

R網路爬蟲

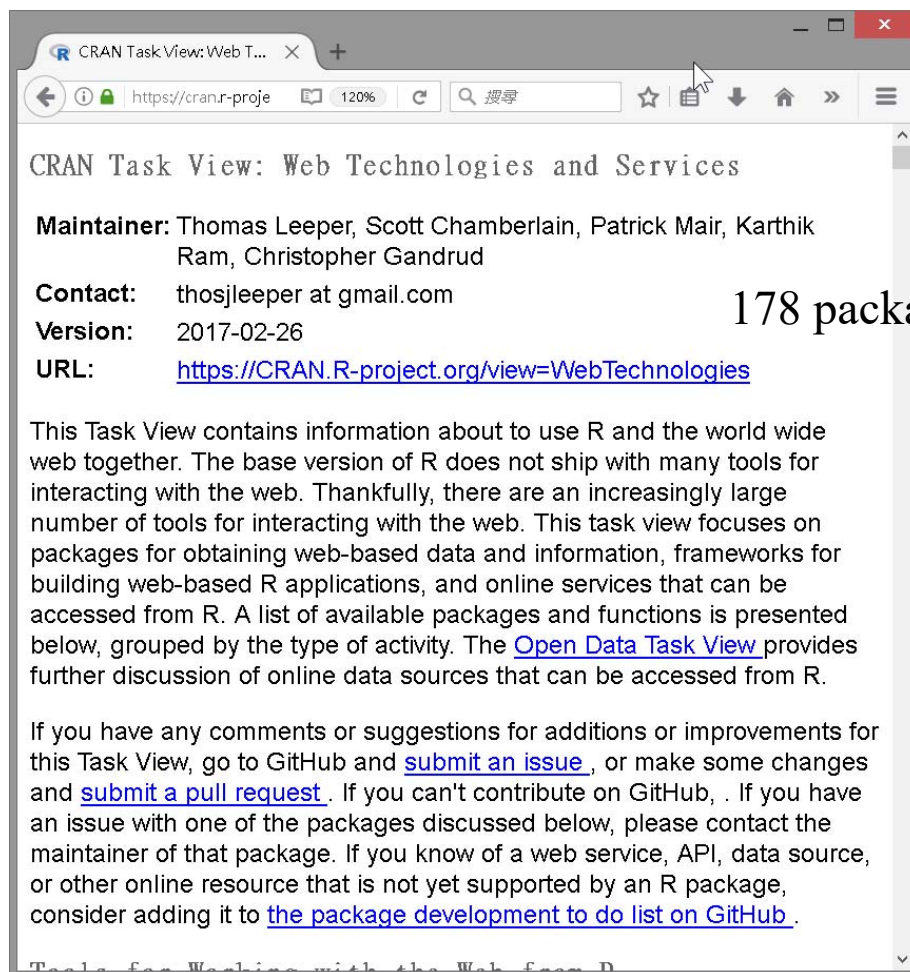
吳漢銘

國立政治大學 統計學系



<https://hmwu.idv.tw>

讀取網路資料: R網路爬蟲(Crawler)



注意事項:

- (1) 先了解網站對於資料的宣告及版權聲明。
- (2) 爬蟲程式是一種駭客行為(Hacking)。

Task View

- Tools for Working with the Web from R
 - Core Tools For HTTP Requests
 - Parsing Structured Web Data
 - Tools for Working with URLs
 - Tools for Working with Scraped webpage Contents
 - Other Useful Packages and Functions
- Web and Server Frameworks
- Web Services
 - Cloud Computing and Storage
 - Document and Code Sharing
 - Data Analysis and Processing Services
 - Social Media Clients
 - Web Analytics Services
 - Other Web Services

r-crawler

台大資工系統訓練班 R 語言與網站爬蟲的教學專案

[View on GitHub](#)

R 語言與網站爬蟲

簡介

這是郭耀仁在台灣大學資訊工程學系資訊系統訓練班所開設的 R 語言與網站爬蟲課程網頁，課程時數 10 小時。

評分標準

- 50% 出席率
- 50% 隨堂練習或作業

課程大綱

- 關於這門課
- R 語言的文字函數
- 正規表達式
- 認識 JSON 與 XML
- 認識 HTML 與 CSS
- readLines() 函數
- 使用 %>% 簡化程式
- rvest 套件
- CSS 與 Xpath 選擇
- 爬蟲練習
- 爬蟲練習 (2)

<https://yaojenkuo.io/r-crawler/>

Automated Data Collection with R

A Practical Guide to
Web Scraping and Text Mining



Simon Munzert | Christian Rubba | Peter Meißner | Dominic Nyhuis

Copyrighted Material

Use R!

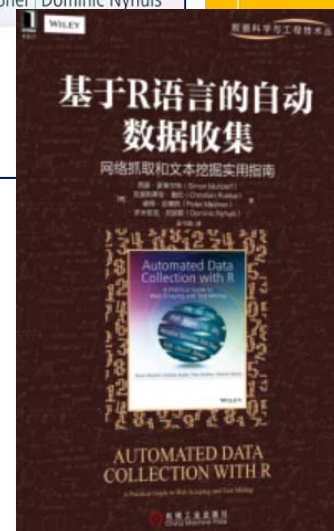
Copyrighted Material

Deborah Nolan
Duncan Temple Lang

XML and Web Technologies for Data Sciences with R

Springer

Copyrighted Material





Easily Harvest (Scrape) Web Pages

- Simple web scraping for R
- Create an html document from a url, a file on disk or a string containing html with `read_html()`.
- Select parts of a document using
 - css selectors: `html_nodes(doc, "table td")`
 - xpath selectors with `html_nodes(doc, xpath = "//table//td")`.
- Extract components with
 - `html_tag()`: the name of the tag,
 - `html_text()`: all text inside the tag,
 - `html_attr()`: contents of a single attribute,
 - `html_attrs()`: all attributes.
- Use `rvest` with XML files: parse with `xml()`, then extract components using `xml_node()`, `xml_attr()`, `xml_attrs()`, `xml_text()` and `xml_tag()`.
- Parse tables into data frames with `html_table()`.
- Extract, modify and submit forms with `html_form()`, `set_values()` and `submit_form()`.

<https://github.com/hadley/rvest>

href: Hypertext Reference
url: A Uniform Resource Locator



層疊樣式表 CSS (Cascading Style Sheets)

- 層疊樣式表 CSS (Cascading Style Sheets)是一種用來為結構化文件(如 HTML 文件或 XML 應用)添加樣式(字型、間距和顏色等)的電腦語言。(wiki)。有了 CSS，我們就可以將資料層及顯示層分開：HTML 文件就只包括資料，而 CSS 則是告訴瀏覽器這些資料應該要如何顯現出來。
- # is the id selector which is used to assign a class to a control having the same id as class name.
- . is the class selector which is used for assign class to the control with class attribute.

ID和 Class這兩者最大的不同，是在於 ID 選擇器在一個 HTML 文件中只能被使用一次，而 Class 選擇器在一個 HTML 文件中可以被使用多次

div 標籤目的是將內容分為不同的區域，而每一個區域可以根據 CSS 中的宣告而有自己的樣式。span和div類似，但span可以涵蓋更細層的元素，例如可以用 來調整單一文字的樣式。

```
<style type='text/css'>
#test {
    color: green;
}
.test1 {
    color:red;
}
</style>
```

```
<div id="test"> text style by #test </div>
<div class="test1"> text style by .test1 </div>
```

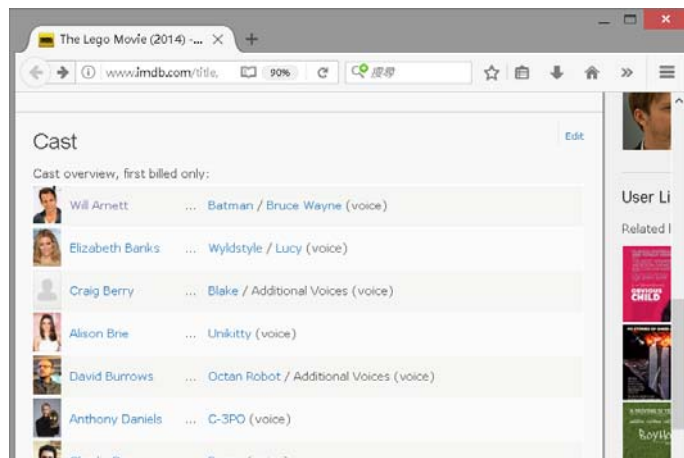
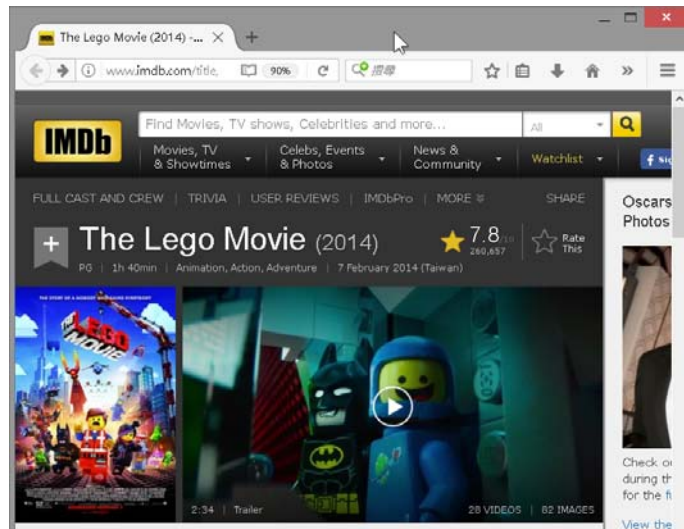
```
#IDName{
    attr: value;
    ...
}
```

```
.ClassName{
    attr: value;
    ...
}
```

