
# 網路程式設計
# 網路爬蟲 動態網頁解析

Instructor: 馬豪尚

# JavaScript

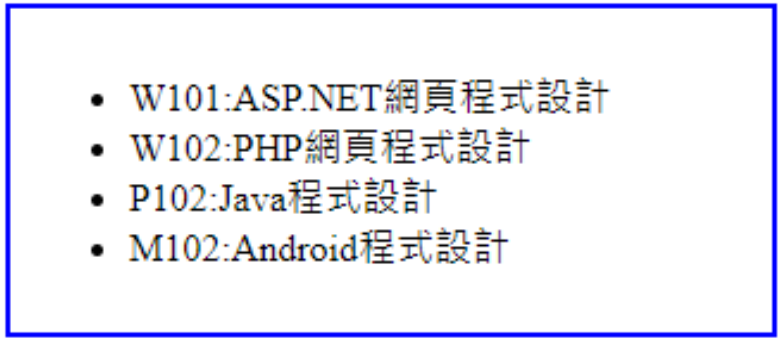
- JavaScript 是一種腳本，也能稱它為程式語言，可以讓你在網頁中實現出複雜的功能。
- JavaScript常用來完成以下任務
  - 嵌入動態文字於HTML頁面
  - 對瀏覽器事件作出回應
  - 讀寫HTML元素
  - 在資料被提交到伺服器之前驗證資料
  - 檢測訪客的瀏覽器資訊
  - 控制Cookie，包括建立和修改等

# Quick JavaScript Switcher

› Chrome瀏覽器的擴充應用程式

# Quick JavaScript Switcher

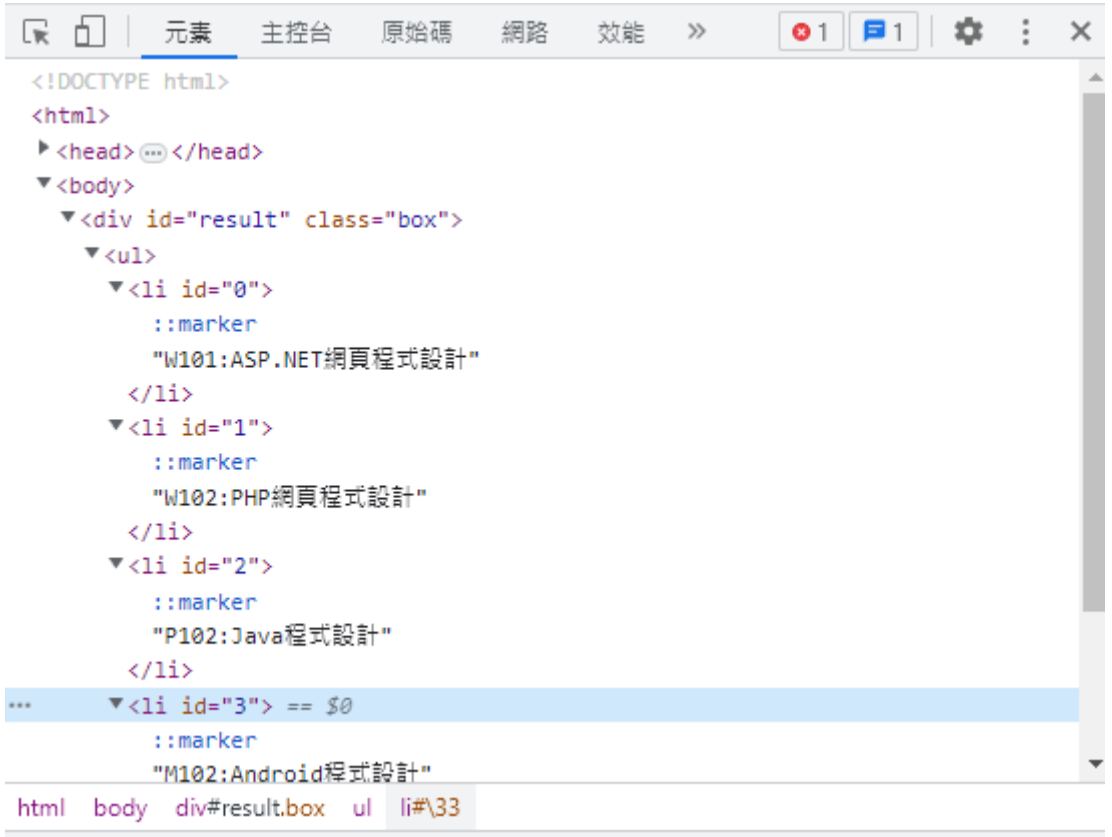
- 測試網址
  - https://fchart.github.io/books.html



