

R-基本的爬蟲教學

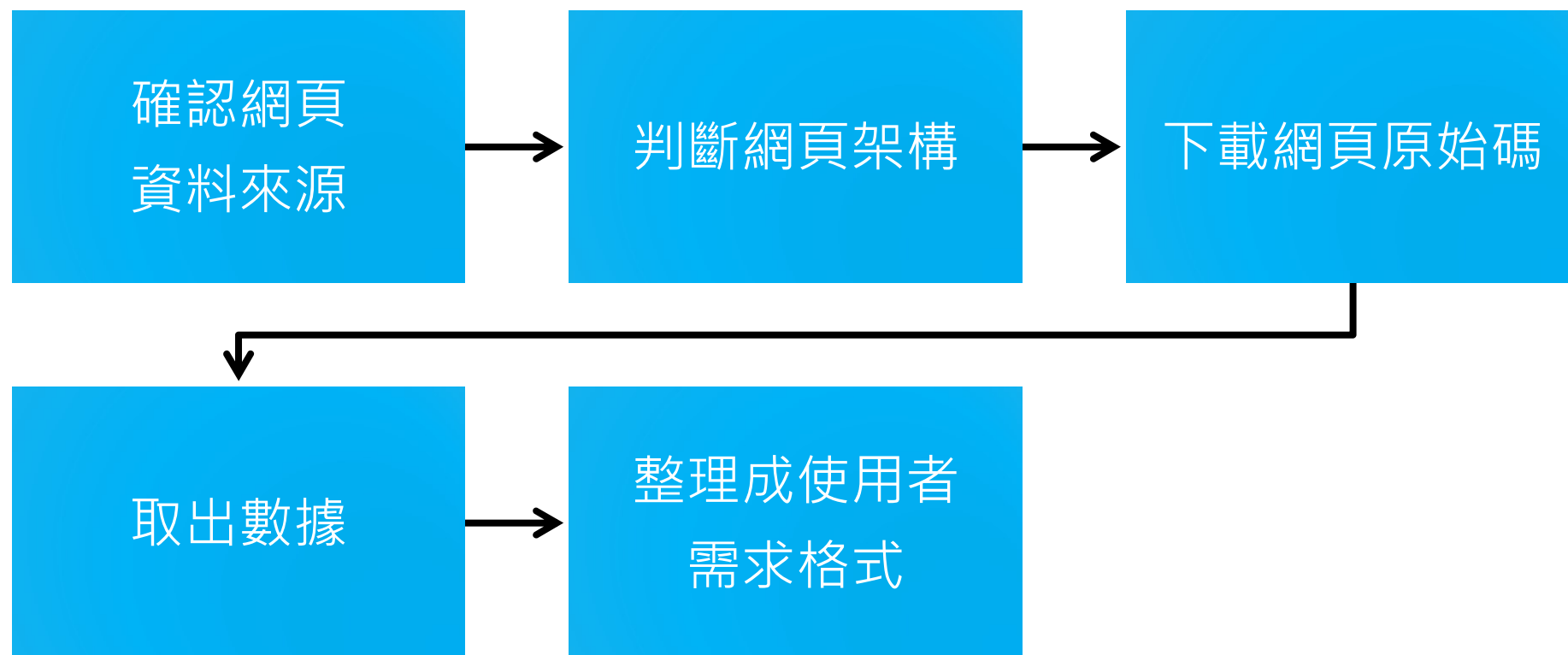
指導教授：王昭文 教授

中山財管所 104級 蘇彥庭

網路爬蟲

- 網路爬蟲(Web crawler)，又稱網路蜘蛛(Web spider)、自動檢索工具(automatic indexer)，是一種「自動化瀏覽網路」的程式，或者說是一種網路機器人。
- 網路爬蟲能夠自動採集所有能夠存取的網頁頁面內容，提供使用者做進一步的分析。
- 網路爬蟲可以說是駭客行為的一種，因為它會大量耗用掉網站的流量，並不被網站擁有者所樂見，因此盡可能偽裝成一般使用者，避免不必要的法律困擾。

網路爬蟲流程



網路爬蟲應注意事項

- 網路爬蟲發出的請求盡量一次就做對，不要有太多異常的請求。
- 盡量偽裝成正常使用者(可利用Sys.sleep和rand函數)，不要短時間大量發出請求。
- 尊重欲爬取資料的網站，不要吃光對方流量，干擾到其正常的服務。
- 注意資料版權問題，是否能夠公開及自由運用。若該網站未宣告，則應假定資料的版權為該網站所有。

