

Selenium Xpath 應用

Outline

- 1 Selenium是什麼？
- 2 Selenium基礎介紹：元素定位的8種方法
- 3 什麼是 Xpath ？
- 4 什麼是相對路徑/絕對路徑定位？
- 5 xpath 文字定位方式？完全比對、模糊比對
- 6 如何抓取Xpath原本路徑full xpath？
- 7 STG後台頁面展示範例:抓取Xpath
- 8 如何縮短原本路徑full xpath？
- 9 欄位變化Xpath絕對路徑範例
- 10 參考文章

Selenium 是什麼？

Selenium 是用於自動化 Web 瀏覽器的工具，也是可以抓網頁內容的動態網頁爬蟲，常搭配 BeautifulSoup 解析 HTML 網頁原始碼。

其主要的 2 大應用為：

自動化測試：

Selenium 自動化測試框架主要用作網頁測試，可以直接驅動瀏覽器，模擬真正的使用者操作網站，包含打開瀏覽器、點擊或連結、輸入文字、取得網頁內容或內容比對驗證等。

網頁抓取 (Web Scraping):

用來自動提取網頁內容的應用程式，並把這些內容轉換成容易理解的格式呈現出來。

Selenium 基礎介紹：元素定位的 8 種方法

定位元素語法	Html 元素	描述
<code>find_element_by_id()</code>	id	ID 屬性，最常用的定位方法，每個元素的 id 應該是唯一的
<code>find_element_by_css_selector()</code>	css selector:	CSS 選擇器
<code>find_element_by_xpath()</code>	xpath:	xpath 表示式定位元素
<code>find_element_by_name()</code>	Name:	NAME 屬性，與 id 定位類似
<code>find_element_by_link_text()</code>	link_text()	僅用於超連結文字
<code>find_element_by_partial_link_text()</code>	partial_link_text()	使用方法和 link text 相同，partial link 只擷取部分文字即可
<code>find_element_by_tag_name()</code>	tag_name	通過 HTML 標籤名定位
<code>find_element_by_class_name()</code>	Class_name	使用類名定位，不能使用複合類名

在這裡主要介紹 xpath 的應用

Selenium 基础介绍：元素定位的 8 種方法

定位元素語法	Html 元素	描述
<code>find_element_by_css_selector()</code>	css selector:	CSS 選擇器
<code>find_element_by_xpath()</code>	xpath:	xpath 表示式定位元素
<code>find_element_by_name()</code>	Name:	NAME 屬性，與 id 定位類似
<code>find_element_by_link_text()</code>	link_text()	僅用於超連結文字
<code>find_element_by_partial_link_text()</code>	partial_link_text()	使用方法和 link text 相同，partial link 只擷取部分文字即可
<code>find_element_by_tag_name()</code>	tag_name	通過 HTML 標籤名定位
<code>find_element_by_class_name()</code>	Class_name	使用類名定位，不能使用複合類名
<code>find_elements</code>	多個元素	尋找多個元素
<code>find_element</code>	單一個元素	尋找單一個元素

Selenium 基础介绍：Xpath 路徑範例

html 模板元素：

```
<html>
<body>
  <form id="loginForm">
    <input name="username" type="text" />
    <input name="password" type="password" />
    <input name="continue" type="submit" value="Login" />
    <input name="continue" type="button" value="Clear" />
  </form>
</body>
</html>
```

Xpath 路徑：

```
1 login_form = driver.find_element_by_xpath("/html/body/form[1]")      # 絕對路徑
2 login_form = driver.find_element_by_xpath("//form[1]")              #html 的第一個 form 元素
3 login_form = driver.find_element_by_xpath("//form[@id='loginForm']") # 用 form 元素 與 id 屬性值來定位
```

什麼是 Xpath ?

