

Install instructions

1. Database, I use MySQL server 5.7 and MySQL workbench 6.3

(1) Create connection.

Set Hostname: 127.0.0.1

Set Port: 3306

Set Username: root

Set Password: 31514115

Show as figure 1

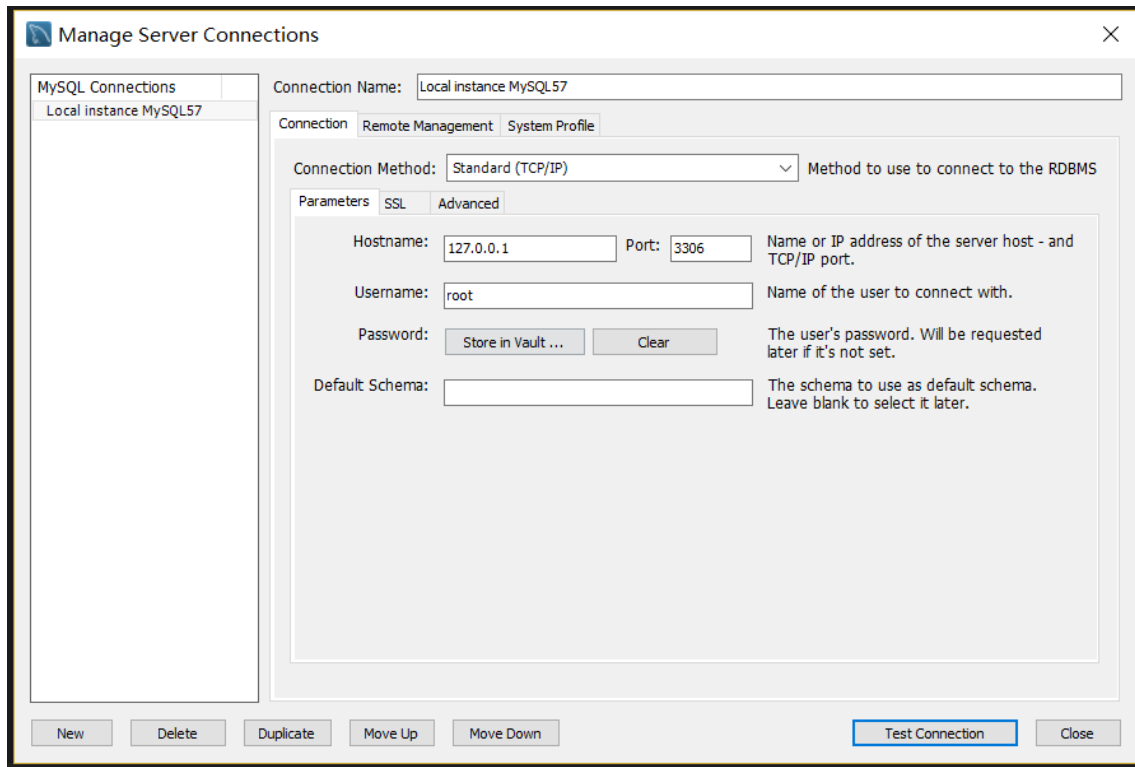


Figure 1 Settings

Click “Test Connection” on the right down corner, if it pops out a window saying it has connection has been successfully connected, and then the connection has been approved. Show as figure 2.

