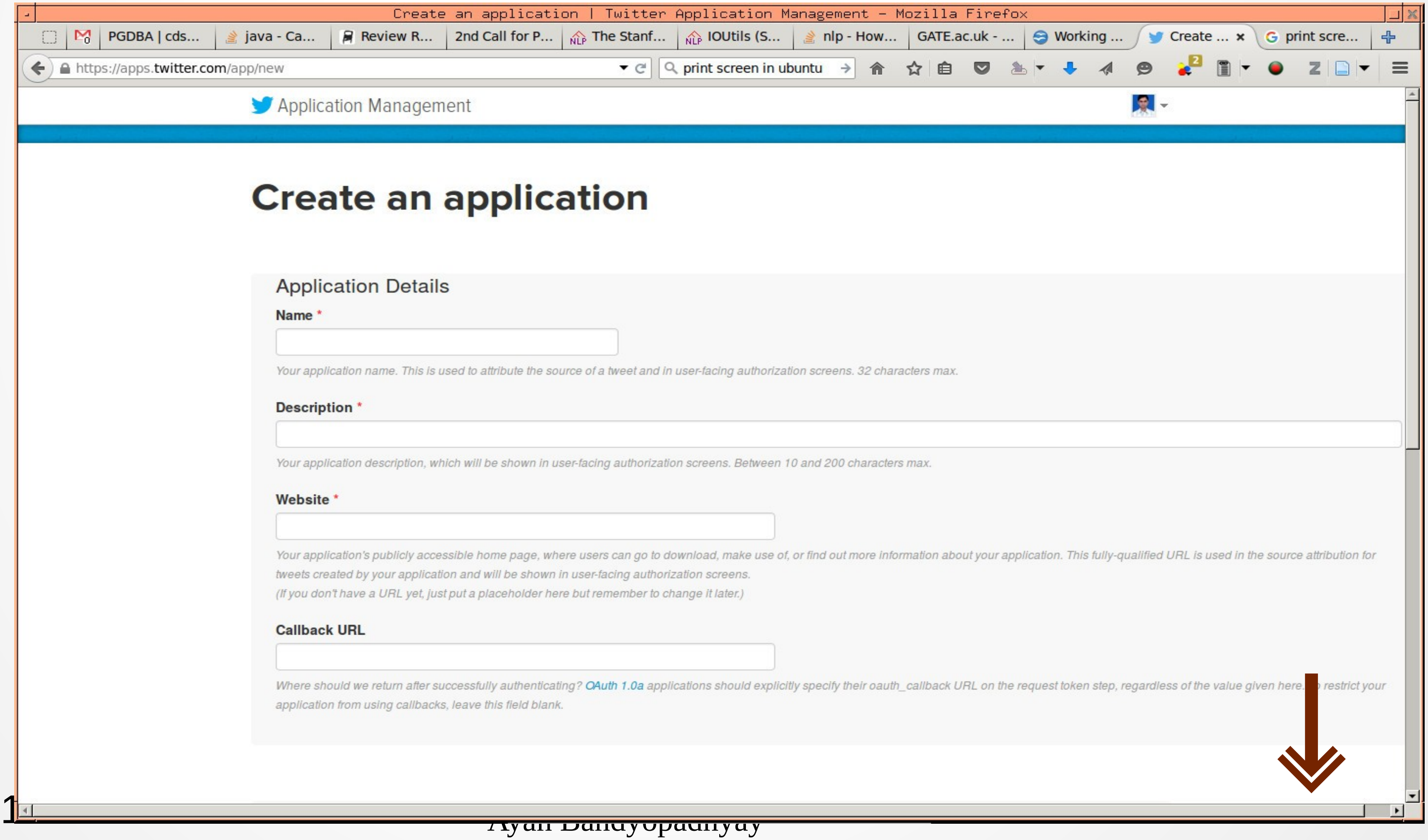
TREC2015_ISIKol_AYAN_cisco
for reasearch

TwitterFourJ
Twitter4j installation

© 2015 Twitter, Inc.

Downloading Tweet Streams

- Fill up the text fields and scroll down



Create an application

Application Details

Name *

Your application name. This is used to attribute the source of a tweet and in user-facing authorization screens. 32 characters max.