# Chrome瀏覽器開發人員工具

› 在瀏覽器的網頁頁面上點右鍵->檢查

- W101:ASP.NET網頁程式設計
- W102:PHP網頁程式設計
- P102:Java程式設計
- M102:Android程式設計

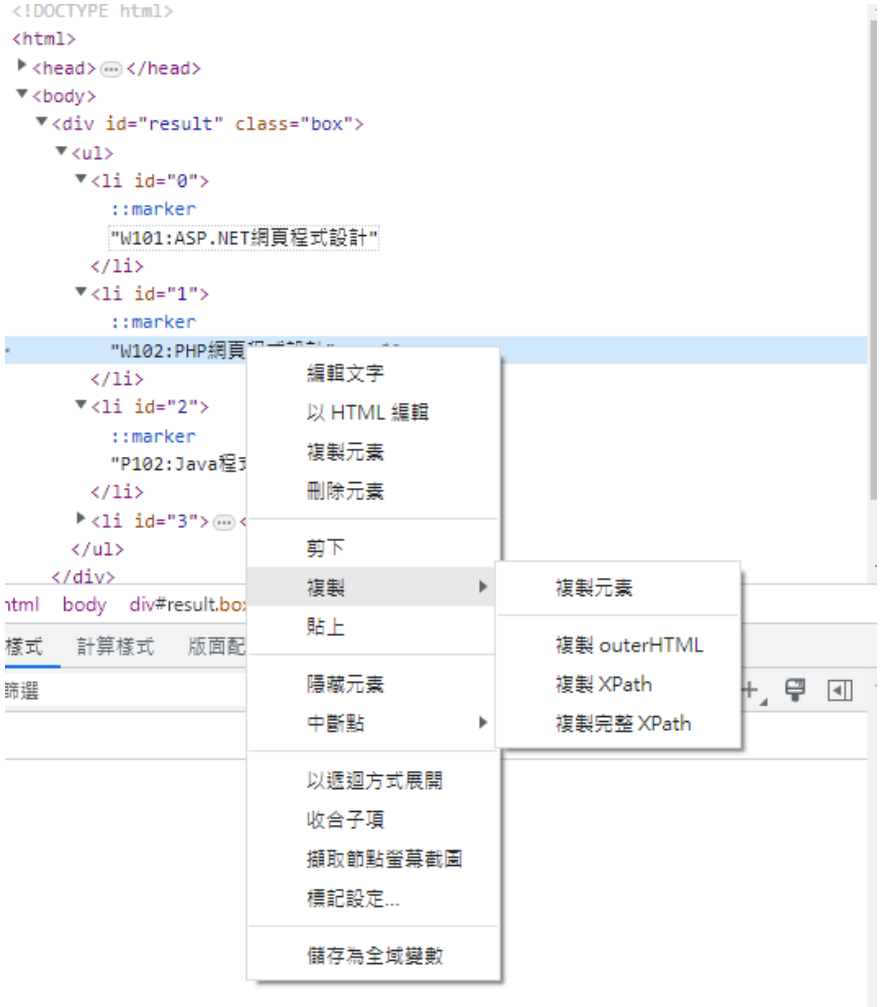



# Chrome瀏覽器開發人員工具

› 可以檢視每一個html元素

# 取得選取元素的網頁定位資料

› 在該選取元速點選右鍵->複製




複製元素

<li id="1">W102:PHP網頁程式設計</li>

# 爬取網站

- 分析該網站的HTML
- 判斷JavaSrcipt是否影響目標網頁內容
- 選擇取得資源的方式
  - Request
  - Selenium
- 撰寫爬蟲和網頁分析語法
  - Beautifulsoap
  - Selenium