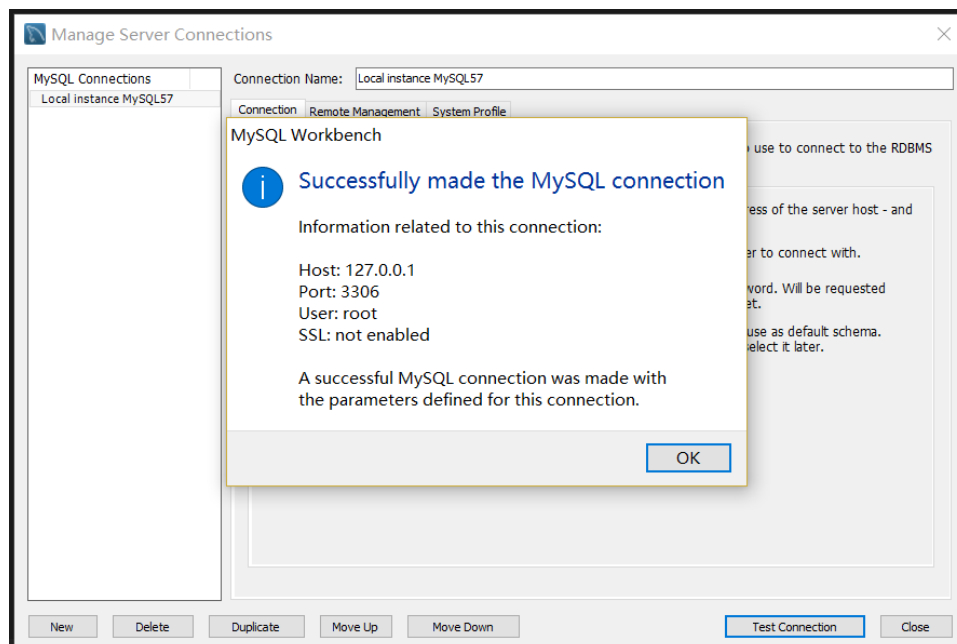
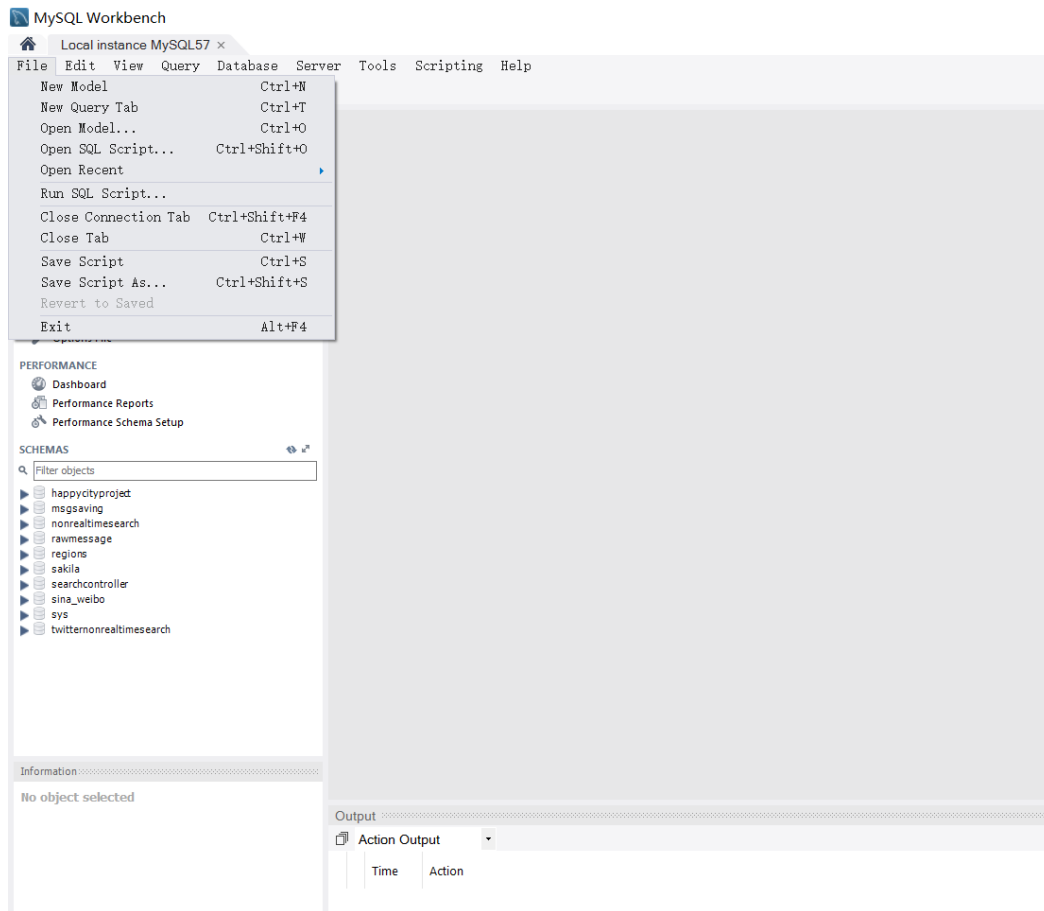