Description *

Your application description, which will be shown in user-facing authorization screens. Between 10 and 200 characters max.

Website *

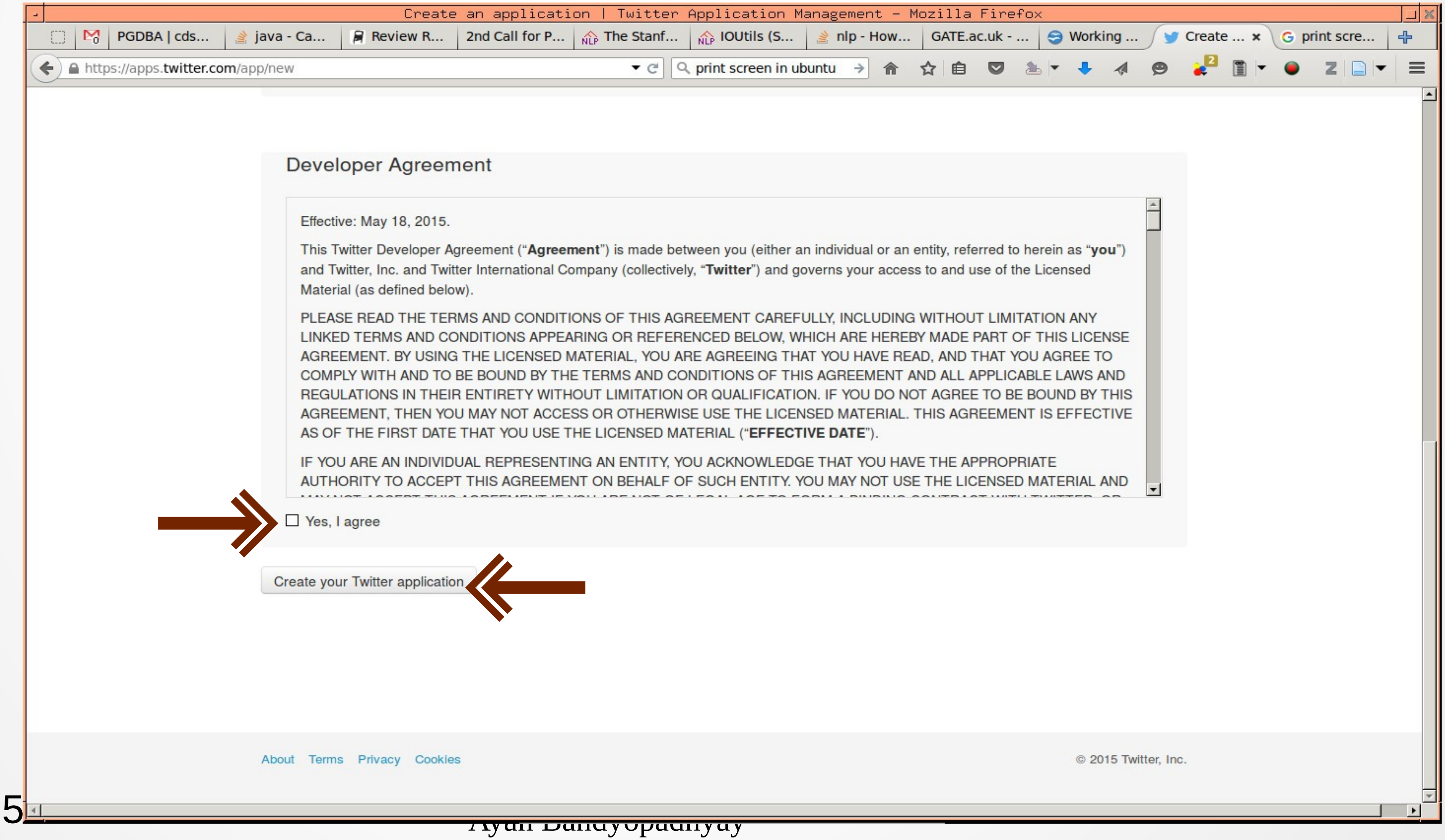
Your application's publicly accessible home page, where users can go to download, make use of, or find out more information about your application. This fully-qualified URL is used in the source attribution for tweets created by your application and will be shown in user-facing authorization screens.
(If you don't have a URL yet, just put a placeholder here but remember to change it later.)

