Assignment 4
Tweet Analysis from HTTP Request
EE422C—The University of Texas at Austin—Fall 2019

Overview– Build a toolbox that can help analyze tweets from a server.

Problem 1: Extracting Tweets from a Server using HTTP Request

Hyptertext Transfer Protocol (HTTP) is an application level protocol for distributed hypermedia information system. This is the foundation for data communication for the internet since the 1990.

This part of the lab will help you understand the following:

- ☐ GET Response from a Web Server
- ☐ De-serialize JSON into its Object Representation

To Do:

In TweetReader.java, Implement List<Tweets> readTweetsFromWeb(URL url). This method will return a list of Tweets.

Our GET url endpoint is: http://kevinstwitterclient2.azurewebsites.net/api/products

Here is an html view: http://kevinstwitterclient2.azurewebsites.net/Views/Product/TwitterCrud.html

Each Tweet will contain:
int Id: *Id will be a number from 1 to 2^32*
String Name: *name only contains a-zA-Z0-9 and _ everything else is invalid*
String Date: *represented in Java parseable format i.e YYYY-MM-DDT00:00:00Z*
String Text: *<= 140 characters*

Note: We CAN have dirty data. If the data does not match the specs, use either exception handling or string comparison to ignore the data.

Feel free to modify any of the classes, add any additional variables, etc.. Do not modify existing method Headers. We will use this to test your code.

FAQ:

1. How do I send an HTTP Request?
https://www.mkyong.com/java/how-to-send-http-request-getpost-in-java/

2. HTTP Request returns a JSON, how do I make that into a class?
https://www.mkyong.com/java/how-to-convert-java-object-to-from-json-jackson/

Install all 3 Jackson Libraries Here
https://mvnrepository.com/search?q=jackson
Note, you can use any deserialization library you want, we recommend Jackson.

3. Will there be any dirty data? i.e. improperly formatted ID, Name, Date, and Text?
   Yes. We are crowdsourcing test cases. (See Below).

Problem 2: Analyzing Tweets

Now that you've extracted the list of tweets, we want to implement common filtering techniques:

This part of the lab will help you understand the following:

- ☐ How to use Java 8's Instant library for DateTime Comparison
- ☐ More practice with string comparison and parsing.
- ☐ Exception Handling

To Do:

Filter Tweets by UserName using the method writtenBy()
Filter Tweets by TimeSpan using the method inTimespan()
Filter Tweets by keywords using the method containing()

If the Timespan's **end** is **before** the start, throw an "InvalidTimespan Exception" with the message "End Date before Start Date"

Also, the dates might likely be formatted incorrectly to where it can not be parsed. Use Exception Handling to prevent your code from failing.

FAQ:

1. Is containing() case insensitive?

   Yes. Turn all tweets and search fields to lowercase when searching.

2. How do I compare strings represented as dates?

   Use Java 8's Instant library
   https://docs.oracle.com/javase/8/docs/api/java/time/Instant.html

3. How do I handle the exception?

   https://www.webucator.com/how-to/how-create-an-exception-class-java.cfm

Problem 3: Inferring Social Network

There are some interesting insights we can gather from twitter data. For example, we can infer social networks.

This part of the lab will help you understand the following:

- ☐ Thinking algorithmically
- ☐ Understanding NP Hard Algorithms

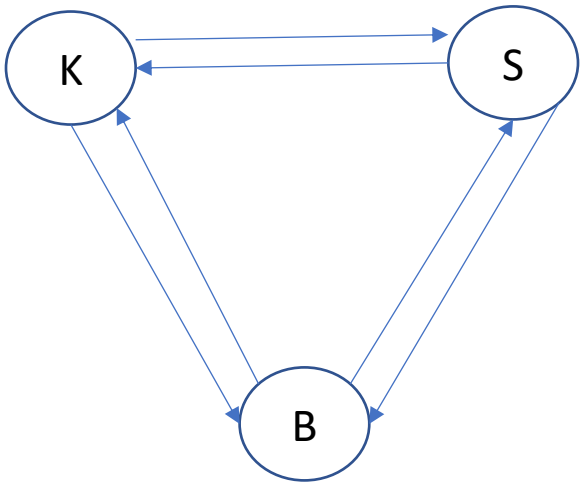
To Do:

Find a list of all cliques.
Find the top k most followed users.

A clique is defined as a set of people that are mutually friends with each other.

Look at the following example:

User Name	**Date**	**Text**
Ethomaz	11/6/2017	**@KevinYee, @BenFu, @ShrikarMurthy**. This lab is due next week. It's hard, start early
Kevinyee	11/12/2017	**@BenFu, @ShrikarMurthy** Can you help me on number 1? I just started.
Benfu	11/12/2017	**@ShrikarMurthy, @KevinYee** I can still graduate with a D right?
ShrikarMurthy	11/12/2017	**@KevinYee, @BenFu** Does the starter code run for you?