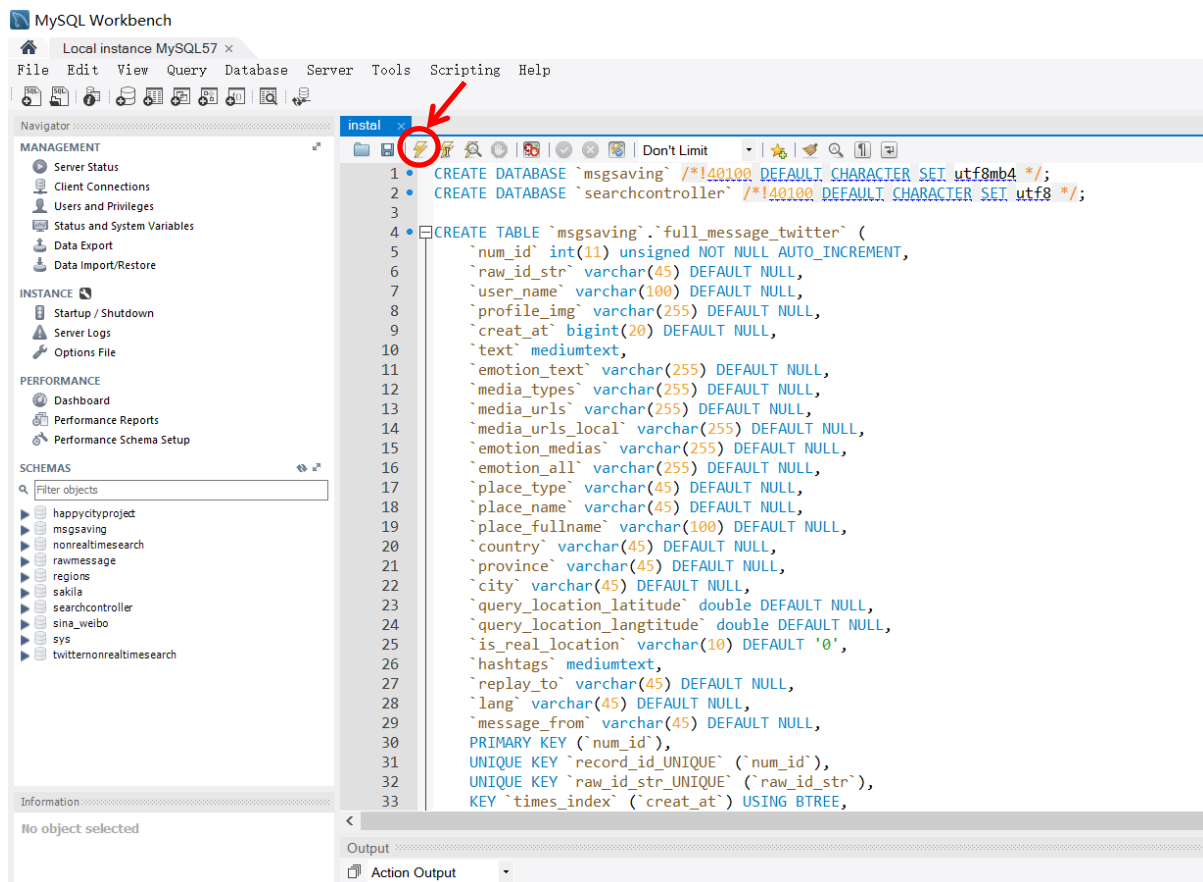
Figure 2 Connection tested successfully

(2) Open the connection and run install.sql

a) Click “File”



b) Click “Open SQL Scripts”. Load to the folder where you download the project, click “install.sql”. Click the first “flash” symbol on the top left of the install scripts window as cycled in the figure 3.



