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

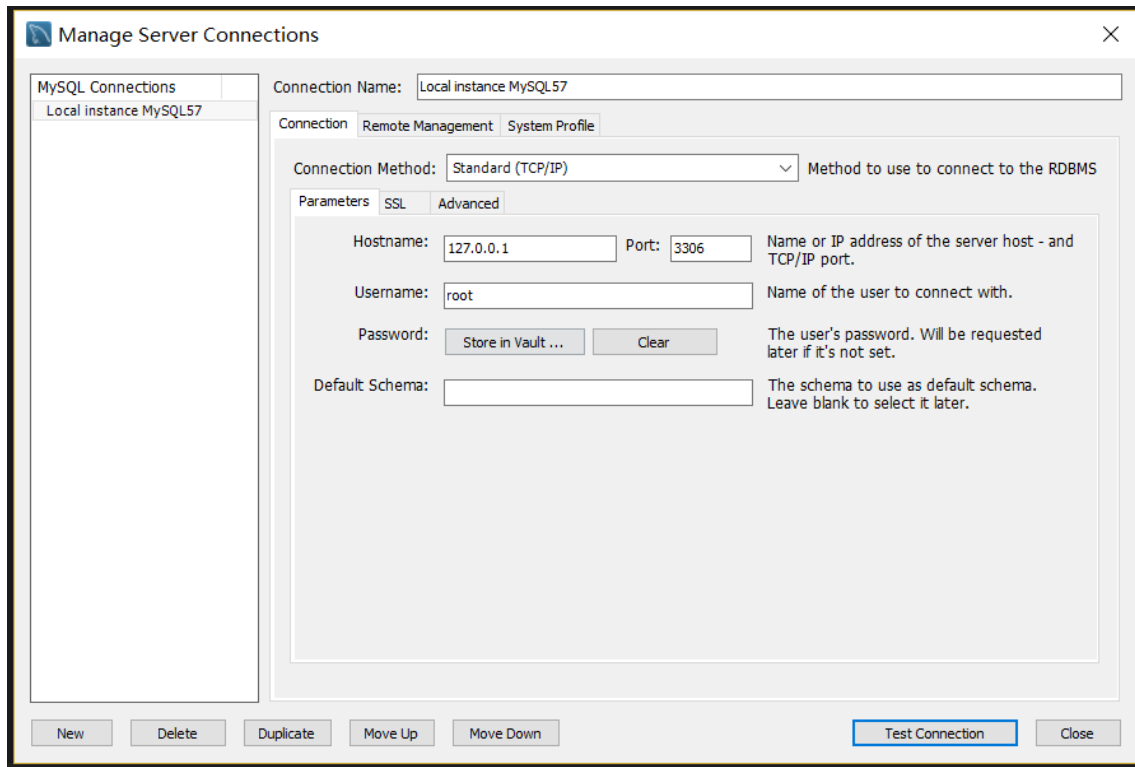Set Password: 31514115

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