爬蟲範例-下載奇摩股市網股票價格資料

Step1. 確認網頁資料來源

是新用戶嗎？請註冊 | 登入 | 服務說明 | Download the new Yahoo Mail app | 信箱 | 拍賣 | Yahoo 奇摩

YAHOO! 股市 奇摩

搜尋 網頁搜尋

雙王帶頭衝 台股收22個月新高 實質利率由負轉正 這樣做讓錢不縮水！ 三大法人買超台股152.67億

首頁 | 投資組合 | 當日行情 | 大盤 | 類股 | 期權 | 港滬深股 | 美股 | 新聞 | Y 選股

股票代號/名稱 1101 --當日個股股價-- 查詢 2017/03/21 21:46 距離美股收盤還有6小時14分鐘 更新

台股 | 亞股 | 歐股 | 美股 >更多

道瓊工業	20963.49	57.63
那斯達克	5922.61	21.08
S&P500	2380.19	6.72
費城半導體	1014.20	1.91

境外基金淨值 境外基金績效排行 全球商品期貨報價

消費情報

- 【指數王】破萬加碼送勝博殿+肯德基
- [機不可失]開戶即送周英老師波浪洩天機!
- 【力道K線】紅買綠賣，新功能AV多空訊號

選擇角色

一般裝備 衣服裝備

地獄南瓜帽	天使翅膀耳飾
巧克力棒	審判外袍 D+2
+5審判鐵鎚	誓言之書 (1)

Step1. 確認網頁資料來源

範圍											資料日期: 106/03/21
股票代號	時間	成交	買進	賣出	漲跌	張數	昨收	開盤	最高	最低	個股資料
1101台泥 加到投資組合	14:30	38.40	38.35	38.40	△0.25	6,755	38.15	38.00	38.40	37.85	成交明細 技術新聞 基本籌碼 個股健診

爬蟲範例-下載奇摩股市網股票價格資料

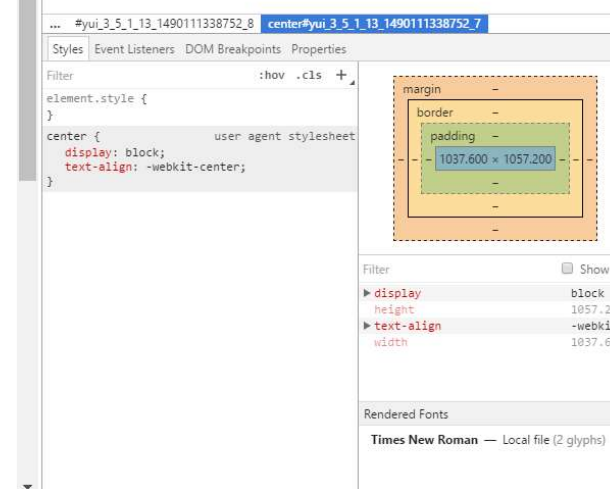
Step1. 確認網頁資料來源



3.



4. 網頁原始碼

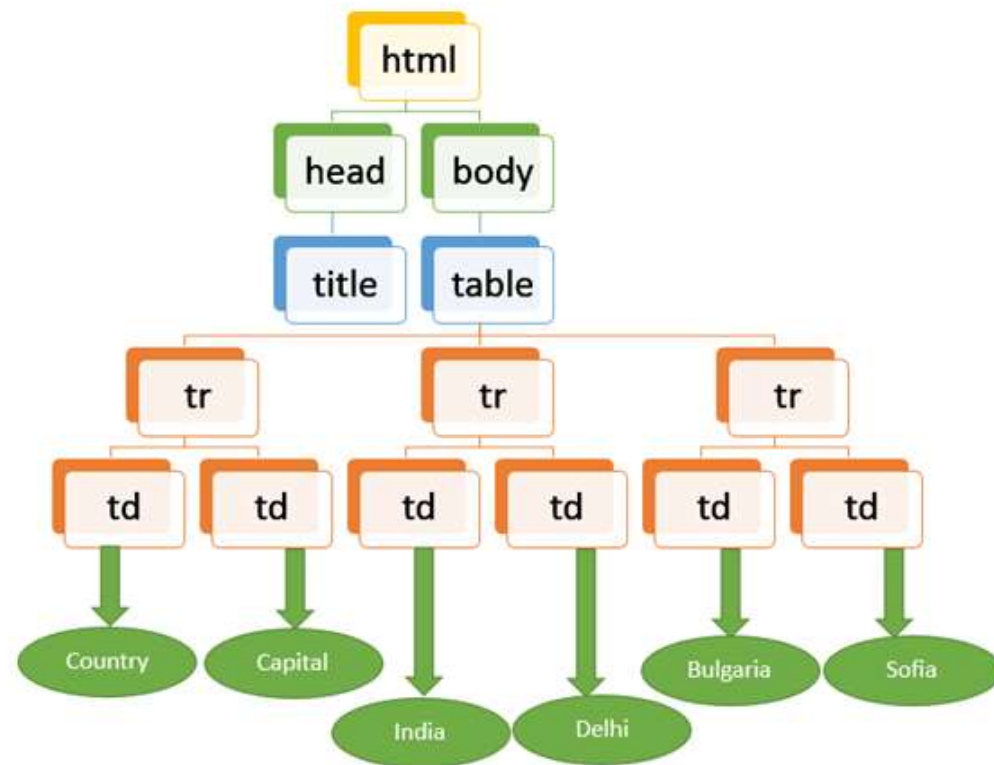


爬蟲範例-下載奇摩股市網股票價格資料

Step2. 判斷網頁架構

網頁架構示意圖

網頁原始碼就像樹一樣，
圖片、資料都各自存放在
不同的分支，我們要在網
頁原始碼中找到存放我們
想要的資料位置



圖片來源：<https://telerikhelper.net/2013/04/11/what-is-document-object-model/>

Step2. 判斷網頁架構

是 newUser 嗎? 請註冊 | 登入 | 服務說明

Download the new Yahoo Mail app

信箱 | 拍賣 | Yahoo 奇摩

YAHOO! 股市 奇摩

搜尋

網頁搜尋

仙境傳說 1轉職業 劍士 (Swordman)