2. Run the programs

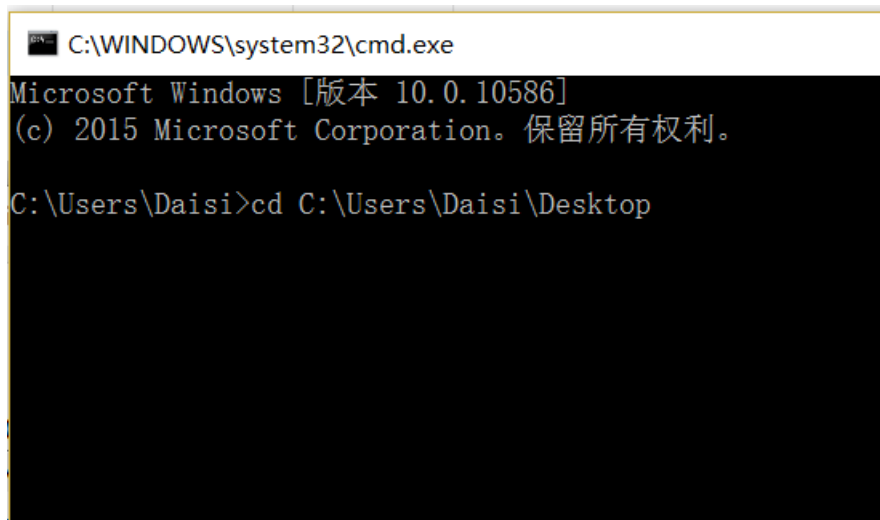
IF you prefer to run it through the source code by IDE, you can have the code and run it directly. However, the IDE should support maven and you have to do the same database operations. The following is how to run the runnable jar file exported by the source code. They are the same.

(1) Open three CMD or POWERSHELL windows

For each window:

(2) Located to the file where the three jar files located.

I put the jar files at the desktop so it will be:



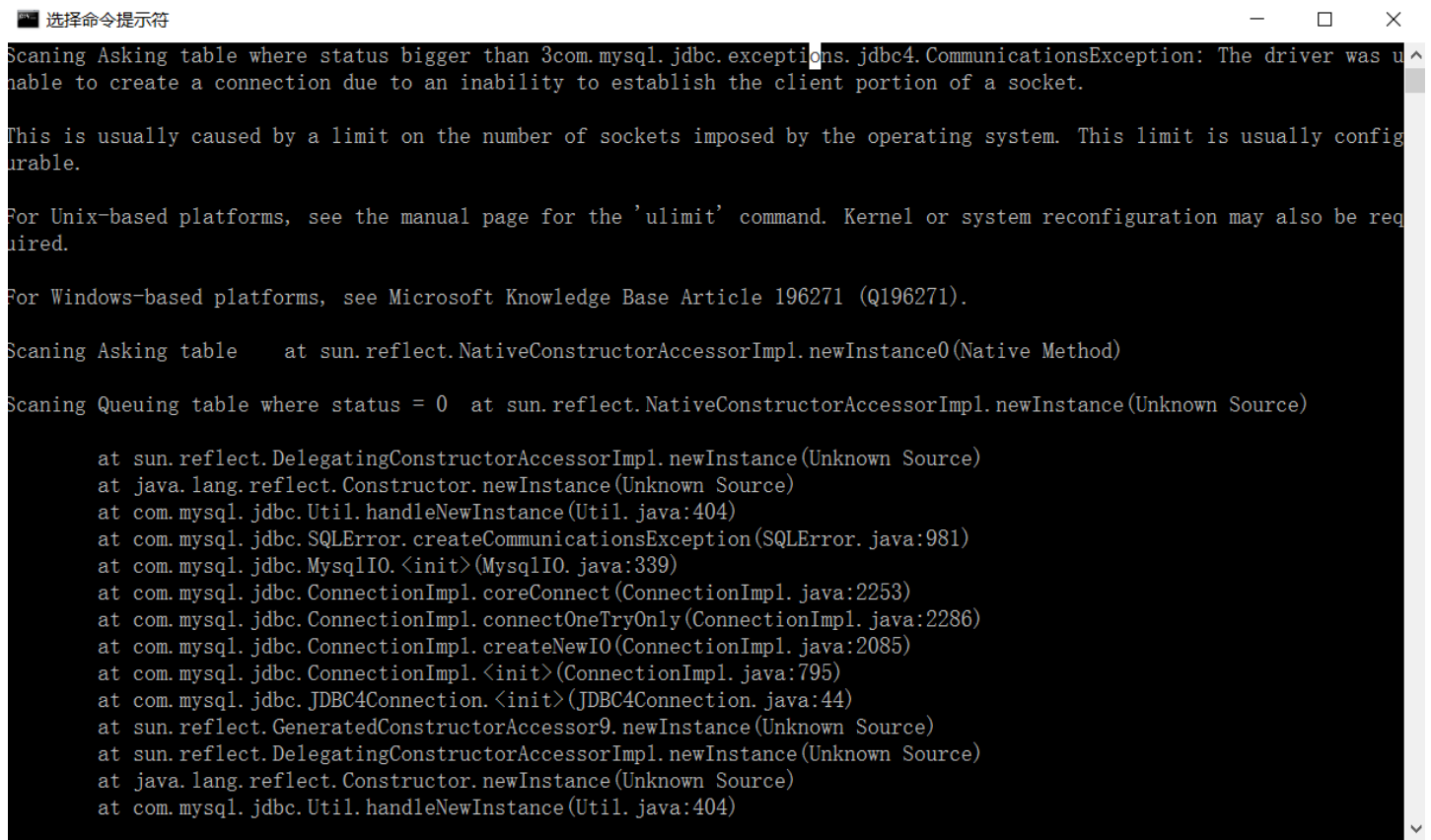
```
C:\WINDOWS\system32\cmd.exe
Microsoft Windows [版本 10.0.10586]
(c) 2015 Microsoft Corporation。保留所有权利。

C:\Users\Daisi>cd C:\Users\Daisi\Desktop
```

(3) Run the following commands on each window.

```
java -jar NonRealTimeInstagram.jar
java -jar NonRealTimeTwitterGetter.jar
java -jar RealTimeInstagramGetter.jar
```

You can run the three commands one by one otherwise you may encounter a problem shown as follow:



```
选择命令提示符
Scanning Asking table where status bigger than 3com.mysql.jdbc.exceptions.jdbc4.CommunicationsException: The driver was unable to create a connection due to an inability to establish the client portion of a socket.

This is usually caused by a limit on the number of sockets imposed by the operating system. This limit is usually configurable.

For Unix-based platforms, see the manual page for the 'ulimit' command. Kernel or system reconfiguration may also be required.

For Windows-based platforms, see Microsoft Knowledge Base Article 196271 (Q196271).

Scanning Asking table    at sun.reflect.NativeConstructorAccessorImpl.newInstance0(Native Method)
Scanning Queuing table where status = 0  at sun.reflect.NativeConstructorAccessorImpl.newInstance(Unknown Source)

    at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(Unknown Source)
    at java.lang.reflect.Constructor.newInstance(Unknown Source)
    at com.mysql.jdbc.Util.handleNewInstance(Util.java:404)
    at com.mysql.jdbc.SQLException.createCommunicationsException(SQLException.java:981)
    at com.mysql.jdbc.MysqlIO.<init>(MysqlIO.java:339)
    at com.mysql.jdbc.ConnectionImpl.coreConnect(ConnectionImpl.java:2253)
    at com.mysql.jdbc.ConnectionImpl.connectOneTryOnly(ConnectionImpl.java:2286)
    at com.mysql.jdbc.ConnectionImpl.createNewIO(ConnectionImpl.java:2085)
    at com.mysql.jdbc.ConnectionImpl.<init>(ConnectionImpl.java:795)
    at com.mysql.jdbc.JDBC4Connection.<init>(JDBC4Connection.java:44)
    at sun.reflect.GeneratedConstructorAccessor9.newInstance(Unknown Source)
    at sun.reflect.DelegatingConstructorAccessorImpl.newInstance(Unknown Source)
    at java.lang.reflect.Constructor.newInstance(Unknown Source)
    at com.mysql.jdbc.Util.handleNewInstance(Util.java:404)
```