Callback URL

Where should we return after successfully authenticating? OAuth 1.0a applications should explicitly specify their oauth_callback URL on the request token step, regardless of the value given here. To restrict your application from using callbacks, leave this field blank.

Downloading Tweet Streams

- Agree to the license agreement and then Click “Create your Twitter Application” Button



Developer Agreement

Effective: May 18, 2015.

This Twitter Developer Agreement (“**Agreement**”) is made between you (either an individual or an entity, referred to herein as “**you**”) and Twitter, Inc. and Twitter International Company (collectively, “**Twitter**”) and governs your access to and use of the Licensed Material (as defined below).

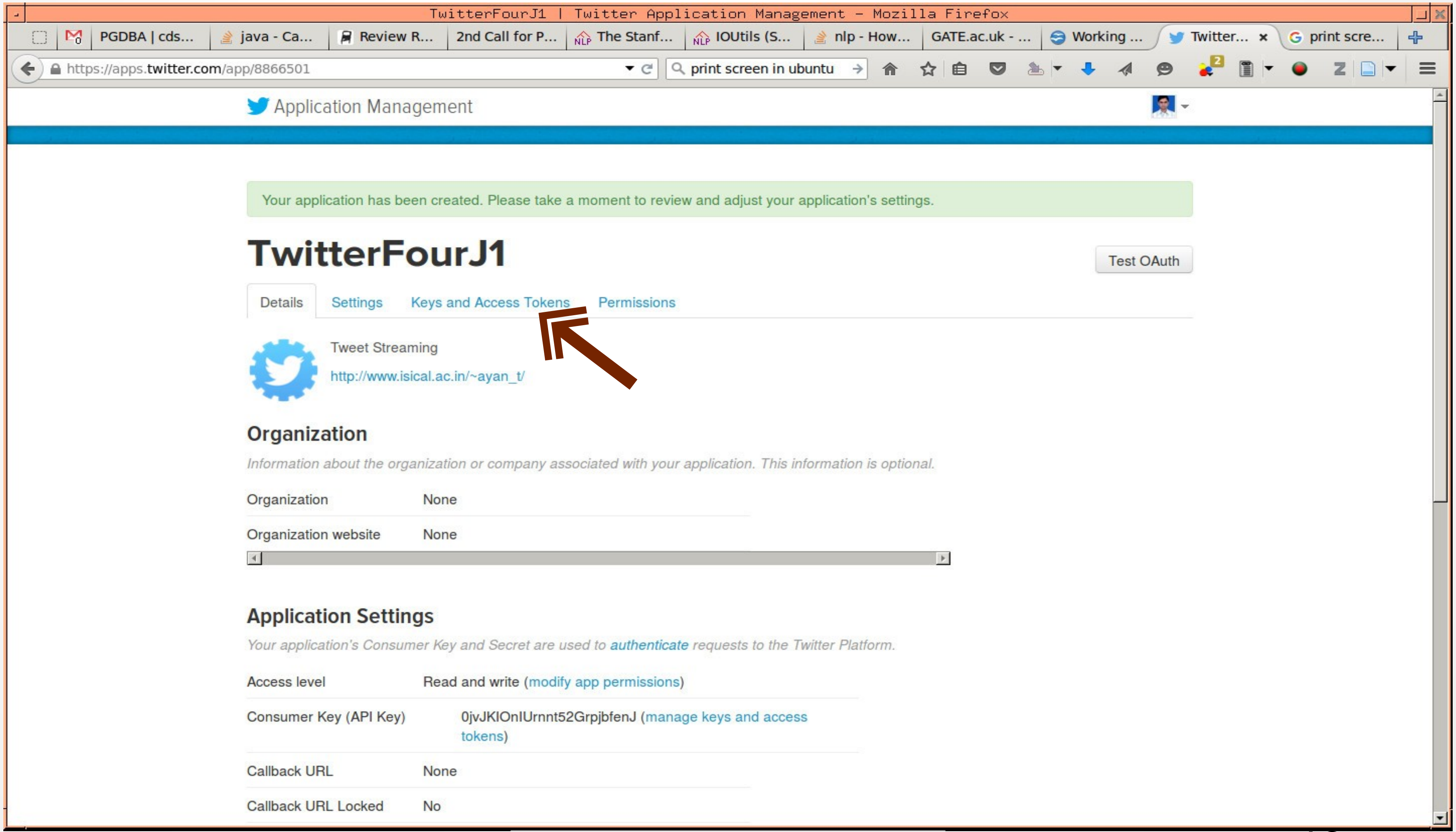
PLEASE READ THE TERMS AND CONDITIONS OF THIS AGREEMENT CAREFULLY, INCLUDING WITHOUT LIMITATION ANY LINKED TERMS AND CONDITIONS APPEARING OR REFERENCED BELOW, WHICH ARE HEREBY MADE PART OF THIS LICENSE AGREEMENT. BY USING THE LICENSED MATERIAL, YOU ARE AGREEING THAT YOU HAVE READ, AND THAT YOU AGREE TO COMPLY WITH AND TO BE BOUND BY THE TERMS AND CONDITIONS OF THIS AGREEMENT AND ALL APPLICABLE LAWS AND REGULATIONS IN THEIR ENTIRETY WITHOUT LIMITATION OR QUALIFICATION. IF YOU DO NOT AGREE TO BE BOUND BY THIS AGREEMENT, THEN YOU MAY NOT ACCESS OR OTHERWISE USE THE LICENSED MATERIAL. THIS AGREEMENT IS EFFECTIVE AS OF THE FIRST DATE THAT YOU USE THE LICENSED MATERIAL (“**EFFECTIVE DATE**”).

