Project 5.29. QQ Zone web scraping (a social media website)

Introduction:

The code to scrap data from this website makes use of requests, beautifulsoup, selenium, time and cookie. The difficulty is the use of selenium and cookie. Selenium has been explained in out tutorials. An **HTTP cookie** is a small piece of data sent from a website and stored on the user's computer by the user's browser while the user is browsing. It lets server programs track what pages a user has visited or what actions the user has performed. Some websites require log-in to read all data, that’s when we need a cookie in our code.

Requirement:

1. Website link: http://user.qzone.qq.com/QQ account
2. Write a code to extract the following data:
   1. Input QQ account
   2. ShuoShuo（which is similar to twitter posts）
   3. Posting time
3. Store the data in an excel
4. Bonus
   1. write PowerPoint slide to introduce the code (50/effective page)

Codes will be graded based on:

1. The format of the codes
2. The organization of the csv file
3. The speed of the codes
4. The extendibility of the codes

Estimated completion time: 20 hours. Time consumption depends on the codes’ quality.

Suggestive points:

Deadline:

Bonus points:

1. Work completion before the deadline will generate an extra 20% bonus.

Penalty:

1. Late delivery/missing deadline will incur -10% of finally determined points of the whole project per missing day.
2. Serious mistakes that made the submitted results not useful will incur -120% of the whole project.