爬蟲程式

#C:\Users\lll12\Desktop  chrom driver site

from ast import keyword

from selenium import webdriver

import time

from selenium.webdriver.common.keys import Keys

import pandas as pd

from selenium import webdriver

from selenium.webdriver.common.by import By

from selenium.webdriver.support.ui import WebDriverWait

from selenium.webdriver.support import expected\_conditions as EC

from selenium.webdriver.common.keys import Keys

path="C:/Users/lll12/Desktop/chromedriver.exe"#電腦中chromedriver的位置

from selenium import webdriver

#打開瀏覽器的用途

driver=webdriver.Chrome(path)

driver.get("https://www.twse.com.tw/zh/page/trading/exchange/STOCK\_DAY.html")

#print(driver.title)

time.sleep(5)

def myfun(df,item):

    df[item[0]] = item[1:]

    return df

items=['2832']

#在矩陣中打入股票代號即可開始爬取資料，一支股票完整的爬取時間可能會將近20-30分鐘(內容包含從'今天'到99年01/01)

for item in items:

    keyword=item

    item1=['日期']

    item2=['成交股數']

    item3=['成交總額']

    item4=['開盤價']

    item5=['最高價']

    item6=['最低價']

    item7=['收盤價']

    item8=['價差']

    item9=['成交筆數']

    searcher = WebDriverWait(driver, 10).until(

        EC.presence\_of\_element\_located((By.CLASS\_NAME, "stock-code-autocomplete")))

    searcher.clear()

    searcher.send\_keys(keyword)

    g=3000

    for i in range(1,2):

        years=driver.find\_element\_by\_xpath('//\*[@id="d1"]/select[1]/option[%s]'% i)

        years.click()

        for j in range(1,3):

            months=driver.find\_element\_by\_xpath('//\*[@id="d1"]/select[2]/option[%s]'%j)

            months.click()

            next=driver.find\_element\_by\_class\_name("stock-code-autocomplete")

            next.send\_keys(Keys.RETURN)

            time.sleep(3) #千萬千萬千萬不要動!!!! IP被鎖我可救不了你!!!!

            for k in range(1,26):

                g-=1

                print(g)

                dates=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[1]'%k)

                for date in dates:

                    print(date.text)

                    item1.append(date.text)

                stock\_counts=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[2]'%k)

                for stock\_count in stock\_counts:

                    item2.append(stock\_count.text)

                stock\_moneys=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[3]'%k)

                for stock\_money in stock\_moneys:

                    item3.append(stock\_money.text)

                open\_values=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[4]'%k)

                for open\_value in open\_values:

                    item4.append(open\_value.text)

                heightests=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[5]'%k)

                for heightest in heightests:

                    item5.append(heightest.text)

                lowests=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[6]'%k)

                for lowest in lowests:

                    item6.append(lowest.text)

                endvalues=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[7]'%k)

                for endvalue in endvalues:

                    item7.append(endvalue.text)

                gaps=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[8]'%k)

                for gap in gaps:

                    item8.append(gap.text)

                dealcounts=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[9]'%k)

                for dealcount in dealcounts:

                    item9.append(dealcount.text)

        print(item1)

    for i in range(2,4):

        years=driver.find\_element\_by\_xpath('//\*[@id="d1"]/select[1]/option[%s]'% i)

        years.click()

        for j in range(1,13):

            months=driver.find\_element\_by\_xpath('//\*[@id="d1"]/select[2]/option[%s]'%j)

            months.click()

            next=driver.find\_element\_by\_class\_name("stock-code-autocomplete")

            next.send\_keys(Keys.RETURN)

            time.sleep(3) #千萬千萬千萬不要動!!!! IP被鎖我可救不了你!!!!

            for k in range(1,26):

                g-=1

                print(g)

                dates=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[1]'%k)

                for date in dates:

                    print(date.text)

                    item1.append(date.text)

                stock\_counts=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[2]'%k)

                for stock\_count in stock\_counts:

                    item2.append(stock\_count.text)

                stock\_moneys=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[3]'%k)

                for stock\_money in stock\_moneys:

                    item3.append(stock\_money.text)

                open\_values=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[4]'%k)

                for open\_value in open\_values:

                    item4.append(open\_value.text)

                heightests=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[5]'%k)

                for heightest in heightests:

                    item5.append(heightest.text)

                lowests=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[6]'%k)

                for lowest in lowests:

                    item6.append(lowest.text)

                endvalues=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[7]'%k)

                for endvalue in endvalues:

                    item7.append(endvalue.text)

                gaps=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[8]'%k)

                for gap in gaps:

                    item8.append(gap.text)

                dealcounts=driver.find\_elements\_by\_xpath('//\*[@id="report-table"]/tbody/tr[%s]/td[9]'%k)

                for dealcount in dealcounts:

                    item9.append(dealcount.text)

    print('一',len(item1))

    print('二',len(item2))

    print('三',len(item3))

    print('四',len(item4))

    print('五',len(item5))

    print('六',len(item6))

    print('七',len(item7))

    print('八',len(item8))

    print('九',len(item9))

    df = pd.DataFrame()

    for i in [item1,item2,item3,item4,item5,item6,item7,item8,item9]:

        df = myfun(df,i)

        print(df)

        df.to\_csv(keyword+".csv",encoding='Big5',index=False)

    a=pd.read\_csv(keyword+'.csv',encoding="Big5")

    #a.head()

    oo=['cdp']

    ee=a['最高價']

    ff=a['最低價']

    gg=a['收盤價']

    hh=a['價差']

    '''=SUMIF(H2:H7,">0")'''

    '''=((G2-MIN(F3:F11))/(MAX(E3:E11)-MIN(F3:F11))\*100)'''

    #for u in range(10,int(len(ee))):

    o=(ee+ff+(2\*gg))/4

    a['CDP(均價)']=o

    oo=a['CDP(均價)']

    p=oo+(ee-ff)

    a['AH(行情最高價)']=p

    pp=['AH(行情最高價)']

    q=2\*oo-ff

    a['NH(近高值)']=q

    qq=['NH(近高值)']

    r=2\*oo-ee

    a['NL(近低值)']=r

    rr=['NL(近低值)']

    s=oo-(ee-ff)

    a['AL(行情最低價)']=s

    ss=['AL(行情最低價)']

    print(a)

    #a=a.drop(['Unnamed: 0'],axis=1)

    a.to\_csv(keyword+'.csv',encoding="Big5",index=False)

    item1.clear()

    item2.clear()

    item3.clear()

    item4.clear()

    item5.clear()

    item6.clear()

    item7.clear()

    item8.clear()

    item9.clear()