IF YOU ARE AN INDIVIDUAL REPRESENTING AN ENTITY, YOU ACKNOWLEDGE THAT YOU HAVE THE APPROPRIATE AUTHORITY TO ACCEPT THIS AGREEMENT ON BEHALF OF SUCH ENTITY. YOU MAY NOT USE THE LICENSED MATERIAL AND MAY NOT ACCEPT THIS AGREEMENT IF YOU ARE NOT OF LEGAL AGE TO FORM A BINDING CONTRACT WITH TWITTER, OR

☐ Yes, I agree

Downloading Tweet Streams

- Click “Keys and Access Token” tab


TwitterFourJ1 | Twitter Application Management - Mozilla Firefox
https://apps.twitter.com/app/8866501
print screen in ubuntu
Application Management

Your application has been created. Please take a moment to review and adjust your application's settings.

TwitterFourJ1

Test OAuth

Details | Settings | Keys and Access Tokens | Permissions

Tweet Streaming
http://www.isical.ac.in/~ayan_t/

Organization

Information about the organization or company associated with your application. This information is optional.

Organization	None
Organization website	None

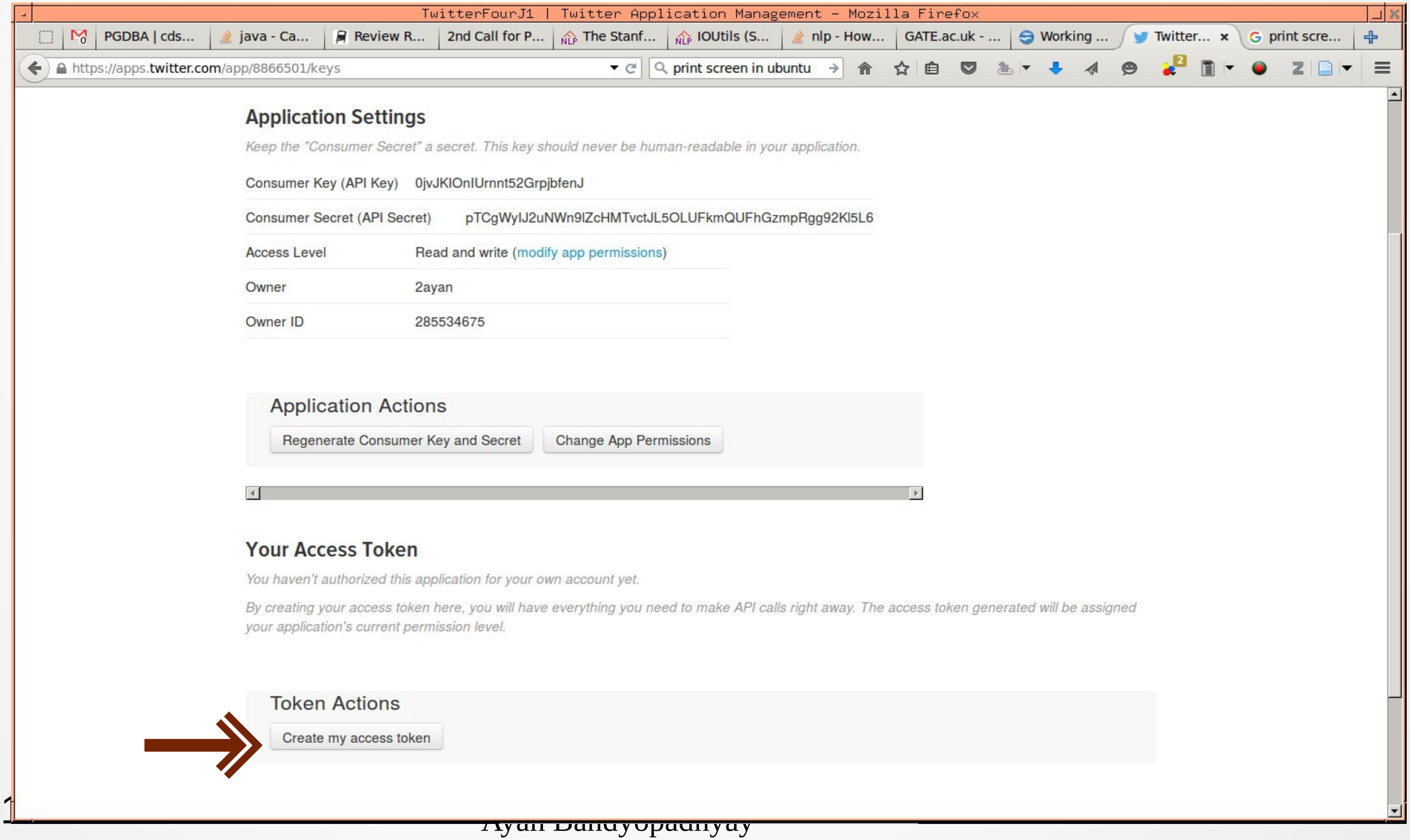
Application Settings

Your application's Consumer Key and Secret are used to authenticate requests to the Twitter Platform.

Access level	Read and write (modify app permissions)
Consumer Key (API Key)	0jvJKIOnIUrnnt52GrpjbfenJ (manage keys and access tokens)
Callback URL	None
Callback URL Locked	No

Downloading Tweet Streams

- Scroll down and Click on the “Create my Access Token” button



Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	0jvJKIOnIUrnnt52GrpjbfenJ
Consumer Secret (API Secret)	pTCgWyIJ2uNWn9lZcHMTvctJL5OLUFkmQUFhGzmpRgg92Kl5L6
Access Level	Read and write (modify app permissions)
Owner	2ayan
Owner ID	285534675