XPath 是一種 XML 路徑，用於瀏覽頁面的 HTML 結構。他是一種語法或者語言用來查詢使用 XML 路徑表達的網頁中的任意元素。

XPath 的基本形式如下：

Xpath=//

tagname [@attribute='value']

- **//**：選中當前節點
- **Tagname**：特定節點的標記名
- **@**：選中屬性的標記符
- **Attribute**：節點的屬性名字
- **Value**：屬性值

什麼是相對路徑 / 絕對路徑定位？

相對定位 -- 以 // 開頭

如：//form//input[@name="phone"]

絕對定位 -- 以 / 開頭，但是要從根目錄開始，比較繁瑣，一般不建議使用 如：/html/body/div/a


xpath 文字定位方式？ 完全比對、模糊比對

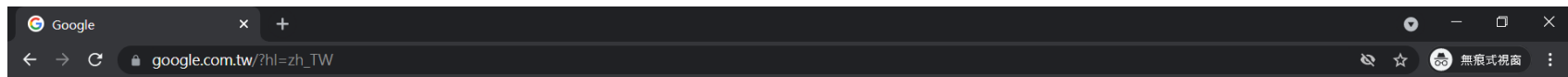
EX: 百度搜索

xpath 寫法為 //a[text()=' 百度搜索 '] // 文字完全比對

或者 //a[contains(text()," 百度搜索 ")] // 文字模糊比對

如何抓取 Xpath 原本路徑 full xpath?

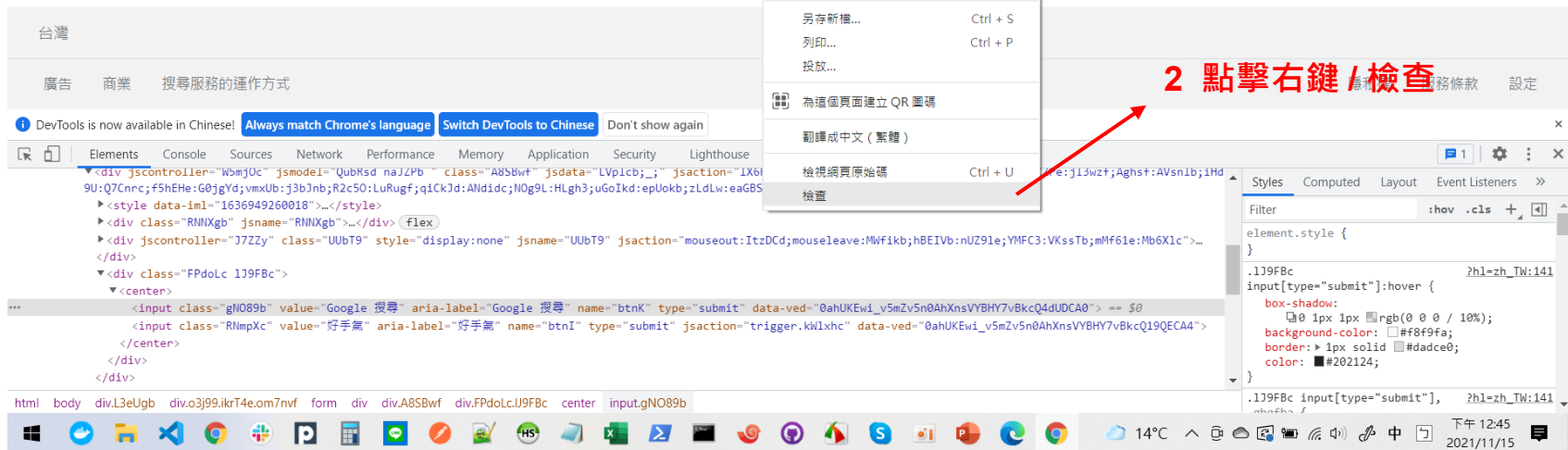
- 1 將滑鼠移動到要操作的網頁元素上面
- 2 點擊右鍵 / 檢查
- 3 將滑鼠移動到元素頁框
- 4 點擊左上角元素選擇器 
- 5 將滑鼠移動到網頁元素頁上作點擊
- 6 將滑鼠移動到元素頁框要查詢的元素文字上面，點擊右鍵
- 7 移動到 Copy -> Copy full XPath
- 8 將 full xpath 貼到範例檔案 `xpath_test.py` 執行測試 // 需先開啟後台 VPN 連線