2轉職業 騎士 (Knight)

3轉職業 盧恩騎士 (Rune Knight)

開始遊戲

首頁 投資組合 當日行情 大盤 類股 期權 港滬深股 美股 新聞 Y 選股

股票代號名稱 1101 查詢 2017/03/22 09:03 距離台股收盤還有4小時27分鐘 更新

table 749.6 x 175.2 : 委託成交 庫存報價 資料日期: 106/03/22

股票代號	時間	成交	買進	賣出	漲跌	張數	昨收	開盤	最高	最低	個股資料
1101台泥 加到投資組合	09:03	37.90	37.85	37.96	0.10	383	38.40	37.95	37.95	37.75	成交明細 技術 新聞 基本 籌碼 個股健診

凱基證券下單 買 賣 張 送出 零股交易

指標停留的位置

爬蟲範例-下載奇摩股市網股票價格資料

Step2. 判斷網頁架構

是新用戶嗎？請註冊 | 登入 | 服務說明 | Download the new Yahoo Mail app | 信箱 | 拍賣 | Yahoo 奇摩

YAHOO! 股市 奇摩

搜尋 網頁搜尋

巧克力棒 審判外袍 D+2 審判戰靴 審判披肩 I+1

審判戰靴 審判戰靴 審判戰靴 審判戰靴 審判戰靴

首頁 投資組合 當日行情 大盤 類股 期權 港滬深股 美股 新聞 選股

股票代號/名稱 1101 當日個股股價 查詢 2017/03/21 23:48 距離美股收盤還有4小時12分鐘 更新

凱基客戶專區：委託成交 庫存報價 資料日期: 106/03/21

股票代號	時間	成交	買進	賣出	漲跌	張數	昨收	開盤	最高	最低	個股資料
63.4 x 17.6											成交明細 技術 新聞 基本 籌碼 個股健診
1101 台泥	14:30	38.40	38.35	38.40	Δ0.25	6,755	38.15	38.00	38.40	37.85	

凱基證券下單 買 賣 張 送出 零股交易

在原始碼中找到資料儲存的位置

```
<td align="center" width="105">
  <a href="/q/bc?s=1101_1101台泥"/>台泥</a>
  <br>
  <a href="/pf/pfsel?stocklist=1101">...</a>
</td>
<td align="center" bgcolor="#FFFFFF" nowrap>14:30</td>
<td align="center" bgcolor="#FFFFFF" nowrap>
  <b>38.40</b>
</td>
<td align="center" bgcolor="#FFFFFF" nowrap>38.35</td>
<td align="center" bgcolor="#FFFFFF" nowrap>38.40</td>
<td align="center" bgcolor="#FFFFFF" nowrap>
  <font color="#ff0000">Δ0.25
  </font>
</td>
<td align="center" bgcolor="#FFFFFF" nowrap>6,755</td>
<td align="center" bgcolor="#FFFFFF" nowrap>38.15</td>
<td align="center" bgcolor="#FFFFFF" nowrap>38.00</td>
<td align="center" bgcolor="#FFFFFF" nowrap>38.40</td>
<td align="center" bgcolor="#FFFFFF" nowrap>37.85</td>
<td align="center" width="137" class="tt">
  <a href="/q/bc?s=1101">成交明細</a>
  <br>
  <a href="/q/ta?s=1101">技術</a>
  <a href="/q/h?s=1101">新聞</a>
  <a href="/q/s/company_1101.html">基本</a>
  <a href="/q/le/condit_1101.html">籌碼</a>
  <a href="/q/le/condit_1101.html">個股健診</a>
</td>
```