CSS Tutorial: <https://www.w3schools.com/css/>

CSS 語法教學: <http://www.1keydata.com/css-tutorial/tw/>

Example 1: Demo



```
> install.packages("rvest")
> library(rvest)
> html <- read_html("http://www.imdb.com/title/tt1490017/")
> cast <- html_nodes(html, "span.itemprop")
> html_text(cast)
[1] "Animation"           "Action"
[3] "Adventure"           "Phil Lord"
[5] "Christopher Miller"  "Phil Lord"
[7] "Christopher Miller"  "Chris Pratt"
[9] "Will Ferrell"        "Elizabeth Banks"
[11] "Will Arnett"         "Elizabeth Banks"
[13] "Craig Berry"         "Alison Brie"
...
[27] "based on toy"        "dual personality"
[29] "good cop bad cop"    "evil businessman"
[31] "Warner Bros."        "Village Roadshow Pictures"
[33] "RatPac-Dune Entertainment"

> cast <- html_nodes(html, "#titleCast .itemprop")
> html_text(cast)
[1] "\n Will Arnett\n"      "
[2] "Will Arnett"
[3] "\n Elizabeth Banks\n"  "
[4] "Elizabeth Banks"
...
[29] "\n Jonah Hill\n"      "
[30] "Jonah Hill"

> cast <- html_nodes(html, "#titleCast span.itemprop")
> html_text(cast)
[1] "Will Arnett"           "Elizabeth Banks" "Craig Berry"
[4] "Alison Brie"           "David Burrows"  "Anthony Daniels"
[7] "Charlie Day"           "Amanda Farinos" "Keith Ferguson"
[10] "Will Ferrell"          "Will Forte"     "Dave Franco"
[13] "Morgan Freeman"       "Todd Hansen"    "Jonah Hill"
```

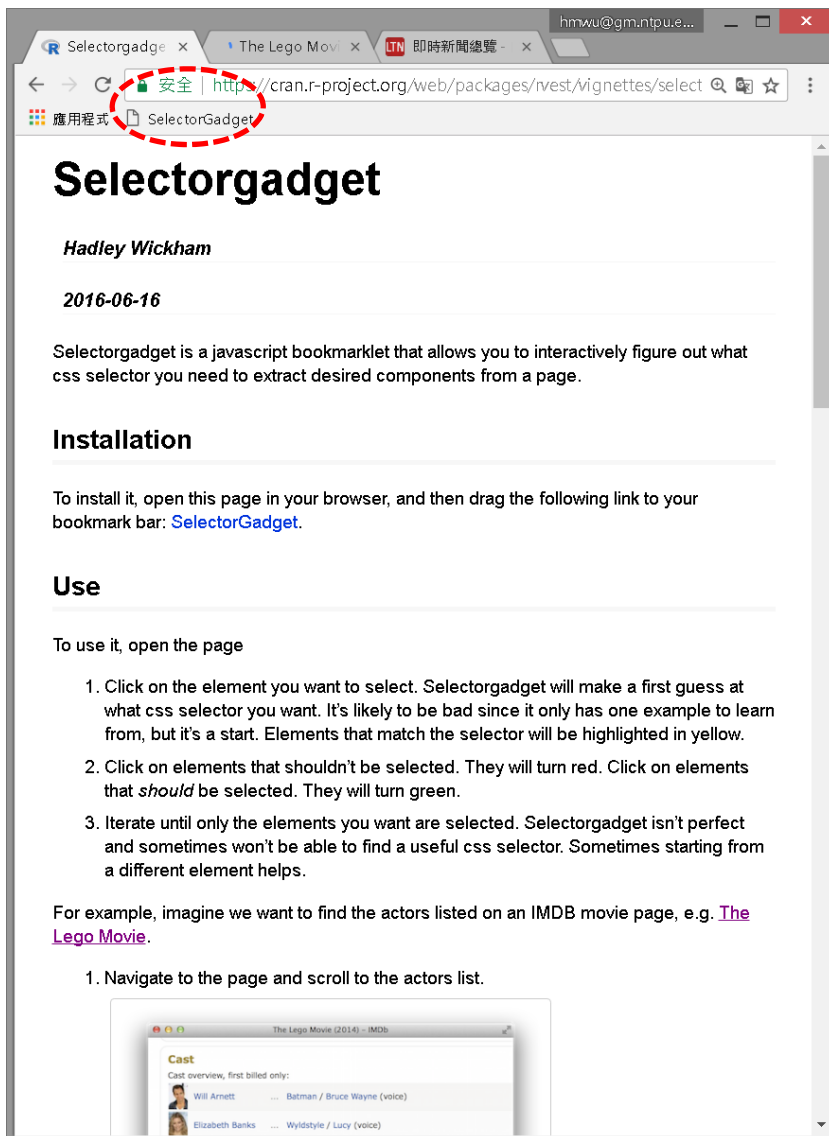
<https://cran.r-project.org/web/packages/rvest/vignettes/selectorgadget.html>

<https://hmwu.idv.tw>

如何查詢html_nodes(tag):

方法1: Selectorgadget

7/44



The screenshot shows the Selectorgadget website in a browser. The address bar shows the URL <https://cran.r-project.org/web/packages/rvest/vignettes/selectorgadget.html>. The page title is "Selectorgadget" by Hadley Wickham, dated 2016-06-16. The page describes Selectorgadget as a javascript bookmarklet for finding CSS selectors. It includes an "Installation" section with instructions to drag a link to the bookmark bar, and a "Use" section with a list of steps. A small inset image shows the tool being used on an IMDb page.

Selectorgadget

Hadley Wickham

2016-06-16

Selectorgadget is a javascript bookmarklet that allows you to interactively figure out what css selector you need to extract desired components from a page.

Installation

To install it, open this page in your browser, and then drag the following link to your bookmark bar: [SelectorGadget](#).


Use

To use it, open the page

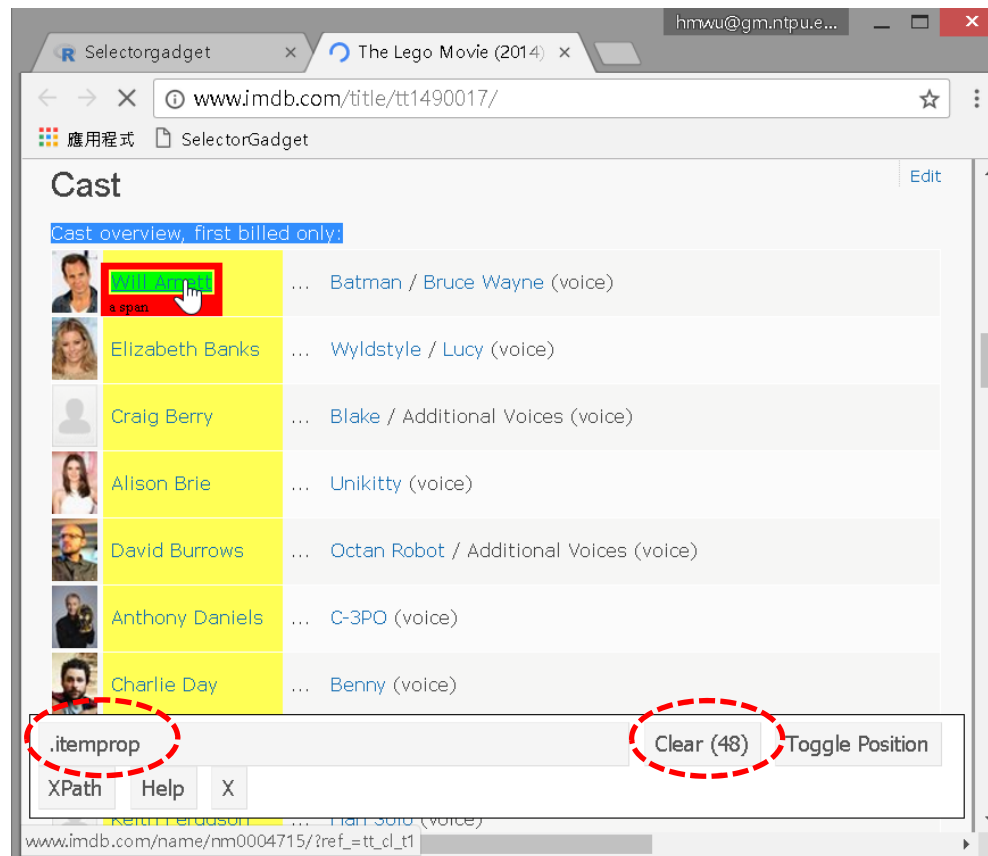
1. Click on the element you want to select. Selectorgadget will make a first guess at what css selector you want. It's likely to be bad since it only has one example to learn from, but it's a start. Elements that match the selector will be highlighted in yellow.
2. Click on elements that shouldn't be selected. They will turn red. Click on elements that *should* be selected. They will turn green.
3. Iterate until only the elements you want are selected. Selectorgadget isn't perfect and sometimes won't be able to find a useful css selector. Sometimes starting from a different element helps.

For example, imagine we want to find the actors listed on an IMDB movie page, e.g. [The Lego Movie](#).

1. Navigate to the page and scroll to the actors list.










(1) Click on the element you want to select. Elements that match the selector will be highlighted in **yellow**.



The screenshot shows the IMDb Cast page for "The Lego Movie (2014)". The "Cast" section is visible, listing actors and their roles. The "Cast overview, first billed only:" section is highlighted. The actors listed are Batman / Bruce Wayne (voice), Elizabeth Banks, Craig Berry, Alison Brie, David Burrows, Anthony Daniels, and Charlie Day. The first row, "Batman / Bruce Wayne (voice)", is highlighted in yellow. A red dashed circle highlights the "a span" text next to the actor's name. At the bottom, the ".itemprop" selector is shown in the input field, and the "Clear (48)" button is highlighted with a red dashed circle.