.
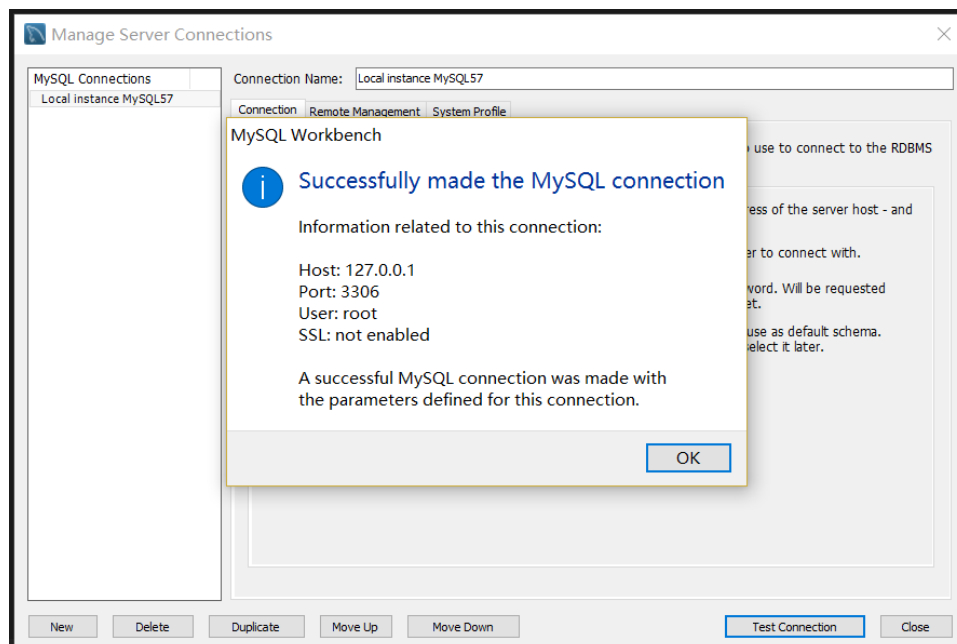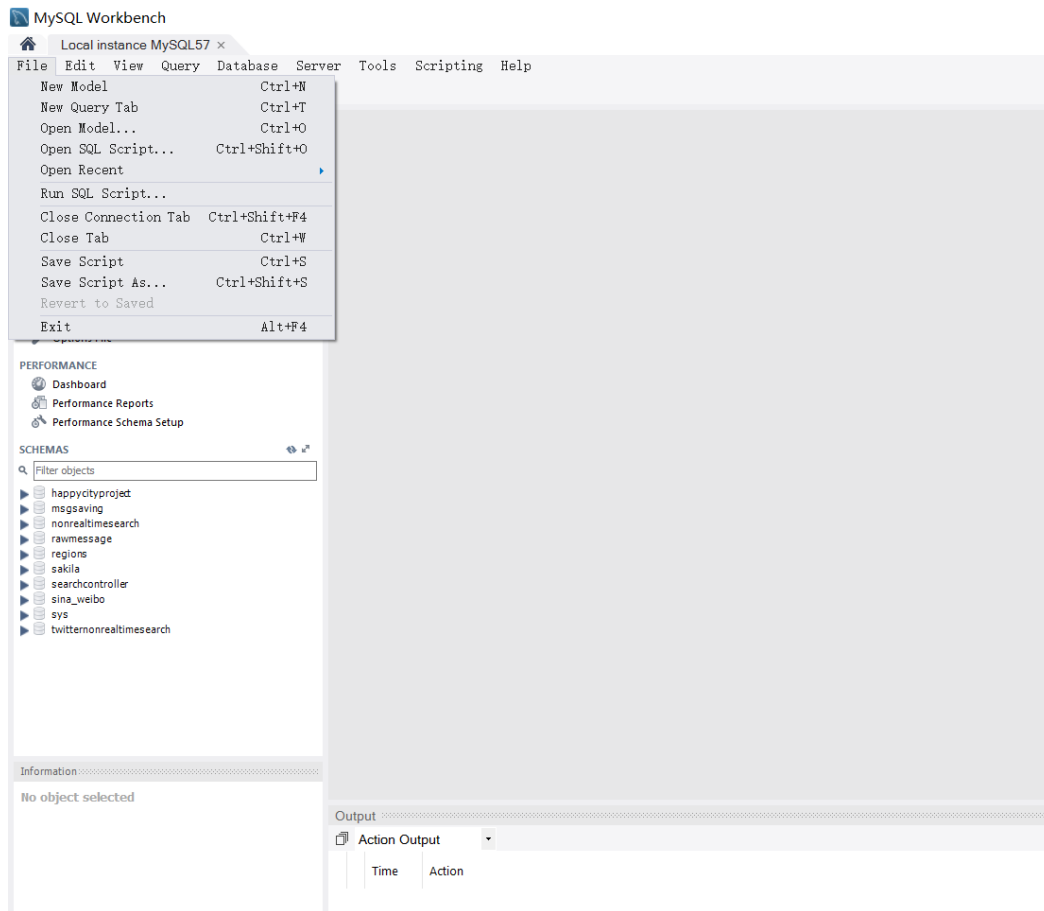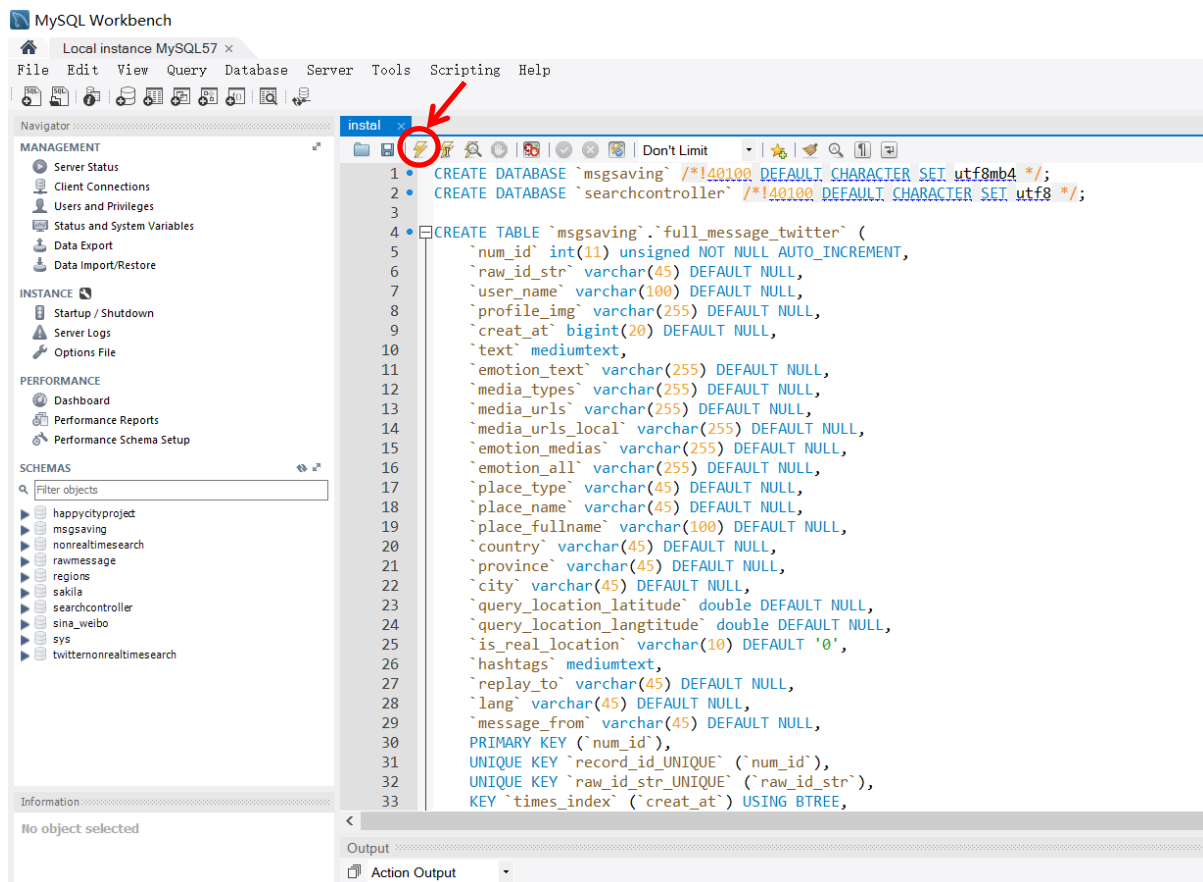