1 將滑鼠移動到要操作的網頁元素上面



2 點擊右鍵 / 檢查



Google

google.com.tw/?hl=zh_TW

關於 Google Google 商店

Gmail 圖片 登入

Google

台灣

廣告 商業 搜尋服務的運作方式

DevTools is now available in Chinese! Always match Chrome

Elements Console Sources Network

4 點擊左上角元素選擇器

5 將滑鼠移動到網頁元素頁上作點擊

6 將滑鼠移動到元素頁框要查詢的元素文字上面，點擊右鍵

7 移動到 Copy -> Copy full XPath

3 將滑鼠移動到元素頁框

1

input.gN089b

Google 搜尋

好手氣

Copy element

Copy outerHTML

Copy selector

Copy JS path

Copy styles

Copy XPath

Copy full XPath

Styles

Computed Layout Event Listeners

Filter

element.style {

.lJ9FBc

input[type="submit"]:hover {

box-shadow:

0 1px 1px rgb(0 0 0 / 10%);

background-color: #f8f9fa;

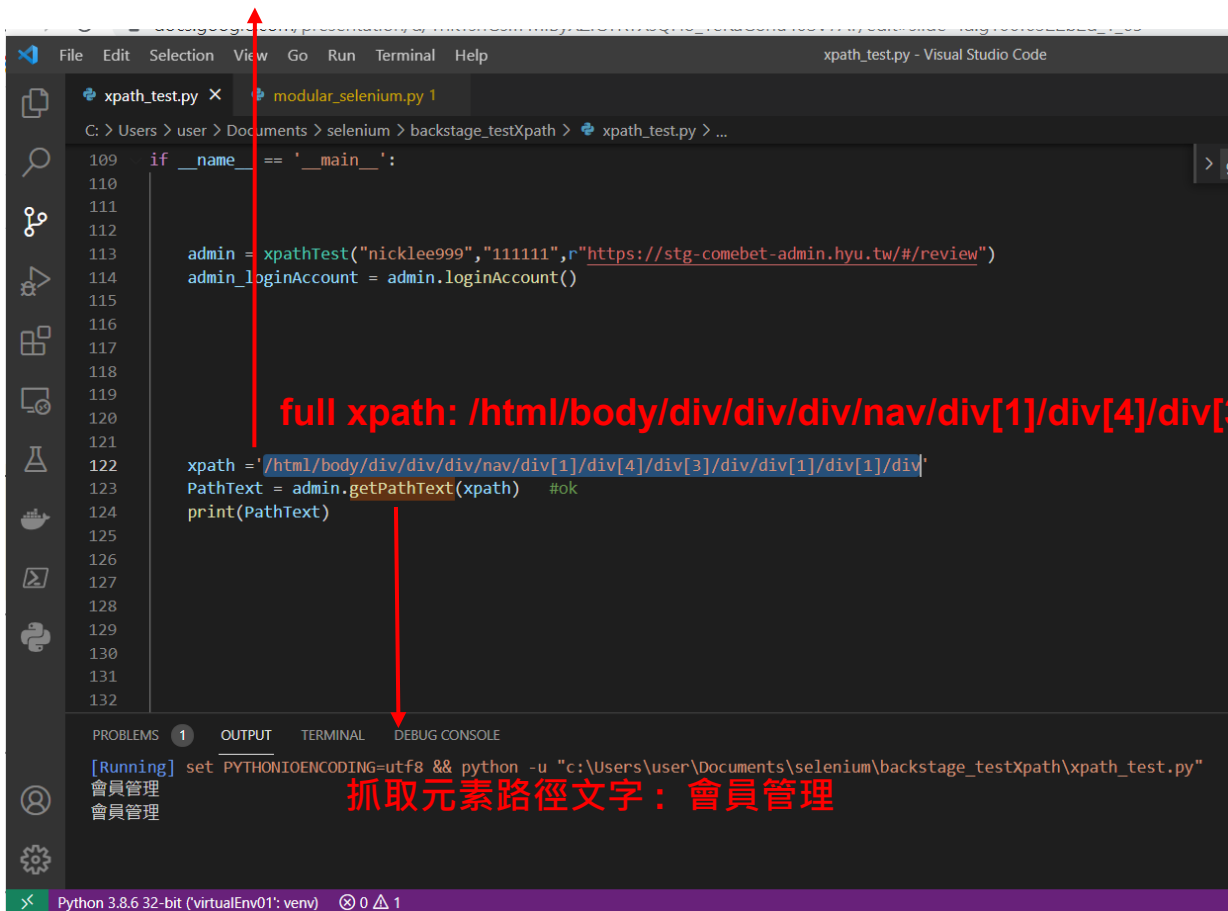
border: 1px solid #dadae0;