Cast

Cast overview, first billed only:

	Batman / Bruce Wayne (voice)
	Elizabeth Banks
	Craig Berry
	Alison Brie
	David Burrows
	Anthony Daniels
	Charlie Day

.itemprop

Clear (48) Toggle Position

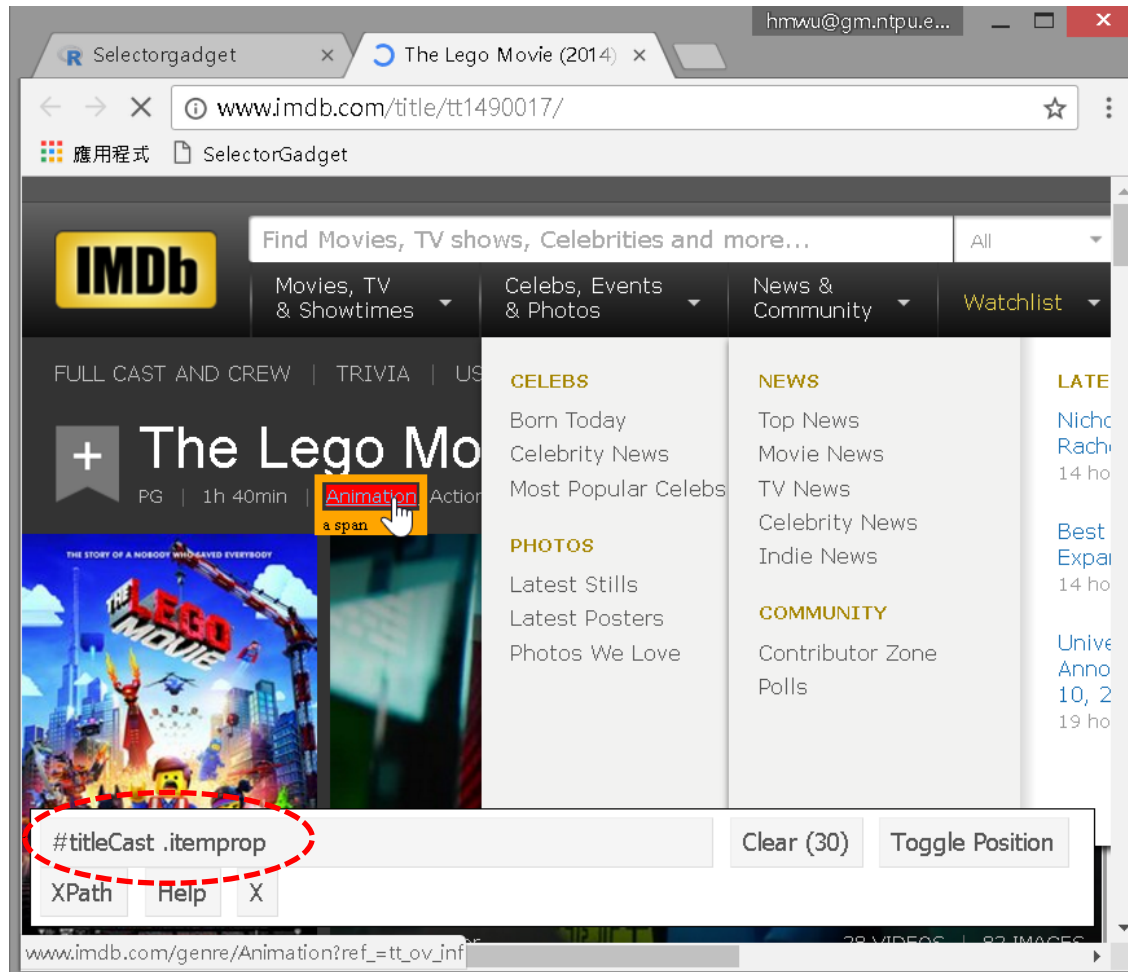
<https://cran.r-project.org/web/packages/rvest/vignettes/selectorgadget.html>

<https://hmwu.idv.tw>

如何查詢html_nodes(tag):

方法1: Selectorgadget

8/44



(2) Click on elements that shouldn't be selected. They will turn **red**. Click on elements that should be selected. They will turn **green**.

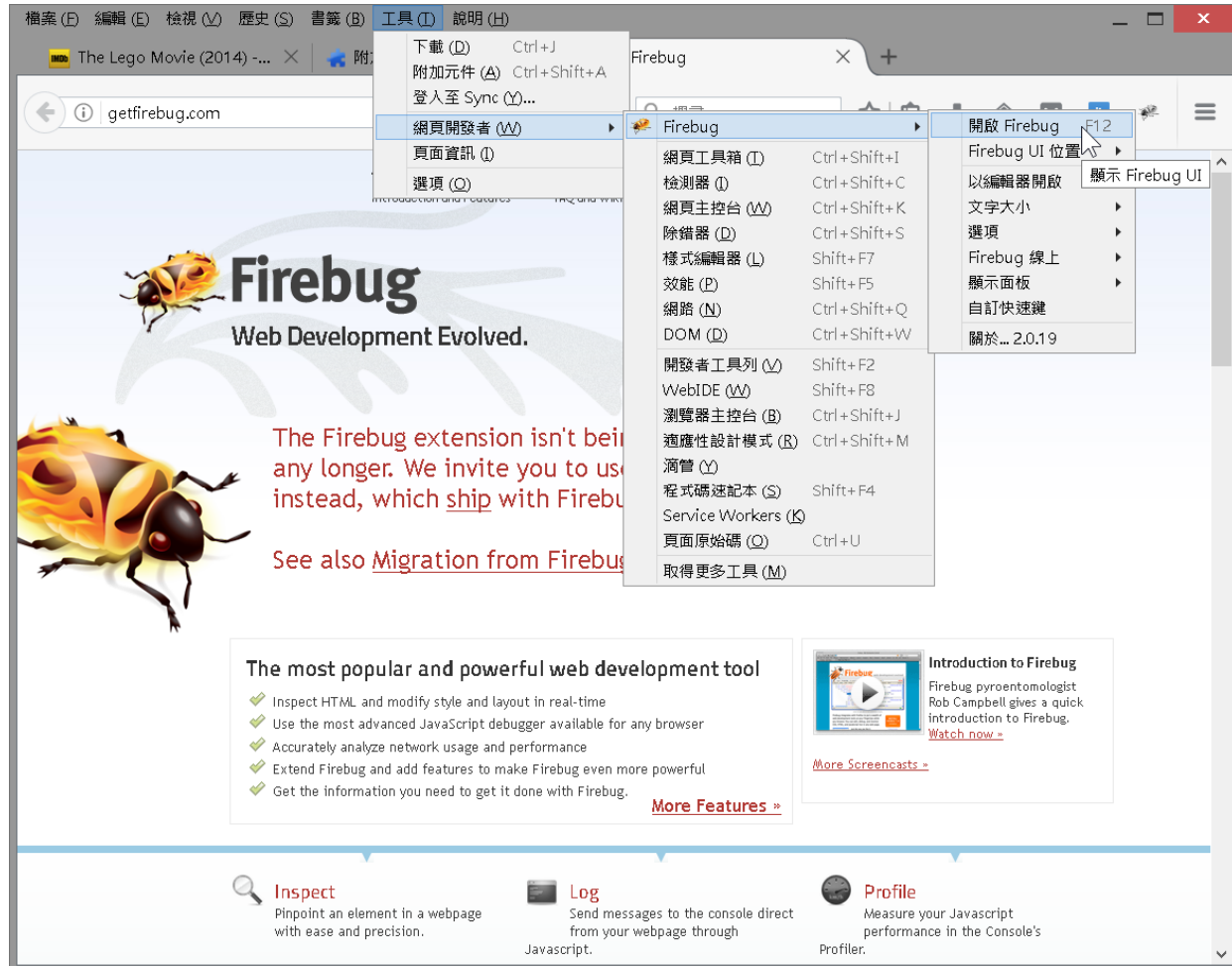
(3) Iterate until only the elements you want are selected. Selectorgadget isn't perfect and sometimes won't be able to find a useful css selector. Sometimes starting from a different element helps.

如何查詢html_nodes(tag):

方法2: firebug for firefox

9/44

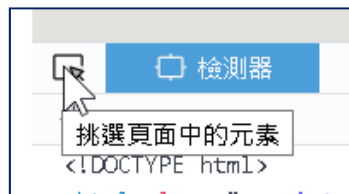
<http://getfirebug.com/>



如何查詢html_nodes(tag):

方法2: firebug for firefox

10/44



CSS ID

The Lego Movie (2014) ... X +

www.imdb.com/title/tt1490017/

Stars: John C. Reilly, Jack McBrayer,...

Cast

Will Arnett ... Batman / Bruce Wayne (voice)

Elizabeth Banks ... Wyldstyle / Lucy (voice)

Craig Berry ... Blake / Additional Voices (voice)

Alison Brie ... Unikitty (voice)

David Burrows ... Octan Robot / Additional Voices (voice)

Anthony Daniels ... C-3PO (voice)

檢測器 主控台 除錯器 {} 樣式編輯器 效能 記憶體 網路 <> DOM

搜尋 HTML

```
<script>/script>
<div id="titleCast" class="article">
  <h2>Cast</h2>
  <table class="cast_list">
    <tbody>
      <tr></tr>
      <tr class="odd">
        <td class="primary_photo"></td>
        <td class="itemprop" itemprop="actor" itemscope="" itemType="http://schema.org/Person">
          <a href="/name/nm0006969/?ref=tt_cl_t2" itemprop="url">
            <span class="itemprop" itemprop="name">Will Arnett</span>
          </a>
        </td>
        <td class="ellipsis">...</td>
        <td class="character"></td>
      </tr>
      <tr class="even">
        <td class="primary_photo"></td>
        <td class="itemprop" itemprop="actor" itemscope="" itemType="http://schema.org/Person">
          <a href="/name/nm0006969/?ref=tt_cl_t2" itemprop="url">
            <span class="itemprop" itemprop="name">Will Arnett</span>
          </a>
        </td>
        <td class="ellipsis">...</td>
        <td class="character"></td>
      </tr>
    </tbody>
  </table>
</div>
```

content-2-wide.flatland > div#main_bottommain > div#titleCastarticle > table.cast_list > tbody > tr.even > td.itemprop > a > span.itemprop