This is because there is too many socket connections to the MySQL server. You can run the jars one by one or set the socket limits to unlimited.

3. You will find:

1). For CMD windwons:

a) For NonReslTimeInstagram

The task I pre-set for it is shown as table 1

Table 1 task for NonRealTimeInstagrram

num_id	city_name	latitude	longitude	radius	start_date	end_date	in_what_lan	required_by	required_at	message_from	getting_status
831	toronto,ON	43.708215	-79.354983	15	2016-04-01	2016-04-02	en	user	user	Instagram	0

This task means a user wants the Instagram posts of (43.708215, -79.354983) from 2016-04-01 to 2016-04-02 with radius 5 KM and it haven't started yet. The radius is fixed as Instagram API only allows 5 KM radius. So far, I didn't do the language filter function but only language detection so the "in_what_lan, required_by, required_at" are only for future works. You will find the following prints on the CMD window which includes the URL sent and the structures message. The SLF4J and log4J warns are for the prints you can ignore them.

```
命令提示符
C:\Users\Daisi\Desktop>java -jar NonRealTimeInstagram.jar
Scann asking table
Insert into queuing table
Scann asking table
Scann asking table
SLF4J: Class path contains multiple SLF4J bindings.
SLF4J: Found binding in [jar:rsrsc:slf4j-log4j12-1.7.19.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: Found binding in [jar:rsrsc:slf4j-simple-1.7.19.jar!/org/slf4j/impl/StaticLoggerBinder.class]
SLF4J: See http://www.slf4j.org/codes.html#multiple_bindings for an explanation.
SLF4J: Actual binding is of type [org.slf4j.impl.Log4jLoggerFactory]
https://api.instagram.com/v1/media/search?min_timestamp=1459655700&max_timestamp=1459656000&distance=5000&lat=43.708215&lng=-79.354983&access_token=546785387.c40b2f5.eacef90ac695483aaf8dd035be32dfbb
log4j:WARN No appenders could be found for logger (org.apache.http.client.protocol.RequestAddCookies).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
Scann asking table
StructuredFullMessage [num_id=0, raw_id_str=1220005390715302846_592307411, user_name=johnutmost, profile_img=https://scontent.cdninstagram.com/t51.2885-19/s150x150/13129880_1073682662702329_1465282148_a.jpg, creat_at=1459655997, text=Incognito, media_type=[photo], media_urls=[https://scontent.cdninstagram.com/t51.2885-15/s640x640/sh0.08/e35/12519088_1179423685401412_1939666508_n.jpg?ig_cache_key=MTIyMDAwNTM5MDcxNTMwMjg0Ng%3D%3D.2], media_urls_local=[], place_type=, place_name=Toronto, Ontario, place_fullname=, query_location_latitude=43.7166, query_location_longitude=-79.3407, is_real_location=true, hashtags=[bnw_society rising_masters blackandwhite vscofilm toronto_insta bnw_captures jj_portrait way2ill insta_daily fartoodope instafocus inspiration toronto_torontoclicks vsco_makemoreportraits mkexplore artofvisuals mask_picoftheday bnw_globe igtoronto instagood makeportraitsnotwar exporecanada agameoftones the6_photooftheday ], replay_to=, lang=it, emotion_text=, emotion_medias=[], emotion_all=, message_from=instagram]
StructuredFullMessage [num_id=0, raw_id_str=1220005370060586241_3075015494, user_name=weary_assasin, profile_img=https://scontent.cdninstagram.com/t51.2885-19/s150x150/12383405_605016502981728_901324633_a.jpg, creat_at=1459655995, text=Walking the line, media_type=[photo], media_urls=[https://scontent.cdninstagram.com/t51.2885-15/s480x480/e35/12918069_1540632999565841_1271994733_n.jpg?ig_cache_key=MTIyMDAwNTM3MDA2MDU4NjIOMQ%3D%3D.2], media_urls_local=[], place_type=, place_name=Toronto, Ontario, place_fullname=, query_location_latitude=43.7166, query_location_longitude=-79.3407, is_real_location=true, hashtags=[bnw_society rising_masters blackandwhite vscofilm toronto_insta bnw_captures jj_portrait way2ill insta_daily fartoodope instafocus inspiration toronto_torontoclicks vsco_makemoreportraits mkexplore artofvisuals mask_picoftheday bnw_globe igtoronto instagood makeportraitsnotwar exporecanada agameoftones the6_photooftheday ], replay_to=, lang=it, emotion_text=, emotion_medias=[], emotion_all=, message_from=instagram]
```