color: #202124;

.lJ9FBc input[type="submit"],

2021/11/15

8 將 full xpath 貼到範例檔案 xpath_test.py 執行測試 // 需先開啟後台 VPN 連線



```
109 if __name__ == '__main__':
110
111
112
113 admin = xpathTest("nicklee999", "111111", r"https://stg-comebet-admin.hyu.tw/#/review")
114 admin_loginAccount = admin.loginAccount()
115
116
117
118
119
120
121
122 xpath = '/html/body/div/div/div/nav/div[1]/div[4]/div[3]/div/div[1]/div[1]/div'
123 PathText = admin.getPathText(xpath) #ok
124 print(PathText)
125
126
127
128
129
130
131
132
```

full xpath: /html/body/div/div/div/nav/div[1]/div[4]/div[3]/div/div[1]/div[1]/div

OUTPUT

```
[Running] set PYTHONIOENCODING=utf8 && python -u "c:\Users\user\Documents\selenium\backstage_testXpath\xpath_test.py"
會員管理
會員管理
```

抓取元素路徑文字：會員管理

Python 3.8.6 32-bit (virtualEnv01: venv) 0 1

xpath 文字定位方式：完全比對

完全比對語法：'`//div[text() = "會員管理"]`'

網頁

時區選擇

2021-11-09 14:50:44

首頁

第三方遊戲平台

會員管理

> 會員列表

> 在線會員列表

今日會員資料統計

註冊人數 : 32	日活 : 1
當前在線 : 8	有效投注額 : 110.00
同IP登錄會員數量 : 10	同裝置登錄會員數量 : 0

會員存款統計

昨日首存 : 1人, 100.00元	今日首存 : 0人, 0.00元
昨日首存率 : 16.67%	今日首存率 : 0.00%

複製元素顯示文字

```
<div></div>
<div>
  <div class="v-list-group v-list-group--active primary--text">
    <div tabindex="0" aria-expanded="true" role="button" class="v-list-group__header v-list-item v-list-item--before">
      <div class="v-list-item__content">
        <div class="v-list-item__title">會員管理</div>
      </div>
    </div>
  </div>
  <div class="v-list-item__icon">
    <div class="v-list-group__header__append-icon"></div>
  </div>
  <div class="v-list-item__after"></div>
  <div class="v-list-group__items"></div>
</div>
```

```
File Edit Selection View Go Run Terminal Help
xpath_test.py modular_selenium.py 1
C: > Users > user > Documents > selenium > backstage_testXPath > xpath_test.py > ...

118
119
120
121
122
123
124
125
126
127 xpath="//div[text()='會員管理']"#OK
128 PathText = admin.getPathText(xpath) #ok
129 print(PathText)
130
131
132
133
134
135
136
137
138

PROBLEMS 1 OUTPUT TERMINAL DEBUG CONSOLE
[Running] set PYTHONIOENCODING=utf8 && python -u "c:\Users\user\Docume
會員管理
```

xpath 文字定位方式：模糊比對

模糊比對語法：'`//div[contains(text(), "會員管理")]`'

模糊比對

複製元素顯示文字

```
121
122
123
124
125
126
127
128
129
130 xpath="//div[contains(text(), \"會員管理\")]"#OK
131 PathText = admin.getPathText(xpath) #ok
132 print(PathText)
133
134
135
136
137
138
139
140
141
```

PROBLEMS 1 OUTPUT TERMINAL DEBUG CONSOLE

[Running] set PYTHONIOENCODING=utf8 && python -u "c:\Users\user\Documents\selenium\backstage_test\XPath > xpath_test.py ...

