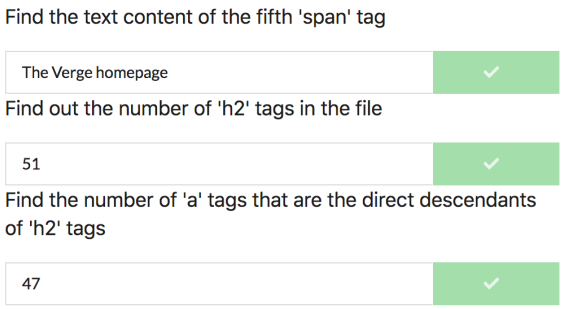
This is the online internship project for web crawler.

The Set up exercise is using requests to make request to https://api.github.com/. And then Use json to parse the request result. The last value of 'notifications_url' is s.

The warm up exercise provides a local html page and asking to do some practices on it.



All the answers has been verified and the code is included in Warmup.py.

The Improve robustness exercise is related to getting to know several anti-crawling mechanisms such as Proxy, Header and SSL certificates. These mechanisms has been learned and several requests has been made.

The Improve Performance exercise asks to use multi-thread to add from 1 to 10000000 and enter the result. I launched 100 threads to solve the problem and the answer is 50000005000000.

The Use Webdriver exercise asks to webdriver to login https://www.linkedin.com/. The variables are initialized to store username and password, the chrome web driver path is included. The login process is automated with selenium by sing XPath for login and password input box. The code is included in multi-threading.py

The last exercise is to use Appium to do a web crawling on Android app: Google News. I used an Android Emulator called Rahulemulator. This is done in Java with several set ups. The code is included in Appium_Google_News.java.