You need to wait a few minutes to have the structured message printed after "Scann asking table" as the second thread will not start very soon after the first one.

b) For RealTimeInstagramGetter

The task I pre-set for it is shown as table 2

Table 2 Task for ReslTimeInstagramGetter

num_id	city_name	latitude	longitude	radius	in_what_lan	required_by	required_at	message_from	getting_status
1	toronto	43.781213	-79.361103	5000	en	user	user	Instagram	0
2	toronto	43.671213	-79.208885	5000	en	user	user	Instagram	0
3	toronto	43.671213	-79.361103	5000	en	user	user	Instagram	0
4	toronto	43.736052	-79.361587	5000	en	user	user	Instagram	0
5	toronto	43.736052	-79.488358	5000	en	user	user	Instagram	0
6	toronto	43.644385	-79.361587	5000	en	user	user	Instagram	0
7	toronto	43.644385	-79.488358	5000	en	user	user	Instagram	0
8	toronto	43.781213	-79.208885	5000	en	user	user	Instagram	0

This task means a user wants the Instagram realtime posts for the 8 tasks. You will find the following prints on the CMD window which includes the URL sent and the messages. "{code":200, "type": "OK", "message": "Upload Success. ", "obj":null}" means this message has been sent to

the server.

```
C:\WINDOWS\system32\cmd.exe
C:\Users\Daisi>cd C:\Users\Daisi\Desktop

C:\Users\Daisi\Desktop>java -jar RealTimeInstagramGetter.jar
program started
https://api.instagram.com/v1/media/search?min_timestamp=1462820895&max_timestamp=1462821195&distance=5000&lat=43.781213&
lng=-79.361103&access_token=546785387.c40b2f5.eacef90ac695483aaf8dd035be32dfbb
log4j:WARN No appenders could be found for logger (org.apache.http.client.protocol.RequestAddCookies).
log4j:WARN Please initialize the log4j system properly.
log4j:WARN See http://logging.apache.org/log4j/1.2/faq.html#noconfig for more info.
num_id1462821195000
Happy to see you back ???
{"code":200,"type":"OK","message":"Upload Success. ","obj":null}
num_id1462821195001
Weekend Photo. Roaming around #torontoigers #torontophoto #happykid #nurselifewhendayoff #
{"code":200,"type":"OK","message":"Upload Success. ","obj":null}
num_id1462821195002
#alexisbittar #banglebracelet shopping is always a priority in life
{"code":200,"type":"OK","message":"Upload Success. ","obj":null}
num_id1462821195003
Haircut
{"code":200,"type":"OK","message":"Upload Success. ","obj":null}
Get Request sent succesfully
https://api.instagram.com/v1/media/search?min_timestamp=1462820895&max_timestamp=1462821195&distance=5000&lat=43.671213&
lng=-79.208885&access_token=546785387.beea524.59e28413250f4e8382a2658b15ab367b
num_id1462821195004
Meathead
{"code":200,"type":"OK","message":"Upload Success. ","obj":null}
Get Request sent succesfully
https://api.instagram.com/v1/media/search?min_timestamp=1462820895&max_timestamp=1462821195&distance=5000&lat=43.671213&
lng=-79.361103&access_token=546785387.ea061b4.77650fdfa9094e96a1clacae41412ef8
```

