法人爬蟲程式碼

#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

def myfun(df,item):

    df[item[0]] = item[1:]

    return df

driver=webdriver.Chrome(path)

items=['2809','2324','2382','2357','3231','2365']

for item in items:

    driver.get("https://concords.moneydj.com/z/zc/zcl/zcl\_1101.djhtm")

    time.sleep(5)

    def myfun(df,item):

        df[item[0]] = item[1:]

        return df

    searcher = WebDriverWait(driver, 10).until(

            EC.presence\_of\_element\_located((By.XPATH, '//\*[@id="SysJustMenuContainer"]/table/tbody/tr/td/table/tbody/tr[4]/td/table/tbody/tr/td/form/input')))

    item1=['日期']

    item2=['外資買賣超']

    item3=['投信買賣超']

    item4=['自營商買賣超']

    item5=['買賣超總計']

    item6=['外資']

    item7=['投信']

    item8=['自營商']

    keyword=item

    searcher.clear()

    searcher.send\_keys(keyword)

    searcher.send\_keys(Keys.RETURN)

    time.sleep(2)

    money=driver.find\_element\_by\_xpath('//\*[@id="SysJustMenuContainer"]/table/tbody/tr/td/table/tbody/tr[36]/td/table/tbody/tr/td/a')

    time.sleep(2)

    money.click()

    mm1=driver.find\_element\_by\_xpath('//\*[@id="oScrollHead"]/td/select[3]/option[1]')

    mm1.click()

    dd1=driver.find\_element\_by\_xpath('//\*[@id="oScrollHead"]/td/select[4]/option[1]')

    dd1.click()

    mm2=driver.find\_element\_by\_xpath('//\*[@id="oScrollHead"]/td/select[6]/option[2]')

    mm2.click()

    dd2=driver.find\_element\_by\_xpath('//\*[@id="oScrollHead"]/td/select[7]/option[28]')

    dd2.click()

    next=driver.find\_element\_by\_xpath('//\*[@id="oScrollHead"]/td/input[3]')

    next.click()

    for i in range(8,41):

        dates=driver.find\_elements\_by\_xpath('//\*[@id="SysJustIFRAMEDIV"]/table[1]/tbody/tr/td/table/tbody/tr[3]/td[4]/table/tbody/tr/td/form/table/tbody/tr/td/table/tbody/tr[%s]/td[1]'%i)

        for date in dates:

            print(date.text)

            item1.append(date.text)

        fis=driver.find\_elements\_by\_xpath('//\*[@id="SysJustIFRAMEDIV"]/table[1]/tbody/tr/td/table/tbody/tr[3]/td[4]/table/tbody/tr/td/form/table/tbody/tr/td/table/tbody/tr[%s]/td[2]'%i)

        for fi in fis:

            print(fi.text)

            item2.append(fi.text)

        iis=driver.find\_elements\_by\_xpath('//\*[@id="SysJustIFRAMEDIV"]/table[1]/tbody/tr/td/table/tbody/tr[3]/td[4]/table/tbody/tr/td/form/table/tbody/tr/td/table/tbody/tr[%s]/td[3]'%i)

        for ii in iis:

            print(ii.text)

            item3.append(ii.text)

        sis=driver.find\_elements\_by\_xpath('//\*[@id="SysJustIFRAMEDIV"]/table[1]/tbody/tr/td/table/tbody/tr[3]/td[4]/table/tbody/tr/td/form/table/tbody/tr/td/table/tbody/tr[%s]/td[4]'%i)

        for si in sis:

            print(si.text)

            item4.append(si.text)

        totals=driver.find\_elements\_by\_xpath('//\*[@id="SysJustIFRAMEDIV"]/table[1]/tbody/tr/td/table/tbody/tr[3]/td[4]/table/tbody/tr/td/form/table/tbody/tr/td/table/tbody/tr[%s]/td[5]'%i)

        for total in totals:

            print(total.text)

            item5.append(total.text)

        fholds=driver.find\_elements\_by\_xpath('//\*[@id="SysJustIFRAMEDIV"]/table[1]/tbody/tr/td/table/tbody/tr[3]/td[4]/table/tbody/tr/td/form/table/tbody/tr/td/table/tbody/tr[%s]/td[6]'%i)

        for fhold in fholds:

            print(fhold.text)

            item6.append(fhold.text)

        iholds=driver.find\_elements\_by\_xpath('//\*[@id="SysJustIFRAMEDIV"]/table[1]/tbody/tr/td/table/tbody/tr[3]/td[4]/table/tbody/tr/td/form/table/tbody/tr/td/table/tbody/tr[%s]/td[7]'%i)

        for ihold in iholds:

            print(ihold.text)

            item7.append(ihold.text)

        sholds=driver.find\_elements\_by\_xpath('//\*[@id="SysJustIFRAMEDIV"]/table[1]/tbody/tr/td/table/tbody/tr[3]/td[4]/table/tbody/tr/td/form/table/tbody/tr/td/table/tbody/tr[%s]/td[8]'%i)

        for shold in sholds:

            print(shold.text)

            item8.append(shold.text)

    print(item1)

    df = pd.DataFrame()

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

        df = myfun(df,i)

        print(df)

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

    item1.clear()

    item2.clear()

    item3.clear()

    item4.clear()

    item5.clear()

    item6.clear()

    item7.clear()

    item8.clear()