# 爬蟲with JavaScript實務#1

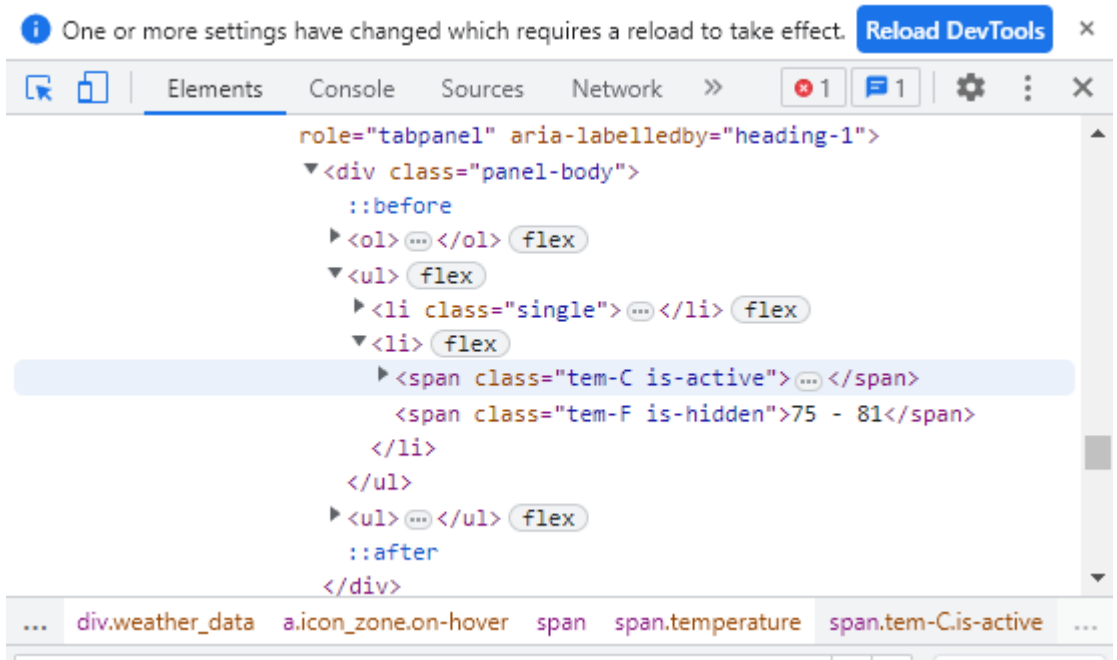
- 爬取氣象局天氣資訊
  - https://www.cwb.gov.tw/V8/C/W/County/County.html?CID=65

# 爬取網站

- 使用chrome瀏覽器開發人員工具分析html
- JavaScript會影響氣象網站內容
- 選擇獲得資源的方式
  - Selenium
- 解析網站的語法
  - Beautifulsoap