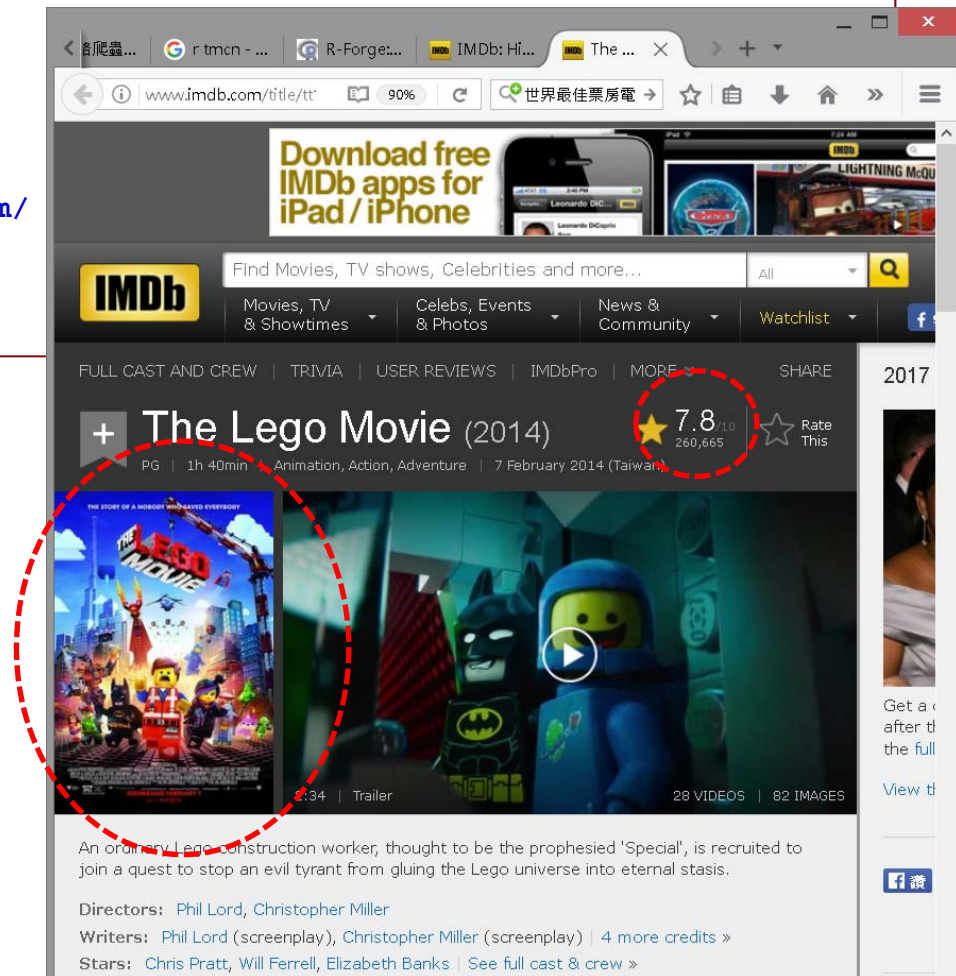
```
> lego.movie <- read_html("http://www.imdb.com/title/tt1490017/")
> rating <- html_nodes(lego.movie, "strong span") %>% html_text()
> rating
[1] "7.8"
> poster <- lego.movie %>%
+   html_nodes(".poster img") %>%
+   html_attr("src")
> poster
[1] "https://images-na.ssl-images-amazon.com/images/M/MV5BMTg4MDk1ODExN15BM15BanBnXkFtZTgwNzIyNjg3MDE@._V1_UX182_CR0,0,182,268_AL_.jpg"
```

NOTE:

InfoLite Extension for Google Chrome:

<https://chrome.google.com/webstore/detail/infolite/ipjbadabbpedegiekhgpiekdImfpgal>

可以在Chrome瀏覽器上直接選擇所要的區塊並顯示標籤。



Example 2: 抓取新聞標題

12/44

LTN 即時新聞總覽 - 自由時報...
news.ltn.com.tw/list/Bre...

自由時報
Liberty Times Net

Mapbox A MAPPING PLATFORM

台日對抗賽Live 打線熾火 中職聯隊8分差

中國爭奪228闡釋權 原來目的是這樣... [2017-03-01 21:38]
中國日前大動作紀念「二二八事件」，難者是中共的革命先烈，甚至是中華運動一部分，中國《南方周末》前新聞部主任、目前流亡至德國的作家長平指出，正因「二二八事件」後，台灣人期待自主統治與民主法...

艾瑪史東奧斯卡封后 爽擒獎座戰服跟著秋 2017-03-01 21:42
娛樂

(影音) 學霸偶像再+1! Chen成「EXO」第3位碩士 2017-03-01 21:42
娛樂

14年後回歸古裝喜劇片 梁朝偉竟喊痛苦? 2017-03-01 21:39
娛樂

台日對抗賽 陳傑惠猛打賞 日本隊4投壓制中職 2017-03-01 21:35
體育

江蕙和陳夫公然出雙入對 前妻再哭訴「一輩子的痛」 2017-03-01 21:34
娛樂

近期網友關注

- 1 台日對抗賽Live: 職聯隊3分勝日本
- 2 包場看格蕾的50道白的液體...
- 3 奧客插隊吃小籠包他...
- 4 為姚元浩失去所有曝光
- 5 女模被摸屁股 直揮巴掌

看新戲抽大獎 LTN好禮

市場動態

春季旅展必搶 慶泰飯店

http://news.ltn.com.tw/list/Br...
view-source: http://news.ltn.c... 80% 搜尋

sf="http://ent.ltn.com.tw/news/breakingnews/1990396" class="picword">艾瑪史東奧斯卡封后 爽擒獎座戰服跟著秋
href="http://ent.ltn.com.tw/news/breakingnews/1990357" class="picword">(影音) 學霸偶像再+1! Chen成「EXO」第3位碩士
href="http://ent.ltn.com.tw/news/breakingnews/1990359" class="picword">14年後回歸古裝喜劇片 梁朝偉竟喊痛苦?
href="http://sports.ltn.com.tw/news/breakingnews/1990387" class="picword">台日對抗賽 陳傑惠猛打賞 日本隊4投壓制中職
href="http://ent.ltn.com.tw/news/breakingnews/1990366" class="picword">江蕙和陳夫公然出雙入對 前妻再哭訴「一輩子的痛」
href="http://ent.ltn.com.tw/news/breakingnews/1990391" class="picword">好「獅」加持 老蕭潮到出水
href="http://sports.ltn.com.tw/news/breakingnews/1990062" class="picword">台日對抗賽Live 打線熾火 中職聯隊8分差
href="http://ent.ltn.com.tw/news/breakingnews/1990388" class="picword">大溪180年老建築「源古本舖」翻修化身藝術基地
href="http://news.ltn.com.tw/news/world/breakingnews/1990229" class="picword">妻翻丈夫童年舊照 驚訝發現自己也入鏡
href="http://ent.ltn.com.tw/news/breakingnews/1990351" class="picword">(影音) 林心如產後亮相 嬌戒吸煙
href="http://ent.ltn.com.tw/news/world/breakingnews/1990361" class="picword">柯林頓夫婦被爆當包租公 白宮林肯臥室1晚收
href="http://ent.ltn.com.tw/news/breakingnews/1990352" class="picword">諸星頂政大高學歷 宣告暫退演藝圈
href="http://news.ltn.com.tw/news/life/breakingnews/1990362" class="picword">改善交通 台南將廣設停車場、目標1萬個車位

news.ltn.com.tw/list/BreakingNews

應用程式 SelectorGadget

「點亮恆春古城」不夠亮 已完成燈光改善 2017-03-01 21:51
生活

艾瑪史東奧斯卡封后 爽擒獎座戰服跟著秋 2017-03-01 21:42
娛樂

(影音) 學霸偶像再+1! Chen成「EXO」第3位碩士 2017-03-01 21:42
娛樂

.picword Clear (24) Toggle Position

XPath Help X

entltn.com.tw/news/breakingnews/1990396

Example 2: 抓取新聞標題

13/44

```
> # install.packages("tmcn", repos="http://R-Forge.R-project.org")
> # "tmcn" 為Windows上用來處理中文字的套件。mac, linux上不需安裝
> library(tmcn)
> html <- read_html("http://news.ltn.com.tw/list/BreakingNews")
> news.title <- html_nodes(html, ".picword")
> news.title.utf8 <- toUTF8(html_text(news.title)) # convert to UTF8
> title.href <- html_attr(news.title, "href")
> my.news <- data.frame(title = news.title.utf8, href = title.href, stringsAsFactors=FALSE)
> my.news
```

title	href
1 艾瑪史東奧斯卡封后 爽擒獎座戰服跟著秋	http://news.ltn.com.tw/news/life/breakingnews/1990401
2 (影音) 學霸偶像再+1! Chen成「EXO」第3位碩士	http://ent.ltn.com.tw/news/breakingnews/1990396
3 14年後回歸古裝喜劇片 梁朝偉竟喊痛苦?	http://ent.ltn.com.tw/news/breakingnews/1990357
4 台日對抗賽》陳傑憲猛打賞 日本隊4投壓制中職	http://ent.ltn.com.tw/news/breakingnews/1990359
5 江蕙和腥夫公然出雙入對 前妻再哭訴「一輩子的痛」	http://sports.ltn.com.tw/news/breakingnews/1990387
6 好「獅」加持 老蕭潮到出水	http://ent.ltn.com.tw/news/breakingnews/1990366
7 台日對抗賽Live》打線熄火 中職聯隊8分差不敵日本	http://ent.ltn.com.tw/news/breakingnews/1990391
8 大溪180年老建築「源古本舖」翻修化身藝術基地	http://sports.ltn.com.tw/news/breakingnews/1990062
9 妻翻丈夫童年舊照 驚訝發現自己也入鏡	http://news.ltn.com.tw/news/life/breakingnews/1990388
10 (影音) 林心如產後亮相 婚戒曝光好搶鏡!	http://news.ltn.com.tw/news/world/breakingnews/1990229
11 柯林頓夫婦被爆當包租公 白宮林肯臥室1晚收1350萬	http://ent.ltn.com.tw/news/breakingnews/1990351
12 諧星頂政大高學歷 宣告暫退演藝圈1年	http://news.ltn.com.tw/news/world/breakingnews/1990361
13 改善交通 台南將廣設停車場、目標1萬個車位	http://ent.ltn.com.tw/news/breakingnews/1990352
14 8年級生成立「逃園」 為桃園畫新藍圖	http://news.ltn.com.tw/news/life/breakingnews/1990362
15 北橫公路明池路段 明、後天施工交管	http://news.ltn.com.tw/news/life/breakingnews/1990337
16 檢察官變行政官? 檢協會質疑司改會議有立場	http://news.ltn.com.tw/news/life/breakingnews/1990377
17 機捷明正式通車 站區商圈推好康搶客	http://news.ltn.com.tw/news/society/breakingnews/1990333
18 機捷串聯林口、新莊 建構新北影視廊帶	http://news.ltn.com.tw/news/life/breakingnews/1990375
19 中正大學「去蔣化」 副校長:難道要改成阿里山?	http://news.ltn.com.tw/news/life/breakingnews/1990334
20 台日對抗賽》恰恰2失誤全隊最多 中職共丟4分冤枉分	http://news.ltn.com.tw/news/life/breakingnews/1990347
21 (影音) 強國女歌手減肥五公斤 一度餓到眼冒金星	http://sports.ltn.com.tw/news/breakingnews/1990291
22 活化霜成分自「隕石萃取」? 衛生局:廣告誇大	http://ent.ltn.com.tw/news/breakingnews/1990376
23 直升機吊掛野放遇亂流 黑熊45公尺高空墜死	http://news.ltn.com.tw/news/society/breakingnews/1990349
24 香港特首選舉提名期結束 3人爭奪戰開打	http://news.ltn.com.tw/news/world/breakingnews/1990340