Figure 2 Connection tested successfully

(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.
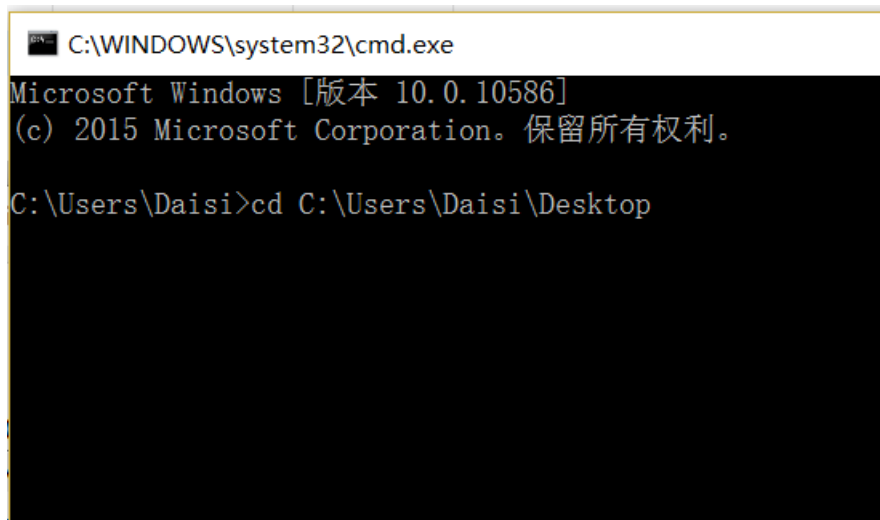
2. Run the programs

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.