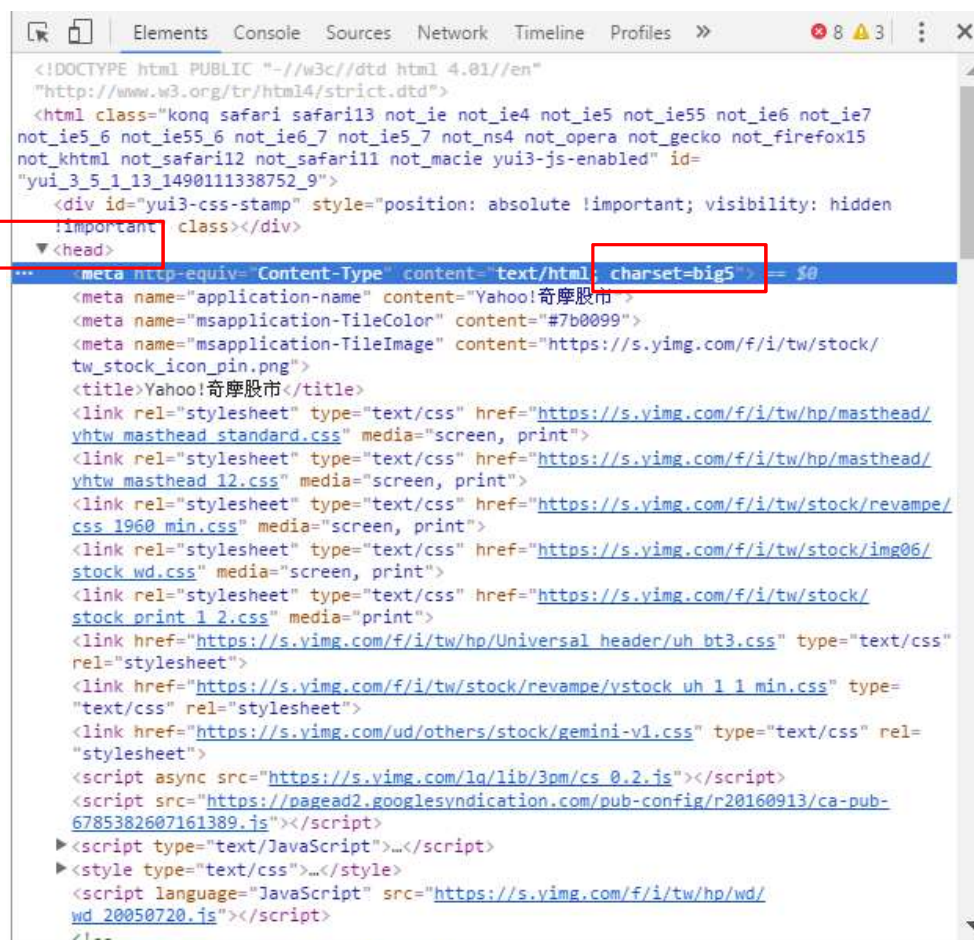
我們的目標：把在原始碼內的資料取出來並整理成我們希望的格式

爬蟲範例-下載奇摩股市網股票價格資料

Step2. 判斷網頁架構

在爬取網頁時要注意字碼(encoding)的問題，若字碼字碼如果用錯，載入到R時，中文字會變成亂碼，造成資料無法判讀。

通常字碼都會放在網頁<head>標籤內的前面幾行，一般台灣網頁常用字碼為big5和utf8。



```
<!DOCTYPE html PUBLIC "-//W3C//DTD HTML 4.01//EN"
"http://www.w3.org/TR/html4/strict.dtd">
<html class="konq safari safari13 not_ie4 not_ie5 not_ie55 not_ie6 not_ie7
not_ie5_6 not_ie55_6 not_ie6_7 not_ie5_7 not_ns4 not_opera not_gecko not_firefox15
not_khtml not_safari12 not_safari11 not_macie yui3-js-enabled" id=
"yui_3_5_1_13_1490111338752_9">
<div id="yui3-css-stamp" style="position: absolute !important; visibility: hidden
!important" class"></div>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=big5" />
<meta name="application-name" content="Yahoo!奇摩股市">
<meta name="msapplication-TileColor" content="#7b0099">
<meta name="msapplication-TileImage" content="https://s.yimg.com/f/i/tw/stock/
tw_stock_icon_pin.png">
<title>Yahoo!奇摩股市</title>
<link rel="stylesheet" type="text/css" href="https://s.yimg.com/f/i/tw/hp/masthead/
yhtw_masthead_standard.css" media="screen, print">
<link rel="stylesheet" type="text/css" href="https://s.yimg.com/f/i/tw/hp/masthead/
yhtw_masthead_12.css" media="screen, print">
<link rel="stylesheet" type="text/css" href="https://s.yimg.com/f/i/tw/stock/revampe/
css_1960_min.css" media="screen, print">
<link rel="stylesheet" type="text/css" href="https://s.yimg.com/f/i/tw/stock/img06/
stock_wd.css" media="screen, print">
<link rel="stylesheet" type="text/css" href="https://s.yimg.com/f/i/tw/stock/
stock_print_1_2.css" media="print">
<link href="https://s.yimg.com/f/i/tw/hp/Universal_header/uh_bt3.css" type="text/css"
rel="stylesheet">
<link href="https://s.yimg.com/f/i/tw/stock/revampe/vstock_uh_1_1_min.css" type=
"text/css" rel="stylesheet">
<link href="https://s.yimg.com/ud/others/stock/gemini-v1.css" type="text/css" rel=
"stylesheet">
<script async src="https://s.yimg.com/lq/lib/3pm/cs_0.2.js"></script>
<script src="https://pagead2.googlesyndication.com/pub-config/r20160913/ca-pub-
6785382607161389.js"></script>
<script type="text/JavaScript">...</script>
<style type="text/css">...</style>
<script language="JavaScript" src="https://s.yimg.com/f/i/tw/hp/wd/
wd_20050720.js"></script>
<!--
```