Application Actions

Your Access Token

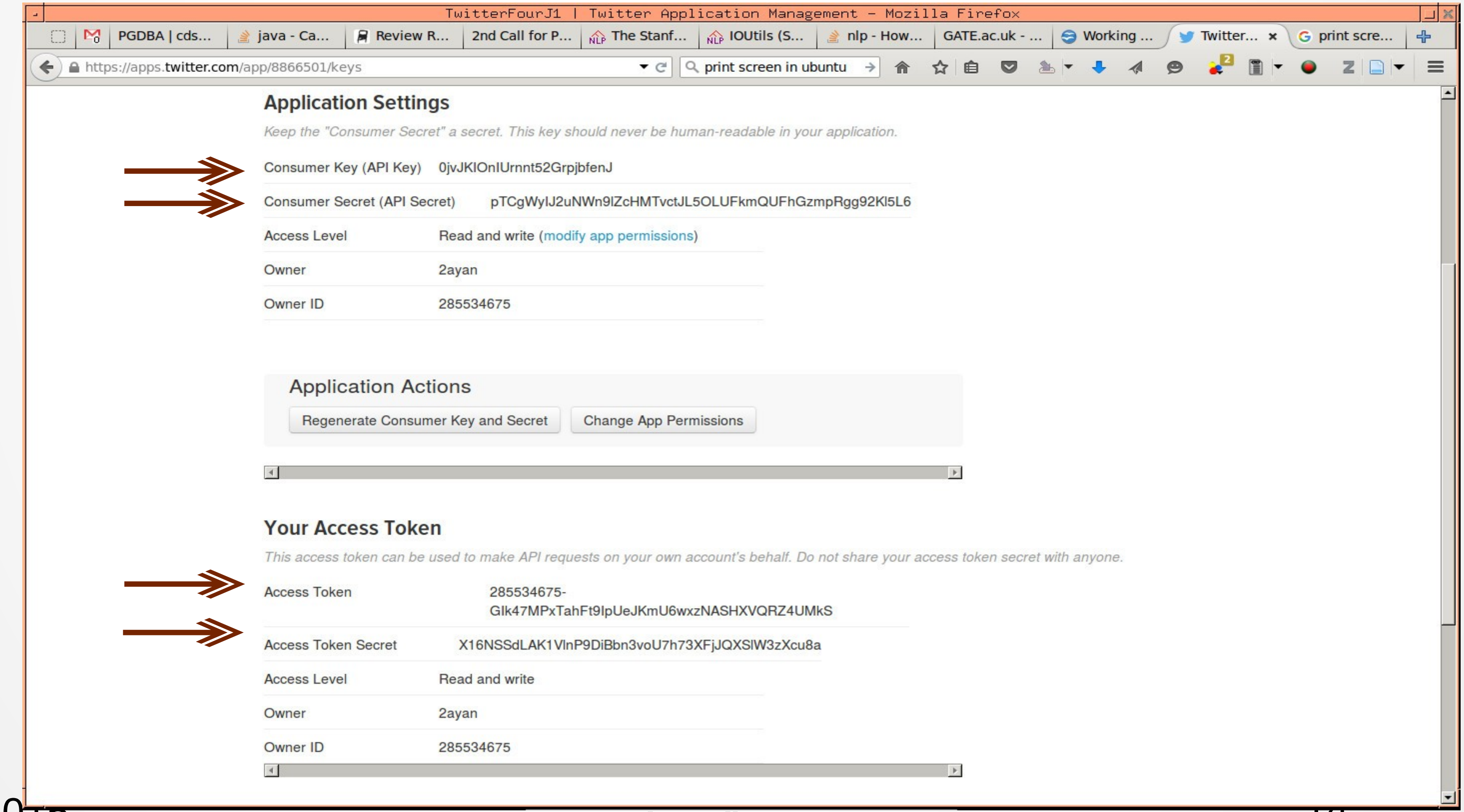
You haven't authorized this application for your own account yet.

By creating your access token here, you will have everything you need to make API calls right away. The access token generated will be assigned your application's current permission level.

Token Actions

Downloading Tweet Streams

- Finally you got the tokens.


TwitterFourJ1 | Twitter Application Management - Mozilla Firefox
https://apps.twitter.com/app/8866501/keys
print screen in ubuntu

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key)	0jvJKIOnIUrnnt52GrpjbfenJ
Consumer Secret (API Secret)	pTCgWyIJ2uNWn9lZcHMTvctJL5OLUFkmQUFhGzmpRgg92Kl5L6
Access Level	Read and write (modify app permissions)
Owner	2ayan
Owner ID	285534675

Application Actions

Regenerate Consumer Key and Secret
Change App Permissions

Your Access Token

This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

Access Token	285534675-GIk47MPxTahFt9IpUeJKmU6wxzNASHXVQRZ4UMkS
Access Token Secret	X16NSSdLAK1VlnP9DiBbn3voU7h73XFjJQXSlW3zXcu8a
Access Level	Read and write
Owner	2ayan
Owner ID	285534675

Downloading Tweet Streams

- Necessary steps
  - Download twitter4j-core-4.0.4.jar and twitter4j-stream-4.0.4.jar ✔

    (http://mvnrepository.com/artifact/org.twitter4j/twitter4j-core/4.0.4 and http://mvnrepository.com/artifact/org.twitter4j/twitter4j-stream/4.0.4)
  - Put these jars in the CLASSPATH or add the jars in the CLASSPATH variable ✔
  - Create New Twitter App and get Access tokens ✔
  - Save these Access tokens in the init.properties file