http://data-sci.info/2015/09/20/rvest_01/

課堂練習: appledaily

14/44

<http://www.appledaily.com.tw/realtimenews/section/new/>

The screenshot shows the Apple Daily website. The main headline is '世新大學為招陸生 簽一中承諾' (Shih Hsin University signs a one-year commitment to recruit mainland students). Below the headline, there is a list of news items for the date 2017/03/02. The items include: '女星變臉成癮 山根隆起鼻型再進化(21975)', '宏碁MR頭戴式顯示器開發者版本3月出貨(25)', '板橋中山國小改建校舍 新增地下停車場(44)', '馬英九談台獨：沒有一個國家宣佈獨立兩次(301)', '3字經謊高鐵車長 富商辯：口頭禪無罪(128)', and '小小童軍訪派出所 試乘重機大呼過癮(5)'.

```
> appledaily <- "http://www.appledaily.com.tw"
> url.main <- paste0(appledaily, "/realtimenews/section/new/")
> apple.news <- read_html(url.main)
>
> # news.rddt <- html_nodes(apple.news, '.rtdtdt')
> # news.time <- html_text(html_nodes(news.rddt, 'time'))
>
> news.time <- html_text(html_nodes(apple.news, '.rtdtdt time'))
> news.title <- html_text(html_nodes(apple.news, '.rtdtdt h1'))
> news.category <- html_text(html_nodes(apple.news, '.rtdtdt h2'))
> news.url <- html_attr(html_nodes(apple.news, '.rtdtdt a'), 'href')
> realtimenews <- data.frame(time=news.time, title=news.title,
+                             category=news.category, url=paste0(appledaily, news.url))
> realtimenews
```

	time	title	category
1	12:59	女星變臉成癮 山根隆起鼻型再進化(21975)	娛樂
2	12:58	宏碁MR頭戴式顯示器開發者版本3月出貨(0)	財經
3	12:56	板橋中山國小改建校舍 新增地下停車場(0)	生活
4	12:55	馬英九談台獨：沒有一個國家宣佈獨立兩次(0)	國際

```
1 http://www.appledaily.com.tw/realtimenews/article/local/20170302/1067249/大鵬灣漁民抗議臨時碼頭空間有限 無法生存
2 http://www.appledaily.com.tw/realtimenews/article/life/20170302/1067250/【有片】不滿遭影射「北港香爐」 美女教授張正琪將控告網路霸凌者
3 http://www.appledaily.com.tw/realtimenews/article/local/20170302/1067216/老翁遇魔神仔？ 15公里騎1天1夜
4 http://www.appledaily.com.tw/realtimenews/article/life/20170302/1067237/張正琪控台大高層要她閉嘴 台大這麼說
```

<https://hmwu.idv.tw>



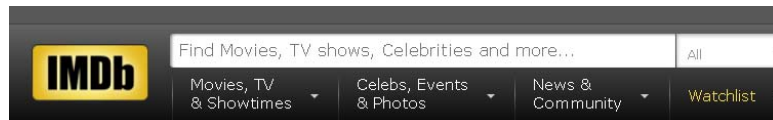
```
> get_content <- function(x){
+   tag <- html_node(read_html(x), '.trans')
+   text <- html_text(tag)
+   text
+ }
```

```
> get_content(as.character(realtimenews$url[1]))  
[1] "\r\n  
(新增影片)日本夯團「AKB48」25歲前成員板野友美，4年前退團單飛，但歌壇成績卻遠不如預期，反倒是一再進化的五官，持續成為注目焦點。日前她參加日本電視台節目《PON!》受訪，再被網友發現她的鼻型又有所不同，山根處明顯隆起，她從出道後一直進廠整修，若和當初兩相比較，根本已很難認出是同一...  
的山根處明顯隆起。翻攝電視畫面前田敦子（左）再傳新戀情，日前被拍到與30歲新男友約會過夜。翻攝《女性SEVEN》網站\r\n\t\t\t\t\t\t\tgoogletag.cmd.push(function(){googletag.display(\"InRead\");})\r\n\t\t\t\t\t\t\tgoogletag.cmd.push(function() {googletag.display('teadstv');})\r\n\t\t\t\t\t\t\tgoogletag.cmd.push(function() {googletag.display('goldenhorse');})\r\n\t\t\t\t\t\t\tgoogletag.cmd.push(function() {googletag.display('textlink');})\r\n有話要說 投稿「即時論壇」\r\n"
```

```
> url.c <- as.character(realtimenews$url)  
> news.content <- sapply(url.c, get_content)
```

課堂練習: IMDb

16/44



Highest Rated Adventure Feature Films With At Least 25000 Votes

1 to 50 of 866 titles | Next »

View Mode: Compact | Detailed

Sort by: Popularity | Alphabetical | **IMDb Rating** ▼ | Number of Votes | US Box Office | Runtime | Year | Release Date



1. The Lord of the Rings: The Return of the King (2003)

PG-13 | 201 min | Adventure, Drama, Fantasy

★ 8.9 ☆ Rate this 94 Metascore

Gandalf and Aragorn lead the World of Men against Sauron's army to draw his gaze from Frodo and Sam as they approach Mount Doom with the One Ring.

Director: Peter Jackson | Stars: Elijah Wood, Viggo Mortensen, Ian McKellen, Orlando Bloom

Votes: 1,276,504 | Gross: \$377.02M



2. Inception (2010)

PG-12 | 148 min | Action, Adventure, Sci-Fi

★ 8.8 ☆ Rate this 74 Metascore

A thief, who steals corporate secrets through use of dream-sharing technology, is given the inverse task of planting an idea into the mind of a CEO.

Director: Christopher Nolan | Stars: Leonardo DiCaprio, Joseph Gordon-Levitt, Ellen Page, Ken Watanabe

Votes: 1,551,811 | Gross: \$292.57M



3. The Lord of the Rings: The Fellowship of the Ring (2001)

PG-13 | 178 min | Adventure, Drama, Fantasy

★ 8.8 ☆ Rate this 92 Metascore

A meek Hobbit from the Shire and eight companions set out on a journey to destroy the powerful One Ring and save Middle Earth from the Dark Lord Sauron.

Director: Peter Jackson | Stars: Elijah Wood, Ian McKellen, Orlando Bloom, Sean Bean

Votes: 1,298,267 | Gross: \$313.84M



4. Star Wars: Episode V - The Empire Strikes

http://www.imdb.com/search/title?genres=adventure&sort=user_rating,desc&title_type=feature&num_votes=25000,&pf_rd_m=A2FGELUUNOQJNL&pf_rd_p=2406822102&pf_rd_r=11HHSKM8Z32G4X0Z4933&pf_rd_s=right-6&pf_rd_t=15506&pf_rd_i=top&ref=chttp_gnr_2