爬蟲範例-下載奇摩股市網股票價格資料

Step3. 下載網頁原始碼

```
爬取奇摩股票價格資料範例.R x
Source on Save
1 # R爬蟲範例
2 library(Rcurl) 使用Rcurl函數包
3
4 # 設定連結頁面
5 url <- "https://tw.stock.yahoo.com/q/q?s=1101"
6
7 # 將網頁原始碼載入R
8 html <- getURL(url,.encoding='big5')
```

爬取網頁原始碼<- getURL(url=爬取網頁網址, .encoding=網頁字碼)

```
Console C:/Users/Su-Yen-Ting/Desktop/R爬蟲教學/
> html
[1] "<!doctype html public "-//w3c//dtd html 4.01//en" "http://www.w3.org/tr/html4/strict.dtd">\n<html>\n<head>\n<meta http-equiv="Content-Type" content="text/html; charset=big5"> \n<meta name="application-name" content="Yahoo!奇摩股市"/>\n<meta name="msapplication-TileColor" content="#7b0099"/> \n<meta name="msapplication-TileImage" content="https://s.yimg.com/f/i/tw/stock/tw_stock_icon_pin.png"/>\n<title>Yahoo!奇摩股市</title>\n<link rel="stylesheet" type="text/css" href="https://s.yimg.com/f/i/tw/hp/masthead/yhtw_masthead_standard.css" media="screen, print" />\n<link rel="stylesheet" type="text/css" href="https://s.yimg.com/f/i/tw/hp/masthead/yhtw_masthead_12.css" media="screen, print" />\n<link rel="stylesheet" type="text/css" href="https://s.yimg.com/f/i/tw/stock/revampe/css_1960_min.css" media="screen, print" />\n<link rel="stylesheet" type="text/css" href="https://s.yimg.com/f/i/tw/stock/img06/stock_wd.css" media="screen, print" />\n<!-- <truncated>
```

爬蟲範例-下載奇摩股市網股票價格資料

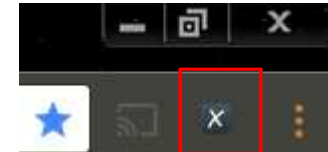
Step4. 取出數據

取出數據時要給程式一個路徑，讓程式知道我們想要的資料在哪裡，但網頁原始碼有時架構太複雜，要花很多時間找路徑，此時可利用Google的擴充工具-Xpath Helper解決這個問題