Downloading Tweet Streams

- Save these Access tokens in the init.properties file

  -Create a file init.properties and save tokens in against the following four variables

```
consumerKey=

consumerSecret=

accessToken=

accessSecret=
```

Downloading Tweet Streams

- Necessary steps
  - Download twitter4j-core-4.0.4.jar and twitter4j-stream-4.0.4.jar ✔

    (http://mvnrepository.com/artifact/org.twitter4j/twitter4j-core/4.0.4 and http://mvnrepository.com/artifact/org.twitter4j/twitter4j-stream/4.0.4)
  - Put these jars in the CLASSPATH or add the jars in the CLASSPATH variable ✔
  - Create New Twitter App and get Access tokens ✔
  - Save these Access tokens in the init.properties file ✔

Downloading Tweet Streams

- Downloading Tweet Streams
  - Basic requirement ✔
  - Necessary steps ✔
  - Write Java code

Downloading Tweet Streams

- Write Java code
  - Use methods, objects from the 2 Twitter4j jar files, + saved access tokens in your Java code

Downloading Tweet Streams

- You can use this project
  - Download zipped project from https://goo.gl/2DPXXs
  - Extract the project
    - cd <Where_project_was_downloaded>
    - unzip TweetStream.zip
  - If you want to use any integrated development environment (IDE)
    - Import project into IDE.
    - Add jar file into the project.

Downloading Tweet Streams

- Import libraries and declare variables
- Read Access tokens from init.properties
- Configure Access tokens using ConfigurationBuilder object
- Configure StatusListener object
- Configure TwitterStream object
- Write the main() function
- Most frequent/possible problems

Downloading Tweet Streams

- Import libraries and declare variables

```
import twitter4j.*;
import twitter4j.conf.*;
import java.io.*;
import java.util.*;
```

Downloading Tweet Streams

- Declare four Access token variables. These variables will store access tokens

```
String consumerKey;
String consumerSecret;
String accessToken;
String accessSecret;
```

- And need to declare three more objects

```
ConfigurationBuilder cb;
StatusListener listener;
TwitterStream twitterStream;
```

Downloading Tweet Streams

- Import libraries and declare variables ✔
- Read Access tokens from init.properties

Downloading Tweet Streams

```
public void getValuesfromPropFile(String propFile) throws IOException {
        prop = new Properties();
        prop.load(new FileReader(propFile));

        consumerKey = prop.getProperty("consumerKey");
        consumerSecret = prop.getProperty("consumerSecret");
        accessToken = prop.getProperty("accessToken");
        accessSecret = prop.getProperty("accessSecret");
}
```

propFile is the path to the init.properties file (e.g.).

Downloading Tweet Streams

- Import libraries and declare variables ✔
- Read Access tokens from init.properties using java project ✔
- Configure Access tokens by ConfigurationBuilder object

Downloading Tweet Streams

```
public void ConfigurationBuild() throws Exception {
        cb = new ConfigurationBuilder();

        cb.setDebugEnabled(true);  //optional
        cb.setJSONStoreEnabled(true);

        cb.setOAuthConsumerKey(consumerKey);
        cb.setOAuthConsumerSecret(consumerSecret);
        cb.setOAuthAccessToken(accessToken);
        cb.setOAuthAccessTokenSecret(accessSecret);
    }
```

Downloading Tweet Streams

- Import libraries and declare variables ✔
- Read Access tokens from init.properties using java project ✔
- Configure Access tokens by ConfigurationBuilder object ✔
- Configure StatusListener object

Downloading Tweet Streams

```
public void StatusListenerConfigure() {
        listener = new StatusListener() {

            @Override
            public void onException(Exception arg0) {
                // TODO Auto-generated method stub
            }

            @Override
            public void
onDeletionNotice(StatusDeletionNotice arg0) {
                // TODO Auto-generated method stub

            }

            @Override
            public void onStallWarning(StallWarning
warning) {
            }
@Override
            public void onScrubGeo(long arg0, long arg1)
{
                // TODO Auto-generated method stub

            }

            @Override
            public void onTrackLimitationNotice(int arg0) {
                // TODO Auto-generated method stub
            }

@Override
            public void onStatus(Status status) {
                printStatus(status);
            }


        };
    }
```

