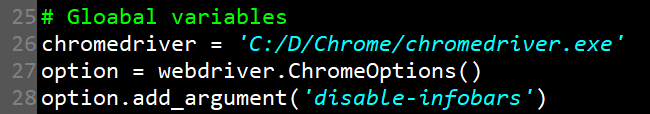
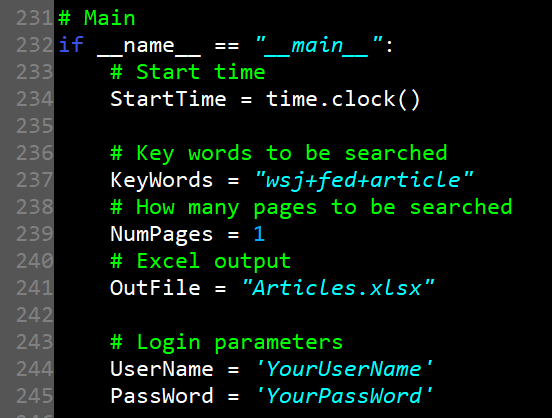
**Read Me for CrawlWSJ.py**

This is the Python code to automatically scrape Google pages of WSJ articles. Here are some tips to run this program.

1. To start with, in order to initiate the web driver provided by Selenium we need to first download the Chrome driver from <https://sites.google.com/a/chromium.org/chromedriver/downloads> and then point to its path of chromedriver.exe in line 26 to set it as a global variable:



1. In the main function starting from line 232, the parameters need to be set for the website information.



KeyWords store the key words to be searched in Google. For example we want to type “wsj fed article” in Google then we need to set it as a string named “wsj+fed+article”. After that please note the number of Google pages you want to search on. If NumPages = 10 it means that it will search the key words on Google from page 1 to page 10.

And please set the Login information of your own account from line 244 to 245.

1. After setting these locations you can run the program. But it is suggested that once it is running please stop using your computer because there will be constantly website bumping out and in case you click on the new websites to damage the procedure please just do nothing at that time.