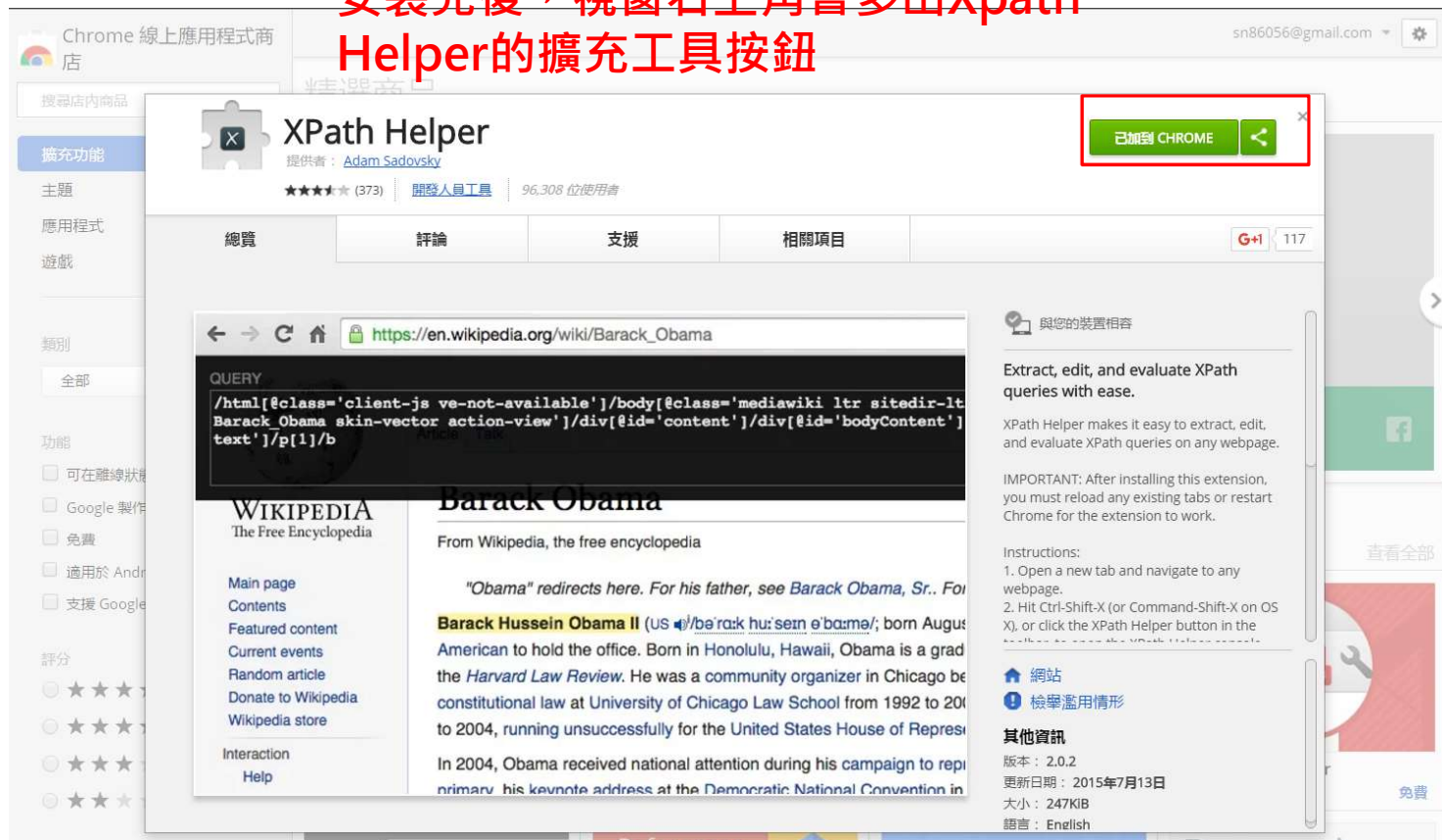


爬蟲範例-下載奇摩股市網股票價格資料

Step4. 取出數據



安裝完後，視窗右上角會多出Xpath
Helper的擴充工具按鈕



爬蟲範例-下載奇摩股市網股票價格資料

Step4. 取出數據

回到要爬取數據的頁面，按下ctrl+shift+x，即可啟動Xpath Helper工具



QUERY RESULTS (0)

首頁 投資組合 當日行情 大盤 類股 期權 港滬深股 美股 新聞 Y 選股

股票代號/名稱 1101 當日個股股價 查詢 2017/03/22 00:09 距離美股收盤還有3小時51分鐘 更新

凱基客戶專區：委託成交 庫存報價 資料日期: 106/03/21

股票代號	時間	成交	買進	賣出	漲跌	張數	昨收	開盤	最高	最低	個股資料
1101台泥 加到投資組合	14:30	38.40	38.35	38.40	△0.25	6,755	38.15	38.00	38.40	37.85	成交明細 技術 新聞 基本 籌碼 個股健診

凱基證券下單 ☐ 買 ☐ 賣 張 零股交易

爬蟲範例-下載奇摩股市網股票價格資料

Step4. 取出數據

QUERY

```
/html[@id='yui_3_5_1_13_1490114291912_9']/body[@id='yui_3_5_1_13_1490114291912_8']/center[@id='yui_3_5_1_13_1490114291912_7']/table[2]/tbody/tr/td/table/tbody/tr[2]/td[3]/b
```

RESULTS (1)

38.40

資料在原始碼內的路徑

首頁 投資組合 當口行情 大盤 類股 期權 港滬深股 美股 新聞 Y 選股

股票代號/名稱 1101 --當日個股股價-- 查詢 2017/03/22 00:09 距離美股收盤還有3小時51分鐘 更新

凱基客戶專區：委託成交 庫存報價 資料日期: 106/03/21

股票代號	時間	成交	買進	賣出	漲跌	張數	昨收	開盤	最高	最低	個股資料
1101台泥 加到投資組合	14:30	38.40	38.35	38.40	△0.25	6,755	38.15	38.00	38.40	37.85	成交明細 技術 新聞 基本 籌碼 個股健診

凱基證券下單 ☐ 買 ☐ 賣 張 送出 零股交易

將滑鼠指標移到有興趣的資料，按下shift鍵，Xpath Helper即會秀出這個資料在原始碼的路徑(查詢語句)。此處以1101台泥的成交價作範例。

爬蟲範例-下載奇摩股市網股票價格資料

Step4. 取出數據

QUERY 註冊用戶名? 請註冊 登入 服務說明 Download the new Yahoo Mail RESULTS (12) 信箱 拍賣 Yahoo 新聞

```
/html[@id='yui_3_5_1_13_1490114291912_9']/body[@id='yui_3_5_1_13_1490114291912_8']/center[@id='yui_3_5_1_13_1490114291912_7']/table[2]/tbody/tr/td/table/tbody/tr[2]/td
```

1101台泥加到投資組合
14:30
38.40
38.35
38.40

15+
守護你我永恆的愛

首頁 投資組合 當日行情 大盤 類股 期權 港滬深股 美股 新聞 Y選股

股票代號/名稱 1101 一一當日個股股價一一 查詢 2017/03/22 00:09 距離美股收盤還有3小時51分鐘 更新

凱基客戶專區：委託成交 庫存報價 資料日期: 106/03/21

股票代號	時間	成交	買進	賣出	漲跌	張數	昨收	開盤	最高	最低	個股資料
1101台泥 加到投資組合	14:30	38.40	38.35	38.40	△0.25	6,755	38.15	38.00	38.40	37.85	成交明細 技術 新聞 基本 籌碼 個股傳診