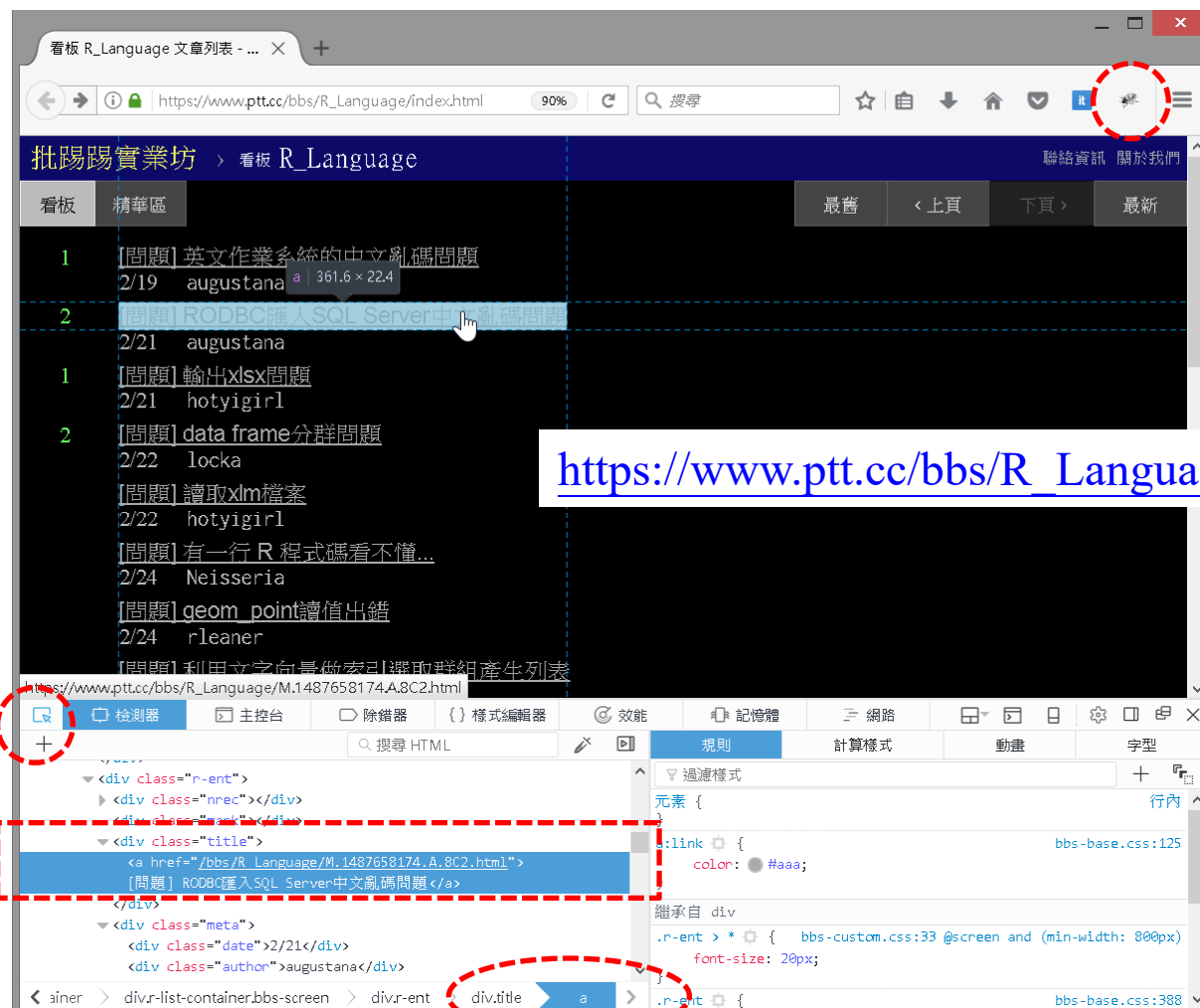


	A	B	C	D	E	F	G
1	Title	ReleaseDate	MPAARatings	Runtime	UserRatings	Votes	Gross
2							
3							
4							
5							
6							
7							
8							

Example 3: 爬「批踢踢實業坊」看板 R_Language目前的文章

17/44

- 讀取.title底下的a tag的 href屬性。
- 使用 `html_attr('href')`取得連結進行爬文章內容。
- 爬多個頁面的文章，使用for迴圈。
- 加`Sys.sleep(sample(3:5, 1))`讓爬蟲休息3~5秒，避免頻繁存取同個網站而遭阻斷。



https://www.ptt.cc/bbs/R_Language/index.html



Example 3: 爬「批踢踢實業坊」看板 R_Language目前的文章

18/44

```
> url.main <- 'https://www.ptt.cc/bbs/R_Language/index.html'
> href.title <- html_nodes(read_html(url.main), ".title a")
> R.hrefs <- html_attr(href.title, 'href')
>
> R.article.data <- c()
> for(i in 1:length(R.hrefs)){
+   article.url <- paste0('https://www.ptt.cc', R.hrefs[i])
+   article <- html_nodes(read_html(article.url), "#main-content")
+   article.content <- html_text(article)
+   article.utf8 <- iconv(article.content, 'utf8')
+   R.article.data <- c(R.article.data, article.utf8)
+   Sys.sleep(sample(3:5, 1))
+ }
> R.article.data
```

```
> R.article.data
[1] "作者augustana (微小的希望) 看板R_Language標題[問題] 英文作業系統的中文亂碼問題時間S
[2] NA
[3] "作者hotyigirl (yiju) 看板R_Language標題[問題] 輸出xlsx問題時間Tue Feb 21 20:48:06 2
[4] "作者locka (locka) 看板R_Language標題[問題] data frame分群問題時間Wed Feb 22 08:14:3

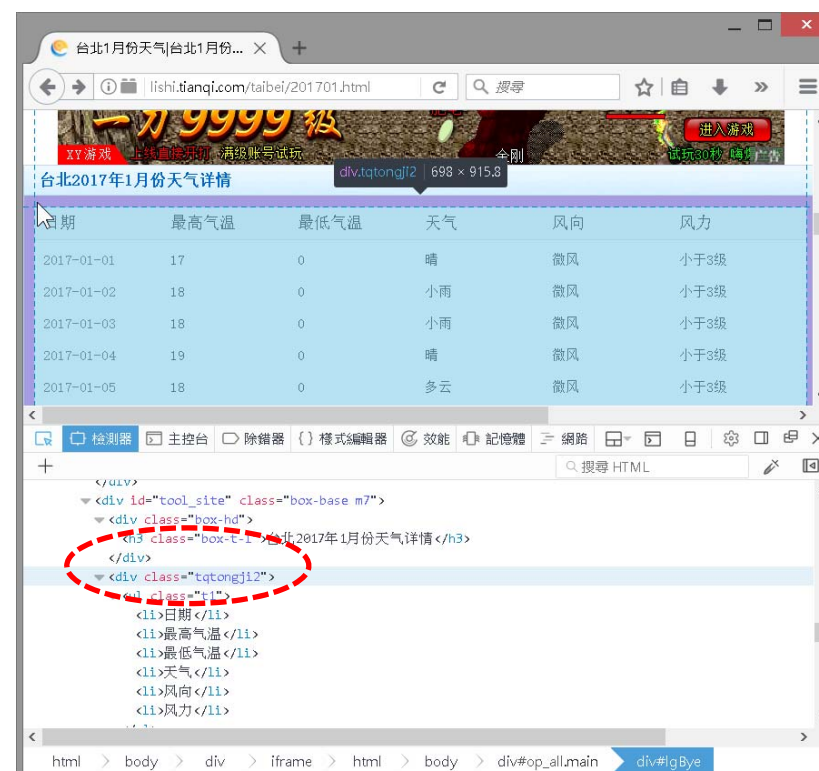
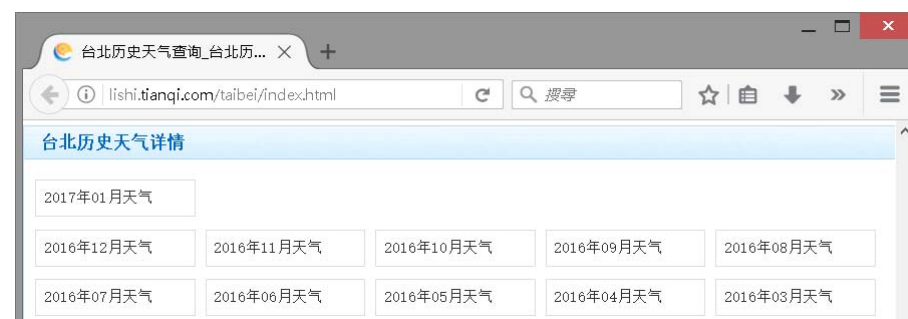
[16] "作者ecdysis (別煩..蝴蝶在飛) 看板R_Language標題[問題] 在用R語言作驗證式因素分析時無
[17] "作者Wush978 (拒看低質媒體) 看板R_Language標題[公告] 初版版務相關事項時間Thu Mar 28
[18] "好讀版: http://www.ptt.cc/bbs/R_Language/M.1382251406.A.957.html\n\n歡迎大家推文提
[19] "作者Wush978 (拒看低質媒體) 看板R_Language標題[情報] R語言翻轉教室時間Wed Apr 6 00:
>
```

`paste0(..., collapse)` is equivalent to `paste(..., sep = "", collapse)`, slightly more efficiently.

Example 4: 大量且有規則之網址

<http://lishi.tianqi.com/taibei/201701.html>

```
> # readHTMLTable {XML} failed
> url <-
getURL("http://lishi.tianqi.com/taibei/201701.html")
> htmlTable1 <- readHTMLTable(url, header=T)
> htmlTable1
named list()
```





Example 4: 大量且有規則之網址

```
library(tmcn)
# don't use for tables that contain missing values
get_tianqi_table <- function(city="taipei", year.from=2015, year.to=2016){

  month <- c(paste0("0", 1:9), "10", "11", "12")
  url.list <- c()
  for (year in year.from:year.to){
    url <- paste0("http://lishi.tianqi.com/taipei/", year, month, ".html")
    url.list <- c(url.list, url)
  }
```

```
tianqi.table <- NULL
for(x in url.list){
  # x <- "http://lishi.tianqi.com/taipei/201701.html"
  html <- read_html(x)
  wdata <- html_text(html_nodes(html, '.tqtongji2'))
  content <- toTrad(wdata)
  content.tmp <- str_replace_all(content, "[\\r\\n\\t]", "")
  content.tmp2 <- strsplit(str_trim(content.tmp), "\\s+")[[1]]
  tmp <- as.data.frame(matrix(content.tmp2[-(1:6)], ncol=6, byrow=T))
  colnames(tmp) <- content.tmp2[1:6]
  tianqi.table <- rbind(tianqi.table, tmp)
}
tianqi.table
```

```
> content
[1] "\\r\\n\\t\\t 日期\\r\\n\\t\\t 最高氣溫\\r\\n\\t\\t 最低氣溫\\r\\n\\t\\t 天氣\\r\\n\\t\\t 風向\\r\\n\\t\\t
```

```
library(stringr)
```

```
> get_tianqi_table(city="taipei", year.from=2016, year.to=2016)
```

	日期	最高氣溫	最低氣溫	天氣	風向	風力
1	2016-01-01	16	0	多雲	微風	小於3級
2	2016-01-02	17	0	中雨	微風	小於3級
3	2016-01-03	19	0	中雨	微風	小於3級
4	2016-01-04	19	0	小雨	微風	小於3級
5	2016-01-05	19	0	小雨	微風	小於3級
6	2016-01-06	16	0	小雨	微風	小於3級

```
> get_tianqi_table(city="taipei", year.from=2016, year.to=2016)
```



爬蟲練習

21/44

- PChome線上購物
<https://shopping.pchome.com.tw>
- Agoda旅館資訊
<https://www.agoda.com>
- 591租屋網
<https://rent.591.com.tw/new/>
- 台灣銀行的牌告匯率
<http://rate.bot.com.tw/xrt?Lang=zh-TW>
- YAHOO理財-國內基金
<https://tw.money.yahoo.com/fund/domestic>
- 台灣證券交易所
<http://www.twse.com.tw/ch/index.php>
- 證券櫃檯買賣中心
<http://www.tpex.org.tw/web/index.php?l=zh-tw>
- 公開資訊觀測站
<http://mops.twse.com.tw/mops/web/index>
- 國發會「物價資訊看板平台」
<http://price.nat.gov.tw>
- 自選