[Done] exited with code=0 in 0.178 seconds

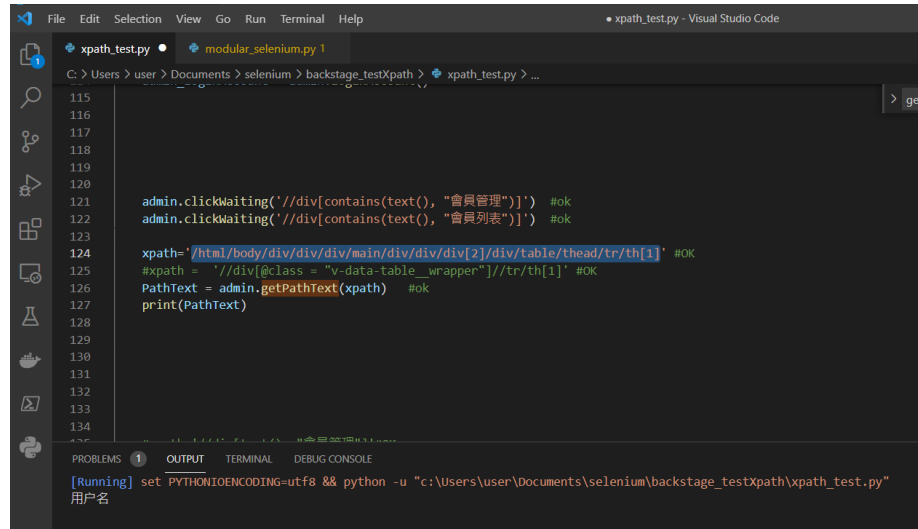
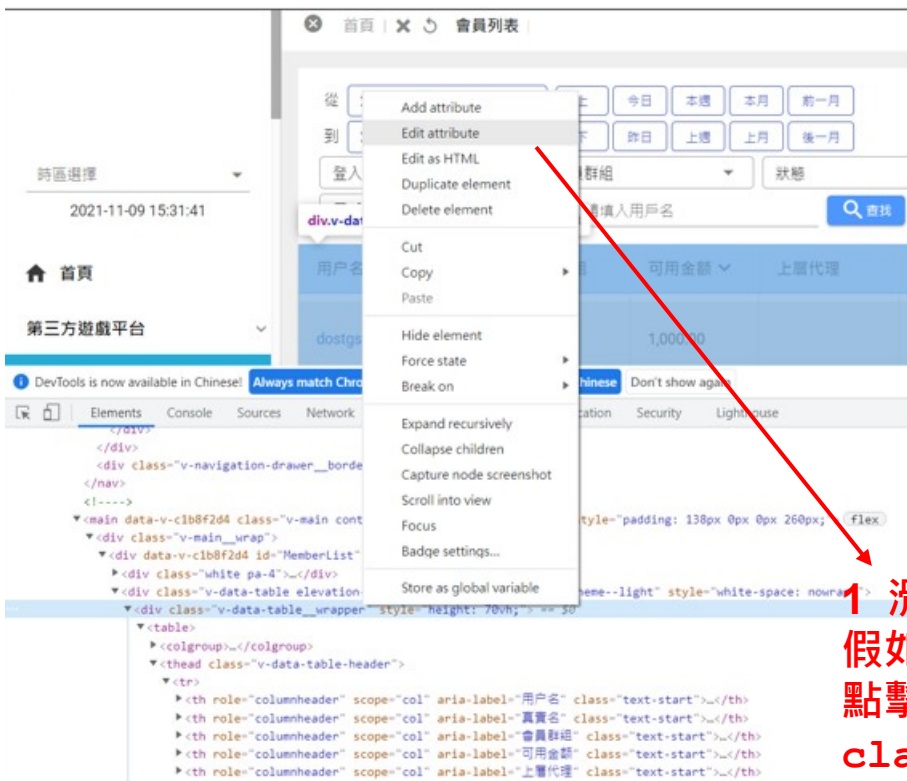
[Running] set PYTHONIOENCODING=utf8 && python -u "c:\Users\user\Documents\selenium\backstage_test\XPath > xpath_test.py ...