凱基證券下單 ☐ 買 ☐ 賣 張 零股交易

此處對查詢語句做「剪枝」，刪除td標籤後的語句，即可查詢到td標籤底下的所有資料。透過刪除部分語句的方法，可以讓我們一次取出想要的數據。

爬蟲範例-下載奇摩股市網股票價格資料

Step4. 取出數據

QUERY 是新用戶嗎? 請註冊 登入 服務說明 Download the new Yahoo Mail RESULTS (13) 信箱 拍賣 Yahoo 奇摩

//tr[2]/td

1101台泥加到投資組合
14:30
38.40
38.35
38.40

守護你我永恆的愛

首頁 投資組合 當日行情 大盤 類股 期權 港滬深股 美股 新聞 Y選股

股票代號/名稱 1101 -- 當日個股股價 -- 查詢 2017/03/22 00:09 距離美股收盤還有3小時51分鐘 更新

凱基客戶專區: 委託成交 庫存報價 資料日期: 106/03/21

股票代號	時間	成交	買進	賣出	漲跌	張數	昨收	開盤	最高	最低	個股資料
1101台泥 加到投資組合	14:30	38.40	38.35	38.40	Δ0.25	6,755	38.15	38.00	38.40	37.85	成交明細 技術 新聞 基本 籌碼 個股健診

凱基證券下單 ☐ 買 ☐ 賣 張 送出 零股交易

有時候路徑太長，可以將查詢語句開頭改為“//” 做精簡，這樣工具會去查詢原始碼內「所有」標籤底下有沒有符合我們的要求，並傳回對應的資料。

爬蟲範例-下載奇摩股市網股票價格資料

Step4. 取出數據



股票代號/名稱: 1101 台泥

股票代號	時間	成交	買進	賣出	漲跌	張數	昨收	開盤	最高	最低	個股資料
1101 台泥 加到投資組合	14:30	38.40	38.35	38.40	△0.25	6,755	38.15	38.00	38.40	37.85	成交明細 技術 新聞 基本 籌碼 個股健診

凱基證券下單: ☐ 買 ☐ 賣 張 零股交易

但查詢語句若簡略的太多，可能會多找出我們不要的資料。

爬蟲範例-下載奇摩股市網股票價格資料

Step4. 取出數據

```
爬取奇摩股票價格資料範例.R *
Source on Save
1 # R爬蟲範例
2 library(RCurl)
3 library(XML) 新增一個XML函數包，用來讀取XPath路徑找出我們要的數據
4
5 # 設定連結頁面
6 url <- "https://tw.stock.yahoo.com/q/q?s=1101"
7
8 # 將網頁原始碼載入R
9 html <- getURL(url,.encoding='big5')
10
11 # 轉換字碼
12 html <- iconv(html,'big5','utf-8') 將字碼轉換成UTF-8，方便程式處理
13
14 # 轉換為XML函數包要求的爬蟲格式
15 html <- htmlParse(html,encoding='utf-8') 將html轉換為XML函數包要求格式
16
17 # 利用XPath路徑表達式爬取資料
18 data <- xpathSApply(html, "//tr[2]/td", xmlValue)
19 我們想要的資料 <- xpathSApply(原始碼, 查詢語句, xmlValue)
20
```

Step4. 取出數據

```
Console C:/Users/Su-Yen-Ting/Desktop/R/爬蟲教學/ ↗
> data
[1] "" "1101台泥加到投資組合"
[3] "09:42" "38.00"
[5] "37.95" "38.00"
[7] "▽0.40\n" "1,179"
[9] "38.40" "37.95"
[11] "38.00" "37.75"
[13] "\n  成交明細技術 新聞基本 籌碼個股健診"
>
```

[illegible]

爬蟲範例-下載奇摩股市網股票價格資料

觀察網址規則



從奇摩股市網的網址可以觀察到，後面的1101是股票代號。
如果改成其他家股票代號，是不是會跳出別家股票的資訊？

爬蟲範例-下載奇摩股市網股票價格資料

觀察網址規則



The screenshot shows the Yahoo! Stock website interface. The URL in the address bar is `https://tw.stock.yahoo.com/q/q?s=2330`. The page displays a table of stock data for the ticker 2330 (Taiwan Power Corp.).

股票代號	時間	成交	買進	賣出	漲跌	張數	昨收	開盤	最高	最低	個股資料
2330台積電 加到投資組合	09:48	192.5	192.0	192.5	▼2.5	6,893	195.0	192.5	193.0	191.5	成交明細 技術 新聞 基本 籌碼 個股博覽

Below the table, there is a section for placing orders (凱基證券下單) with buttons for Buy (買), Sell (賣), and a field for quantity (張).