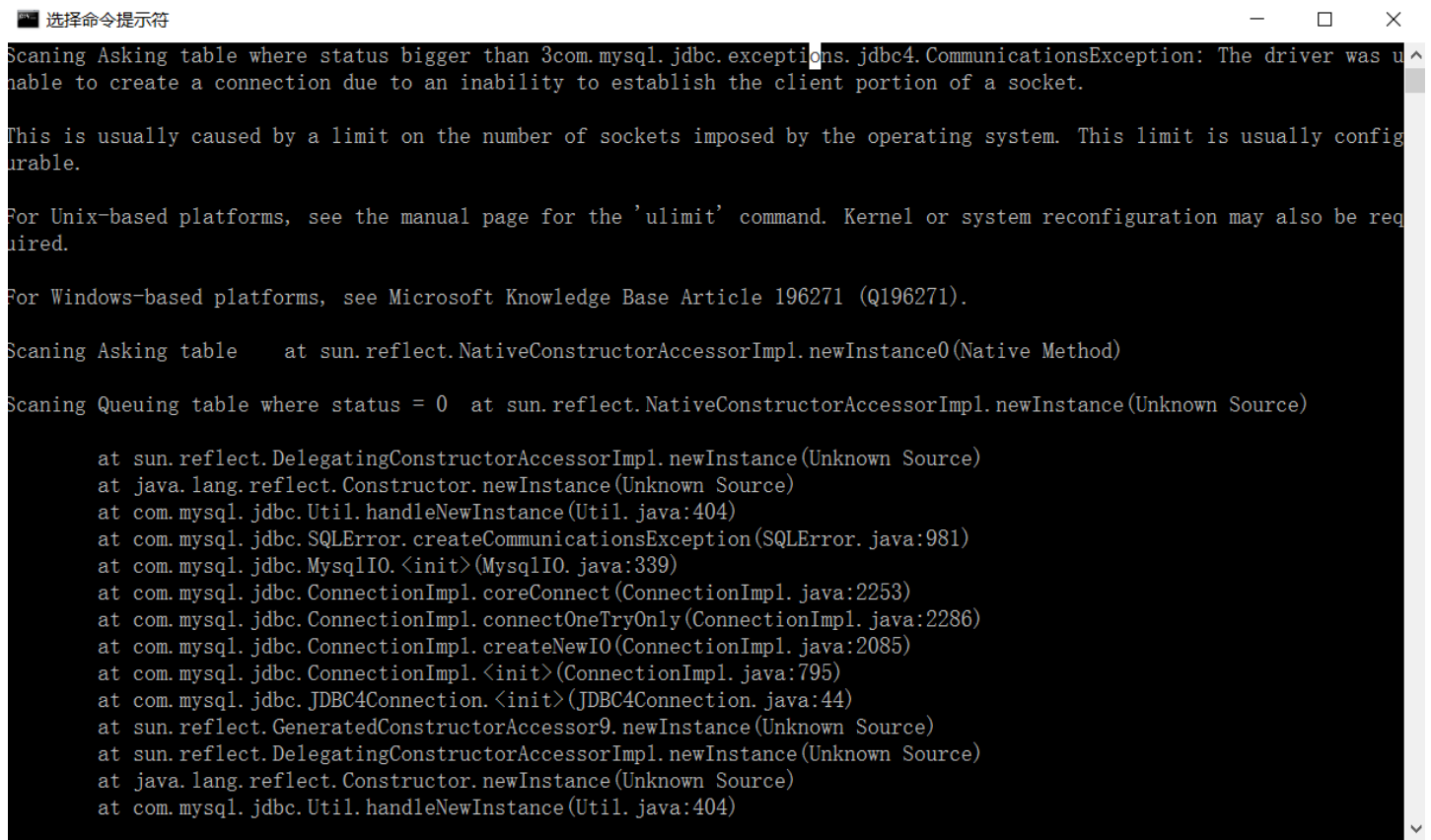I put the jar files at the desktop so it will be:

(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 commonds one by one otherwise you may encounter a problem shown as follow:

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 Instgram 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.

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 Instgram 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) 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:

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