If you mention someone in a tweet, you assume you are friends with someone.
However, the friendship is not mutual unless they have mentioned you.

In the above example, Ethomaz mentioned KevinYee, BenFu, and ShrikarMurthy.
He thinks he is friends with them.
However, Kevinyee, BenFu, and ShrikarMurthy does not mention Ethomaz.

As a result, Ethomaz does not form a clique with anyone.

Kevin, Ben, and Shrikar all mutually mention each other. As a result, they form a clique.

Modify *List<List<String>> findCliques()* to return a list of all cliques on twitter.

Modift *List<User> findKMostFollowers* to return a list of the most followed users

FAQ:

1. Are usernames case insensitive?

   Yes. We can assume @ethomaz is the same as @EThomaz

2. What if there are less users than the "K" in findKMostFollower?

   Return a list of all users sorted by popularity.

3. What is considered a valid twitter mention?

   Twitter mentions cannot be preceded by any valid twitter characters.
   Example:
   Invalid:
   ethomaz@utexas.edu is preceded by z and u. therefore @utexas is not a valid id.

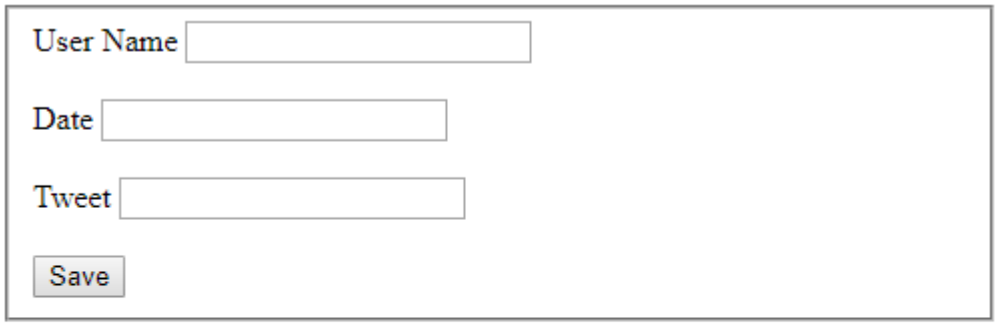
   Valid:
   Retweet @kevinyee is preceded by a space (a nonvalid character) therefore @kevinyee

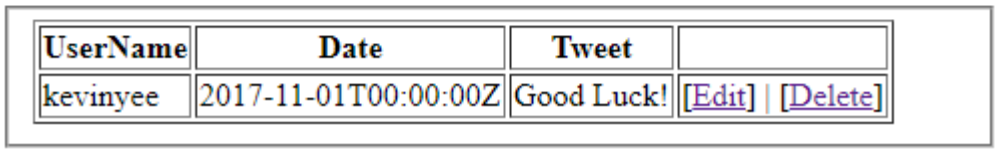
   Also Valid:

   ^@kevinyee
   &@kevinyee

Crowdsourcing Test Cases:

http://kevinstwitterclient2.azurewebsites.net/Views/Product/TwitterCrud.html





UserName	Date	Tweet	
kevinyee	2017-11-01T00:00:00Z	Good Luck!	[Edit] | [Delete]

Submit your own Tweets. You can post on Piazza the post you've added, and the expected test cases + outputs you have created to help other students out.

Note:
There is no data validation in the text boxes. i.e. you can submit anything in the User, Date, and Tweet. However, do note that these are expected to be handled in your code.

String Id: *Id will be a number from 1 to 2^32*
String Name: *name only contains a-zA-Z0-9 and _ everything else is invalid*
String Date: *represented in Java parseable format i.e YYYY-MM-DDT00:00:00Z*

String Text: <= *140 characters*

What to submit:
1. The package directory ‘assignment4’ and all Java files it contains

***** Zip these two item together and name the zip file ‘Project4_EID.zip’ *****

When unzipped, the folder structure should be:

```
Project4_EID(folder)
->assignment4(folder)
--->Tweets.java
--->Main.java
```

Instructor’s Note:

1. We have provided Tweets.java and Timespan.java, you should not need to modify these two classes. If you feel like these classes have to be modified in order to make your program work, please post on piazza and explain why a modification is needed to make your program work before proceeding.

2. We recommend making JUNIT test cases to test your code – constantly querying a web server may be extremely slow.

3. Disclaimer: Our web server was not well made. It is neither secure nor built against attacks. We understand that there are many students taking this class talented enough to hack the website and completely alter it’s contents. Please do the right thing.

4. Warning against procrastinating. If the server is down due to high traffic at 11:30 PM before the night it is due – that would be troublesome.

5. In the event the server **does** go down. You are still expected to finish the lab. Use JUNIT to test our code.

Final Checklist:

- ☐ Implement readTweetsFromWeb()
- ☐ Implement writtenBy()
- ☐ Implement inTimespan ()
- ☐ Implement containing()
- ☐ Implement findCliques()
- ☐ Implement findKMostFollowers()