c) For NonRealTimeTwitterGetter

The task I pre-set for it is shown as table 3

Table 3 Task for ReslTimeInstagramGetter

num_id	city_name	latitude	longitude	radius	start_date	end_date	in_what_lan	required_by	required_at	message_from	getting_status
1	mississauga,ON	43.625708	-79.665695	15	2016-04-01	2016-04-01	en	user	user	Twitter	0
2	vancouver,BC	49.255216	-123.168950	15	2016-04-01	2016-04-03	en	user	user	Twitter	0

The result is as follow:

```
C:\Users\Daisi\Desktop>java -jar NonRealTimeTwitterGetter.jar
Scanning Asking table
Scanning Queuing table where status = 0
Scanning Asking table where status bigger than 3
Start crawling
Crawling url is https://twitter.com/search?f=tweets&vertical=default&q=place%3A009d493a575b8930%20since%3A2016-03-31%20u
ntil%3A2016-04-02&src=typd&lang=en
五月 09, 2016 3:55:40 下午 org.apache.http.client.protocol.ResponseProcessCookies processCookies
警告: Invalid cookie header: "set-cookie: fm=0; Expires=Mon, 09 May 2016 19:55:26 GMT; Path=/; Domain=.twitter.com; Secu
re; HTTPOnly". Invalid 'expires' attribute: Mon, 09 May 2016 19:55:26 GMT
content.size() 9
has_more_items true
Get Request sent succesfully 1
Second Crawling Url is https://twitter.com/search?f=tweets&vertical=default&q=place%3A009d493a575b8930%20since%3A2016-03-
31%20until%3A2016-04-02&src=typd&lang=en&include_available_features=1&include_entities=1&last_note_ts=1&max_position=TW
EET-715500174697754624-715720126109093888-BD1U02FFu9QAAAAAAAAETAAAAAcAAAAASAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
&reset_error_state=false&src=typd&lang=en
五月 09, 2016 3:55:42 下午 org.apache.http.client.protocol.ResponseProcessCookies processCookies
警告: Invalid cookie header: "set-cookie: fm=0; Expires=Mon, 09 May 2016 19:55:28 GMT; Path=/; Domain=.twitter.com; Secu
re; HTTPOnly". Invalid 'expires' attribute: Mon, 09 May 2016 19:55:28 GMT
content.size() 0
has_more_items false
Crawl Finished
url is https://twitter.com/search?f=tweets&vertical=default&q=near%3A%22West%20Humber-Clairville%2c%20Toronto%22%20withi
```

Again You need to wait a few minutes to have the structured message printed after “Writing into queuing table” as the second thread will not start very soon after the first one. The “Invalid cookie header” warn is because I didn’t set cookie for it as I don’t need so far.

2) For the database

a) The getting_status of the asking table will be changed to 1 if the task has been saved into queuing table. (Queuing)

The task_id tasks saved in queuing table will be the num_id of asking table. One asking column is relevant with the columns of queuing table if its num_id is the task_id of the columns in queuing table.

b) The getting_status of the asking table will be changed to 2 if one of the relevant columns has status 1. (Crawling)

c) The getting_status of the asking table will be changed to 3 if all of the relevant columns have status 2 or 9. (Finished with results)

d) The getting_status of the asking table will be changed to 2 if all of the relevant columns only have status 9. (Found no results)

(e) Message collected by the NonRealTimeInstagram.jar will be saved into full_message_ig table

(f) Message collected by the NonRealTimeTwitterGetter.jar will be saved into full_message_twitter table

(g) Message collected by the RealTimeInstagramGetter.jar will be saved into full_message_ig_realtime table

4. Make sure:

Mysql 5.7 and java 8 are installed