會員管理

Python 3.8.6 32-bit (VirtualEnv01: venv) 0 1

如何縮短原本路徑 full xpath?

範例：會員管理 > 會員列表 > 抓取欄位名稱 and 欄位資料



1 滑鼠在元素頁面上滑動，瀏覽器對應的區塊會變色框住
假如有框到要抓取的文字區塊，則可以停下滑鼠位置，
點擊右鍵 > Edit attribute > 複製屬性與屬性值
class="v-data-table__wrapper"

2 比對原本路徑 full xpath

```
'/html/body/div/div/div/main/div/div/div[2]/div/table/tbody/tr[1]/td[1]/a'
```

縮短路徑：

```
'//div[@class = "v-data-table__wrapper"]//tr[1]/td[1]/a'
```

比對區塊標籤元素

tagname

將屬性與屬性值

class="v-data-table__wrapper"

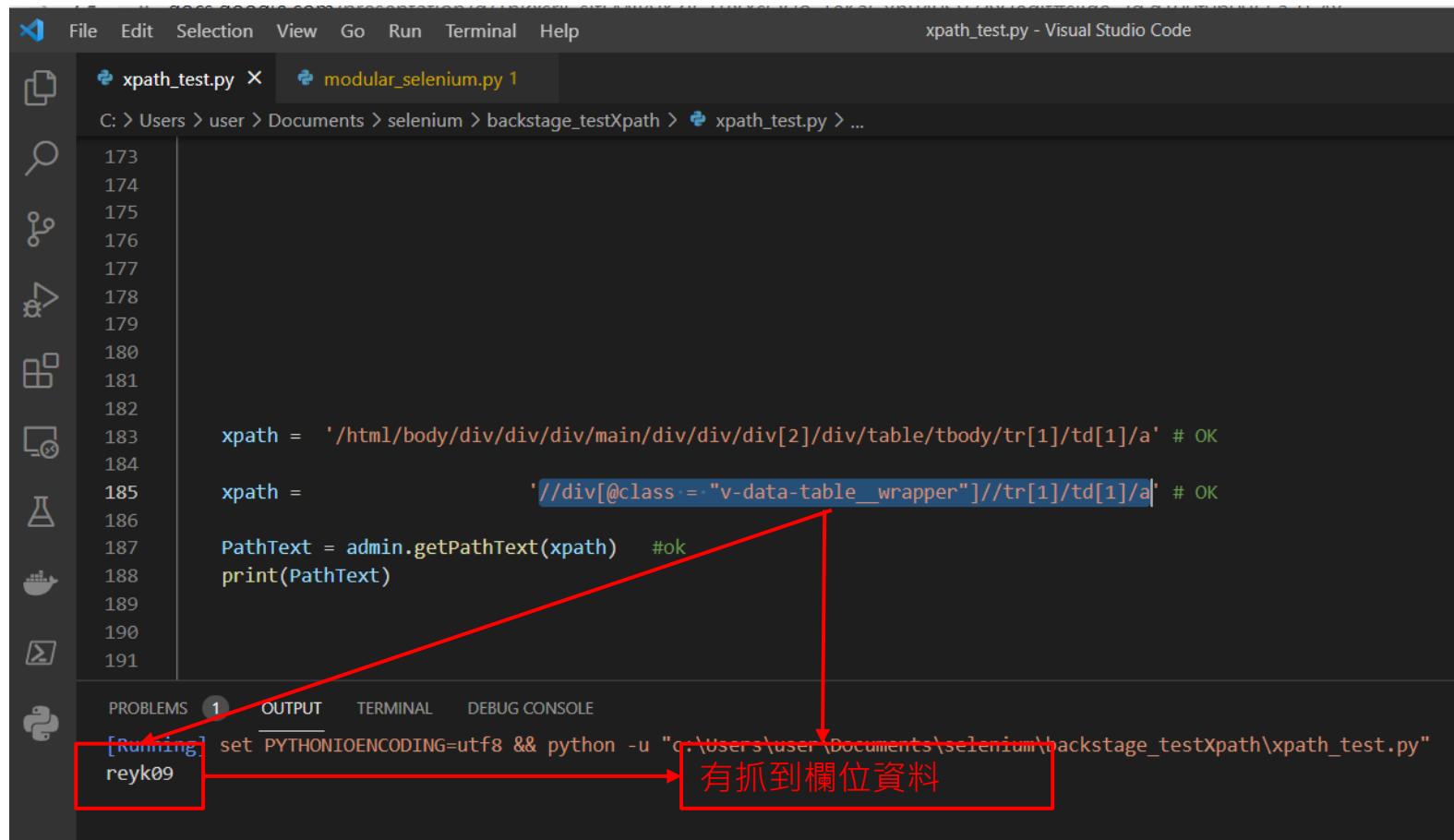
放置中括號 [] 裡面

末端路徑定位

// 縮短路徑

視元素區塊可更換標籤 ex: span 、 div or 去除標籤 tr[1]/td[1] 測試

將縮短的路徑貼到範例檔案測試



The screenshot shows the Visual Studio Code interface with two open files: `xpath_test.py` and `modular_selenium.py 1`. The active file is `xpath_test.py`, which contains the following code:

```
173
174
175
176
177
178
179
180
181
182
183     xpath = '/html/body/div/div/div/main/div/div/div[2]/div/table/tbody/tr[1]/td[1]/a' # OK
184
185     xpath = '//div[@class = "v-data-table__wrapper"]//tr[1]/td[1]/a' # OK
186
187     PathText = admin.getPathText(xpath) #ok
188     print(PathText)
189
190
191
```