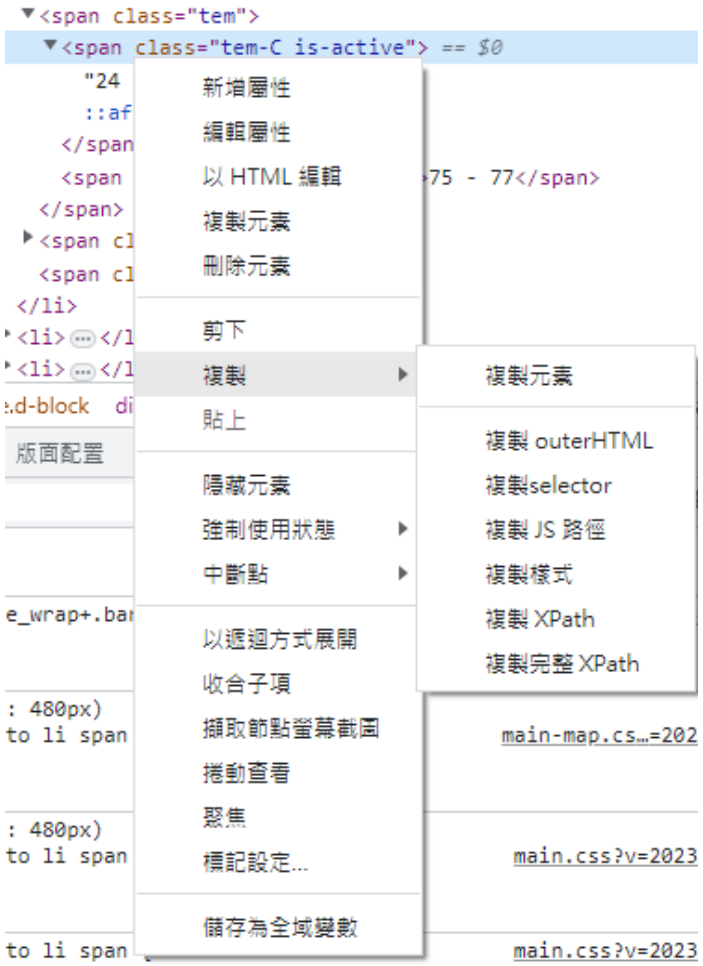
# Chrome瀏覽器開發人員工具

› 可以檢視每一個html元素

# 取得選取元素的網頁定位資料

› 在該選取元速點選右鍵->複製




複製元素

24 - 25

複製css selector

body > div.wrapper > main > div > div:nth-child(1) > div.d-xl-none.d-block > div.banner_wrap > ul > li:nth-child(1) > span.tem > span.tem-C.is-active

# 爬蟲with JavaScript實務#2

- 爬取momo購物網NBA球衣的商品資料
  - https://www.momoshop.com.tw/search/searchShop.jsp?keyword=nike NBA&searchType=1&curPage=1&_isFuzzy=0&showType=chessboardType

# 爬取網站

- 使用chrome瀏覽器開發人員工具分析html
- JavaScript會影響momo購物網站內容
- 選擇獲得資源的方式
  - Selenium
- 解析網站的語法
  - Beautifulsoap
- 與網頁互動來獲取更多資料
  - Selenium

# 取得想要爬取元素的網頁定位資料


- 商品名稱
  - 複製元素的css selector
  - #BodyBase > div.bt_2_layout.searchbox.searchListArea > div.searchPrdListArea.bookList > div.listArea > ul > li:nth-child(1) > a > div.prdInfoWrap > div.prdNameTitle > h3

# 取得想要爬取元素的網頁定位資料


- 商品價錢
  - 複製元素的css selector
  - #BodyBase > div.bt_2_layout.searchbox.searchListArea.selectedtop > div.searchPrdListArea.bookList > div.listArea > ul > li:nth-child(1) > a > div.prdInfoWrap > p.money > span.price

# 取得下一頁的資料


- 找到下一頁按鈕的元素定位
  - 複製元素的css selector
  - #BodyBase > div.bt_2_layout.searchbox.searchListArea.selectedtop > div:nth-child(6) > dl > dd > a
- .click()操作

# 練習

- 爬取momo網站其他任何類別的物品
  - 商品名稱
  - 商品價格
- 存成json檔
  - id: 流水編號
  - title: 商品名稱
  - price: 商品價格