NOTE: 可能需要搭配其它R套件

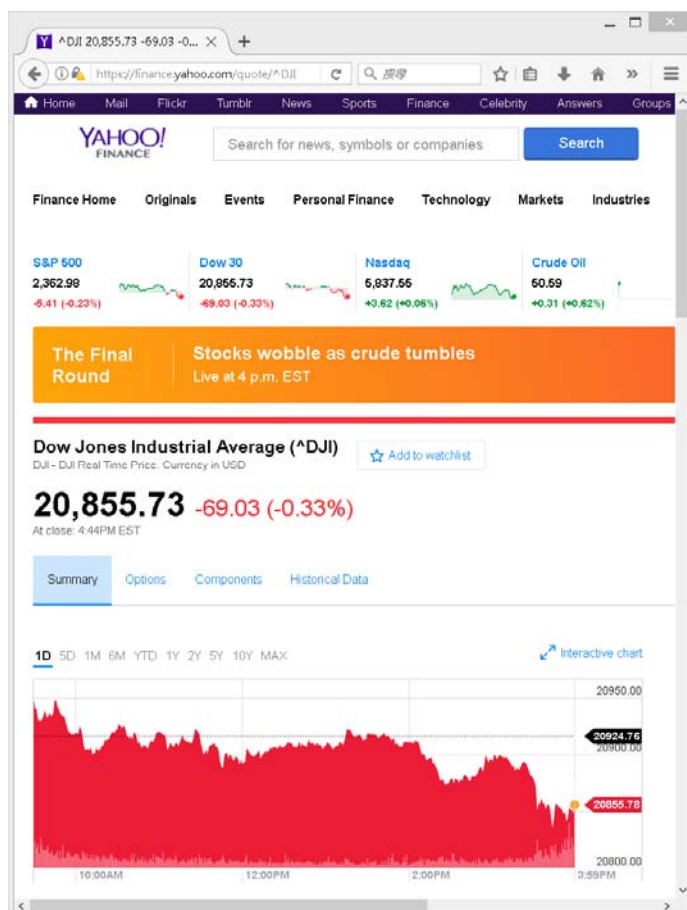
FACEBOOK

- <http://counterintuitiontaiwan.blogspot.tw/2017/03/facebook-r-crawler.html>
- <https://cran.r-project.org/web/packages/Rfacebook/Rfacebook.pdf>
- <http://jialin128.pixnet.net/blog/post/36644622-%5Br%5D-facebook-fanpage-crawler-in-r>

利用quantmod抓取股市資料

22/44

- **quantmod**: Quantitative Financial Modelling Framework
- **getSymbols {quantmod}**: Load and Manage Data from Multiple Sources
 - Current **src** methods available are: yahoo, google, MySQL, FRED, csv, RData, and oanda.
- <http://www.quantmod.com/>
- <http://statmath.wu.ac.at/~hornik/QFS1/quantmod-vignette.pdf>





利用quantmod抓取股市資料

23/44

```
> install.packages("quantmod")
> library(quantmod)
> getSymbols("AAPL", src="yahoo")
[1] "AAPL"
> # Data is loaded silently without user assignment by default.
> # The assigned variable name is that of the respective Symbols value.
> head(AAPL, 3)
      AAPL.Open AAPL.High AAPL.Low AAPL.Close AAPL.Volume AAPL.Adjusted
2007-01-03    86.29    86.58    81.90    83.80    309579900    10.85709
2007-01-04    84.05    85.95    83.82    85.66    211815100    11.09807
2007-01-05    85.77    86.20    84.40    85.05    208685400    11.01904
> # Open: 當日開盤價; High: 當日最高價; Low: 當日最低價; Close: 當日收盤價; Volume: 成交量
> # Adjust: 調整後股價; # 開高低收: OHLC
> # or
> myApple <- get(getSymbols("AAPL", src="yahoo"))
> head(myApple, 3)
      AAPL.Open AAPL.High AAPL.Low AAPL.Close AAPL.Volume AAPL.Adjusted
2007-01-03    86.29    86.58    81.90    83.80    309579900    10.85709
2007-01-04    84.05    85.95    83.82    85.66    211815100    11.09807
2007-01-05    85.77    86.20    84.40    85.05    208685400    11.01904
> class(AAPL)
[1] "xts" "zoo"
> # xts: eXtensible Time Series
> # zoo: S3 Infrastructure for Regular and Irregular Time Series (Z's Ordered Observations)
```

```
> AAPL.df <- as.data.frame(AAPL)
> class(AAPL.df)
[1] "data.frame"
```


畫趨勢圖

24/44

```
> chartSeries(AAPL)
> chartSeries(AAPL, subset='2016-09::2016-12')
> chartSeries(AAPL, subset='2016-09::2016-12', type = "line", theme = "white")
```





選取股市某時期的資料

25/44

```
> start.date <- as.Date("2012-02-01")
> end.date <- as.Date("2012-02-28")
> APPL.201202 <- AAPL[index(AAPL) > start.date & index(AAPL) < end.date]
> head(APPL.201202, 3)
      AAPL.Open AAPL.High AAPL.Low AAPL.Close AAPL.Volume AAPL.Adjusted
2012-02-02    455.90    457.17    453.98    455.12    46699100    58.96514
2012-02-03    457.30    460.00    455.56    459.68    71649900    59.55593
2012-02-06    458.38    464.98    458.20    463.97    62353200    60.11174
> tail(APPL.201202, 3)
      AAPL.Open AAPL.High AAPL.Low AAPL.Close AAPL.Volume AAPL.Adjusted
2012-02-23    515.08    517.83    509.50    516.39    142006900    66.90325
2012-02-24    519.67    522.90    518.64    522.41    103768000    67.68320
2012-02-27    521.31    528.50    516.28    525.76    136895500    68.11723
> AAPL.201604 <- get(getSymbols("AAPL", from = as.Date("2016-04-01"),
+                               to = as.Date("2016-04-30"), src="yahoo"))
> # note: "AAPL" is changed
> index(AAPL.201604)
[1] "2016-04-01" "2016-04-04" "2016-04-05" "2016-04-06" "2016-04-07" "2016-04-08"
...
[19] "2016-04-27" "2016-04-28" "2016-04-29"
```

```
> # multiple stocks
> getSymbols(c("VZ", "AAPL", "MMM", "IBM"))
[1] "VZ"    "AAPL" "MMM"   "IBM"
> head(MMM, 3)
      MMM.Open MMM.High MMM.Low MMM.Close MMM.Volume MMM.Adjusted
2007-01-03    77.53    78.85    77.38    78.26    3781500    59.92042
2007-01-04    78.40    78.41    77.45    77.95    2968400    59.68306
2007-01-05    77.89    77.90    77.01    77.42    2765200    59.27726
```



儲存股市資料

26/44

```
> getwd()
[1] "D:/myR"
> # 儲存資料(Rdata格式)
> # saveSymbols(file.path=".") # save all Symbols in the current directory
> saveSymbols(c("AAPL", "MMM"), file.path=".")
> list.files()
[1] "AAPL.RData"      "Data"      "MMM.RData"    "mysql-connector-odbc-5.3.7-winx64.msi"
> # 刪除此R session的Symbols資料 # removeSymbols() 刪除全部
> removeSymbols("AAPL")
> # 讀取在電腦端的資料(Rdata格式)
> load("AAPL.RData")
>
> # getSymbols取得的股票資料儲存成csv檔
> write.zoo(IBM, "IBM.csv", sep = ",", qmethod = "double") # or
> # 讀取股票資料csv檔進入R，成為xts類別物件 (假設工作目錄下已有IBM.csv檔案)
> removeSymbols("IBM") # > showSymbols()
> getSymbols.csv("IBM", env=globalenv())
[1] "IBM"
> head(IBM, 3)
      IBM.Open IBM.High IBM.Low IBM.Close IBM.Volume IBM.Adjusted
2007-01-03   97.18   98.40   96.26   97.27   9196800   77.73997
2007-01-04   97.25   98.79   96.88   98.31   10524500   78.57116
2007-01-05   97.60   97.95   96.91   97.42    7221300   77.85985
>
> # 以data.frame儲存成csv檔
> write.csv(as.data.frame(IBM), "IBM_2.csv")
> getSymbols.csv('IBM_2', env=globalenv())
[1] "IBM_2"
```

`showSymbols()`: # Objects loaded by getSymbols with auto.assign=TRUE can be viewed with showSymbols



下載台股股價歷史資料

27/44

- 下載股價資料前需知道台股的代碼，quantmod內建資料庫來源是yahoo finance.
- <https://finance.yahoo.com/>
- <https://tw.stock.yahoo.com/h/kimosel.php>
- <http://investexcel.net/all-yahoo-finance-stock-tickers/>
- http://teachercannon.weebly.com/uploads/6/8/3/8/6838442/smg_tickers.pdf
- <https://tw.stock.yahoo.com/h/getclass.php>

```
> symbols <- stockSymbols() # Get investment data from the internet
Fetching AMEX symbols...
Fetching NASDAQ symbols...
Fetching NYSE symbols...
> head(symbols)
```

	Symbol	Name	LastSale
1	AAMC	Altisource Asset Management Corp	76.6000
2	AAU	Almaden Minerals, Ltd.	0.9901
3	ABE	Aberdeen Emerging Markets Smaller Company Opportunities Fund I	12.0900
4	ACU	Acme United Corporation.	24.5700
5	ACY	AeroCentury Corp.	10.2000
6	ADGE	American DG Energy Inc.	0.3119

	MarketCap	IPOyear	Sector	Industry	Exchange
1	\$117.23M	NA	Finance	Real Estate	AMEX
2	\$89.16M	2015	Basic Industries	Precious Metals	AMEX
3	\$116.6M	NA	<NA>	<NA>	AMEX
4	\$81.69M	1988	Capital Goods	Industrial Machinery/Components	AMEX
5	\$15.98M	NA	Technology	Diversified Commercial Services	AMEX
6	\$15.81M	NA	Public Utilities	Power Generation	AMEX