將網址的股票代號改成
2330，網頁確實跳出台
積電股票資訊

爬蟲範例-下載奇摩股市網股票價格資料

觀察網址規則

```
爬取奇摩股票價格資料範例.R x
Source on Save
1 # R爬蟲範例
2 library(RCurl)
3 library(XML)
4
5 # 設定股票代碼
6 stockCode <- 2330
7
8 # 設定連結頁面
9 url <- paste0("https://tw.stock.yahoo.com/q/q?s=", stockCode)
10
11 # 將網頁原始碼載入R
12 html <- getURL(url, .encoding='big5')
13
14 # 轉換字碼
15 html <- iconv(html, 'big5', 'utf-8')
16
17 # 轉換為XML函數包要求的爬蟲格式
18 html <- htmlParse(html, encoding='utf-8')
19
20 # 利用Xpath路徑表達式爬取資料
21 data <- xpathApply(html, "//tr[2]/td", xmlValue)
22
```

透過網址的規則，可將程式更加應用化，使用者

只要指定股票代碼，就可以抓出對應的股價資訊。

如果有股票代碼的清單，就可以透過迴

圈股票代碼來自動化下載全市場股票的

價格資訊。(可以自己動手試試看!)

爬蟲範例-下載奇摩股市網股票價格資料

定時爬取資料

另外也可以寫成每一分鐘自動去
抓股票價格資料的程式

程式執行結果

	股票代 碼	時間	開盤價	最高價	最低價	收盤價
recordData	2330	10:29	192.5	193.0	191.5	192.0
recordData	2330	10:30	192.5	193.0	191.5	192.5
recordData	2330	10:31	192.5	193.0	191.5	192.0
recordData	2330	10:32	192.5	193.0	191.5	192.0
recordData	2330	10:33	192.5	193.0	191.5	192.0
recordData	2330	10:34	192.5	193.0	191.5	192.0
recordData	2330	10:35	192.5	193.0	191.5	192.5
recordData	2330	10:36	192.5	193.0	191.5	192.0
recordData	2330	10:37	192.5	193.0	191.5	192.0
recordData	2330	10:38	192.5	193.0	191.5	192.0
recordData	2330	10:39	192.5	193.0	191.5	192.0
recordData	2330	10:40	192.5	193.0	191.5	192.5

```
RStudio Source Editor
爬取奇摩股票價格資料範例.R
Source on Save

1 # R爬蟲範例
2 library(RCurl)
3 library(XML)
4 |
5 # 設定股票代碼
6 stockCode <- 2330
7
8 # 紀錄資料表
9 recordTable <- matrix(0, nrow=0, ncol=6)
10 colnames(recordTable) <- c("股票代碼","時間","開盤價","最高價","最低價","收盤價")
11
12 # 當前時間
13 timeNow <- as.POSIXlt(Sys.time())
14
15 # 查詢執行時間起始值
16 queryTime <- Sys.time()
17
18 # 執行此程式到13:30收盤前
19 while((timeNow$hour*100+timeNow$min)<1330){
20
21   # 若查詢時間等於當前的時間，則進行查詢
22   if(Sys.time()>=queryTime){
23
24     # 提醒程式目前運作狀況
25     cat(paste0("在時間點:",queryTime,"，爬蟲程式啟動運作\n"))
26
27     # 設定連結頁面
28     url <- paste0("https://tw.stock.yahoo.com/q/q?s=",stockCode)
29
30     # 將網頁原始碼載入R
31     html <- getURL(url,.encoding='big5')
32
33     # 轉換字碼
34     html <- iconv(html,'big5','utf-8')
35
36     # 轉換為XML函數包要求的爬蟲格式
37     html <- htmlParse(html,encoding='utf-8')
38
39     # 利用Xpath路徑表達式爬取資料
40     data <- xpathApply(html, "//tr[2]/td", xmlValue)
41
42     # 紀錄資料
43     recordData <- c(stockCode, data[c(3,10:12,4)])
44     recordTable <- rbind(recordTable, recordData)
45
46     # 重設下次查詢的時間(一分鐘後)
47     queryTime <- Sys.time()+60
48   }
49
50   # 更新當前的時間
51   timeNow <- as.POSIXlt(Sys.time())
52 }
```

相關資料進一步研讀

這份教學PPT僅是最基礎的爬蟲教學，事實上有些網站並不是那麼好爬，對爬蟲有興趣的同學可以進一步去閱讀相關資料：

1. R 講題分享 - 用 R 進行中文 text Mining (作者:陳嘉葳@Taiwan R User Group)

<http://programmermagazine.github.io/201402/htm/article6.html>

2. R 講題分享 – SpideR -- 用R自製網路爬蟲收集資料 (作者：Taiwan R User Group)

<http://programmermagazine.github.io/201311/htm/article6.html>

3. 第一次爬蟲就上手 rvest_tutorial

備註：rvest 是R新出來的爬蟲套件，聽說

<https://rpubs.com/SatoshiLiang/159348>

更好上手，有興趣的人可以去玩玩看

4. [爬蟲] 觀察技巧 with Chrome開發人員工具

<http://tech-marsw.logdown.com/blog/2016/01/10/crawler-tips-mining-chrome>

5. [R crawler] 公開資訊觀測站 (觀察篇)

<http://leoluyi.logdown.com/posts/400371-crawler-mops-1>

教學結束 謝謝大家