Downloading Tweet Streams

```
   private void printStatus(Status status) {
      // to print Json format output.
      System.err.println(TwitterObjectFactory.getRawJSON(status));
 }
```

Downloading Tweet Streams

- Import libraries and declare variables ✔
- Read Access tokens from init.properties using java project ✔
- Configure Access tokens by ConfigurationBuilder object ✔
- Configure StatusListener object ✔
- Configure TwitterStream object ✔

Downloading Tweet Streams

```
void listenToStream() throws Exception {
   myTwitterStream = new  TwitterStreamFactory(cb.build()).getInstance();
    myTwitterStream.addListener(myListener);


    // sample() method internally creates a thread which monitors myTwitterStream
and calls the registered listener methods (myListener in this case) whenever a Tweet
is received.
    myTwitterStream.sample();
}
```

Downloading Tweet Streams

- Import libraries and declare variables ✔
- Read Access tokens from init.properties using Java project ✔
- Configure Access tokens by ConfigurationBuilder object ✔
- Configure StatusListener object ✔
- Configure TwitterStream object ✔
- Write the main() function

Downloading Tweet Streams

main() function is like ..

```
public static void main(String[] args) {
    try {
         TweetStream TS = new TweetStream();

        TS.getValuesfromPropFile("/c/TwitterStream/etc/init.properties");
        TS.ConfigurationBuild();
        TS.StatusListenerConfigure();
        TS.listenToStream();

    } catch (Exception ex) {
        ex.printStackTrace();
    }
}
```

Downloading Tweet Streams

- Import libraries and declare variables ✔
- Read Access tokens from init.properties using Java project ✔
- Configure Access tokens by ConfigurationBuilder object ✔
- Configure StatusListener object ✔
- Configure TwitterStream object ✔
- Write the main() function ✔
- Most frequent/possible problems

Downloading Tweet Streams

- Most frequent/possible problems
  - Access token problem/tokens not valid

    Try creating another set of access tokens
  - Time out problem

    Due to Internet speed and availability time out occurs.

    In this case you have to do nothing but wait.

Downloading Tweet Streams

- Import libraries and declare variables ✔
- Read Access tokens from init.properties using Java project ✔
- Configure Access tokens by ConfigurationBuilder object ✔
- Configure StatusListener object ✔
- Configure TwitterStream object ✔
- Write the main() function ✔
- Most frequent/possible problems ✔

Downloading Tweet Streams

- Downloading Tweet Streams
  - Basic requirement ✔
  - Necessary steps ✔
  - Write Java code ✔

Downloading Tweet Streams and Parsing

- Downloading Tweet Streams ✔
- Tweet Parsing (JSON format)

Downloading Tweet Streams and Parsing

- Tweet Parsing (JSON format)

Tweet Parsing (JSON format)

- What is **JSON** format?
  - **JSON**, or **JavaScript Object Notation**, is a minimal, readable format for structuring data.
  - used primarily to transmit data between a server and web application.
  - alternative to XML.

Tweet Parsing (JSON format)

- Structure of JSON is represented as
  - **Key/Value Pair:** A key value pair follows a specific syntax, with the key followed by a colon followed by the value. Key/value pairs are comma separated.
  - **Key:** A key is always a string enclosed in quotation marks.
  - **Value:** A value can be a string, number, boolean expression, array, or object.
  - **Starts With:** “{”
  - **Ends With:** ”}”
- Example

```
{"collection" : {
 "title" : "Blog",
  "categories" : [ "Category-1", "Category-2" ]
}}
```

Downloading Tweet Streams and Parsing

- Sample JSON format tweet

{"in_reply_to_status_id_str":null,"in_reply_to_status_id":null,"created_at":"Wed Sep 30 09:47:45 +0000 2015","in_reply_to_user_id_str":null,

……………………….

"location":null,"profile_sidebar_fill_color":"DDEEF6","notifications":null}

Tweet Parsing (JSON format)

```
import twitter4j.JSONException;

import twitter4j.JSONObject;


public String parseJSON(String JSONstring2parse, String tag2search)
throws JSONException

    {

        JSONObject myJObject=new JSONObject(JSONstring2parse);

        return myJObject.getString(tag2search);

    }
```

Downloading Tweet Streams and Parsing

- Downloading Tweet Streams ✔
- Tweet Parsing (JSON format) ✔