The output panel at the bottom shows the command executed in the terminal:

```
[Running] set PYTHONIOENCODING=utf8 && python -u "c:\Users\user\Documents\selenium\backstage_testXpath\xpath_test.py"
```

The output shows the result of the script execution, which is "有抓到欄位資料" (Successfully captured the field data).

Red arrows indicate the flow of information: one arrow points from the XPath expression in the code to the terminal output, and another arrow points from the terminal output to the text "有抓到欄位資料".

欄位變化 xpath 絕對路徑範例

後台會員管理模組 子分頁 tableXPath 會員列表

td[*] index 數字變化 1~10 可定位到欄位

用戶名	真實名	會員群組	可用金額 ▾	上層代理	註冊IP/註冊時間 ▾	最後登入/最後訪問時間 ▾	狀態	備註	操作
jarytest	GARYTEST	VIPO	298.00	上級: mi4567	35.187.204.224 2021-07-13 10:05:00	35.187.204.224 2021-10-26 10:38:52	離線 啟用	test	<div><div>⊕ 加值</div><div>✎ 帳號</div></div> <div><div>✎ 編輯</div><div>📄 報表</div></div>

表格變化 xpath 參數

'/html/body/div/div/div/main/div/div/div[2]/div/table/tbody/tr[*]/td[*]'

tr[*] index 數字變化可定位到表格內容第幾列

td[*] index 數字變化可定位到表格內容第幾欄

參考文章：

[Python + Selenium網頁元素定位XPath](#)

[selenium 通過xpath進行元素定位--By.xpath\(\)](#)

[後台範例檔案下載](#)