抓取台積電(2330)資料並繪出股價圖

28/44

```
> TSMC <- getSymbols("2330.TW", auto.assign = FALSE)
> head(TSMC)
```

	2330.TW.Open	2330.TW.High	2330.TW.Low	2330.TW.Close	2330.TW.Volume	2330.TW.Adjusted
2007-01-02	67.4627	67.6617	66.9653	67.3632	26160600	47.3376
...						
2007-01-09	65.5722	66.1692	65.5722	65.6717	26187900	46.1489

```
> # Must use auto.assign=TRUE for multiple Symbols requests
> getSymbols(c("2330.TW", "2303.TW", "2337.TW"))
[1] "2330.TW" "2303.TW" "2337.TW"
> head(2330.TW) # error
錯誤: unexpected symbol in "head(2330.TW"
> head("2330.TW") # wrong
[1] "2330.TW"
> head('2330.TW') # wrong, Enter鍵旁的「'」
[1] "2330.TW"
> head(`2330.TW`) # correct, Tab 鍵上方的「`」
```

	2330.TW.Open	2330.TW.High	2330.TW.Low	2330.TW.Close	2330.TW.Volume	2330.TW.Adjusted
2007-01-02	67.4627	67.6617	66.9653	67.3632	26160600	47.3376
...						
2007-01-09	65.5722	66.1692	65.5722	65.6717	26187900	46.1489

```
>
> TSMC <- get("2330.TW") # or TSMC <- `2330.TW`
> UMC <- get("2303.TW")
> MXIC <- get("2337.TW")
> head(UMC) # View(UMC)
```

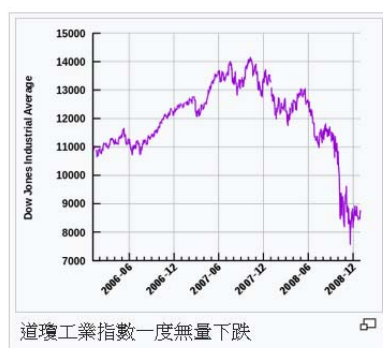
	2303.TW.Open	2303.TW.High	2303.TW.Low	2303.TW.Close	2303.TW.Volume	2303.TW.Adjusted
2007-01-01	20.2501	20.2501	20.2501	20.2501	0	20.6975
...						
2007-01-08	19.9002	20.0001	19.7001	19.7001	41413700	20.1353

抓取台積電(2330)資料並繪出股價圖

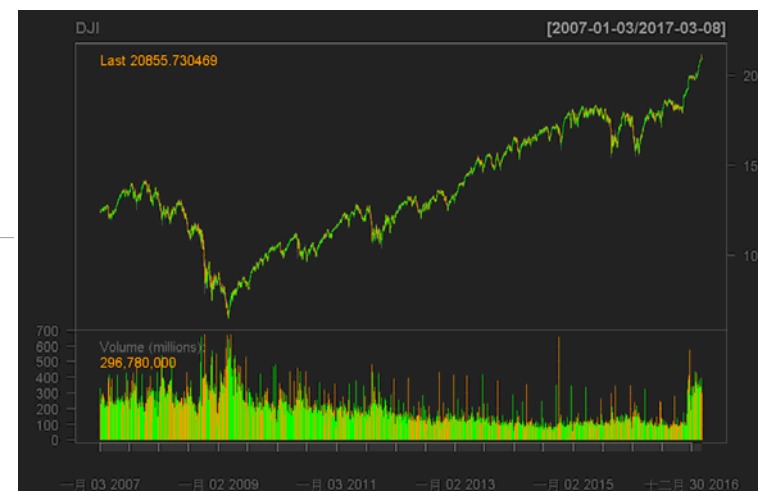
chartSeries(TSMC)



chartSeries(UMC)



getSymbols("^DJI")
chartSeries(DJI)



本文介紹的是2000年代後期發生後延續至今的全球金融危機。關於由此引發的經濟危機，請見「[經濟大衰退](#)」。

2007年–2008年全球金融危機（英語：Financial crisis of 2007–08），又稱2008年世界金融危機、次貸危機、信用危機、2008年華爾街金融危機、2008年金融崩潰，在2008年又出現了金融海嘯及華爾街海嘯等名稱，是一場在2007年8月9日開始浮現的金融危機。自次級房屋信貸危機爆發後，投資者開始對抵押證券的價值失去信心，引發流動性危機。即使多國中央銀行多次向金融市場注入巨額資金，也無法阻止這場金融危機的爆發。直到2008年9月^[1]，這場金融危機開始失控，並導致多間相當大型的金融機構倒閉或被政府接管，並引發經濟衰退。



日、月、年資料

30/44

```
> TSMC.2012.daily <- getYahooData("2330.TW", start = 20120101, end = 20121231)
> head(TSMC.2012.daily)
      Open      High      Low      Close      Volume Unadj.Close Div Split Adj.Div
2012-01-02 64.85687 64.85687 64.85687 64.85687         0       75.8   NA    NA     NA
...
2012-01-09 64.34349 64.68575 63.83012 64.68575 32288430       75.6   NA    NA     NA
> TSMC.2012.monthly <- getYahooData("2330.TW", start = 20120101, end = 20121231,
+                                   freq = "monthly", adjust = F)
> head(TSMC.2012.monthly)
      Open High Low Close  Volume
2012-01-02 75.8 79.5 74.6  78.5 32195800
...
2012-06-01 81.7 81.7 76.1  81.3 42518200
> # https://finance.yahoo.com/world-indices
> # e.g., S&P 500 (^GSPC), Dow Jones Industrial Average (^DJI)
> # 抓取台灣加權指數 TSEC weighted index (^TWII) 歷史資料
> getSymbols("^TWII", src="yahoo")
[1] "TWII"
> head(TWII)
> chartSeries(TWII)
```



See also: R-Code Yahoo Finance Data Loading

<https://www.r-bloggers.com/r-code-yahoo-finance-data-loading/>

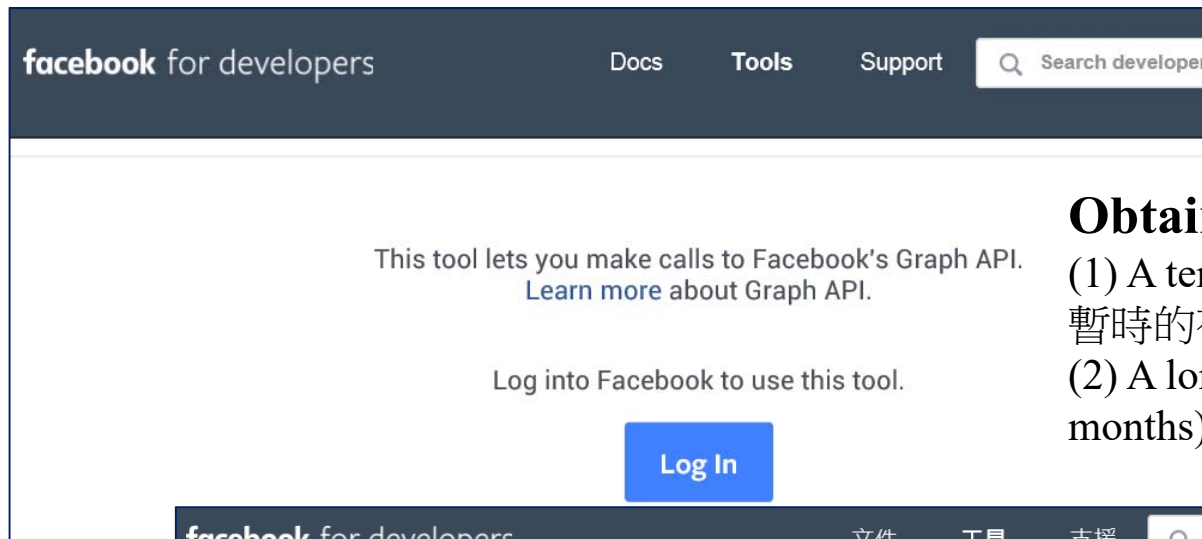


Rfacebook: Access to Facebook API via R^{31/44}

<https://github.com/pablobarbera/Rfacebook>

<https://cran.r-project.org/web/packages/Rfacebook/Rfacebook.pdf>

Get a temporary token here: <https://developers.facebook.com/tools/explorer>



Obtain an access token:

- (1) A temporary token (valid for two hours): 暫時的存取權杖(token)。
- (2) A long-lived token (valid for two months): 使用 facebook app 的方式。



選擇要開放的權限

32/44

選擇權限 v3.0 ✕

用戶數據權限

☐ email
☐ user_age_range
☐ user_birthday
☐ user_friends
☐ user_gender

☐ user_hometown
☐ user_likes
☐ user_link
☐ user_location
☐ user_photos

☐ user_posts
☐ user_status
☐ user_tagged_places
☐ user_videos

活動、社團和粉絲專頁

☐ ads_management
☐ ads_read
☐ business_management
☐ groups_access_member_info
☐ manage_pages
☐ pages_manage_cta

☐ pages_manage_instant_articles
☐ pages_messaging
☐ pages_messaging_payments
☐ pages_messaging_phone_number
☐ pages_messaging_subscriptions
☐ pages_show_list

☐ publish_pages
☐ publish_to_groups
☐ read_page_mailboxes
☐ user_events
☐ user_managed_groups

其他

☐ instagram_basic
☐ instagram_manage_comments

☐ instagram_manage_insights
☐ read_audience_network_insights

☐ read_insights

預設已包含公開的個人檔案。

取得存取權杖 清除

使用 Facebook 帳號登入 - Internet Explorer

https://www.facebook.com/v3.0/dialog/oauth?response_type=token&display=popup&client_id=14563499550

使用 Facebook 帳號登入



Graph API Explorer 將收到：
姓名和大頭貼照、朋友名單、生日、收到的讚、性別和年齡範圍。

[編輯](#)

以漢銘的身分連結

取消

這不會讓應用程式在 Facebook 上發佈貼文

應用程式條款 · 隱私政策

取得FB朋友資料

33/44

```
> # from CRAN
> install.packages("Rfacebook")
```

```
> # from GitHub
> install.packages("devtools")
> library(devtools)
> install_github("pablobarbera/Rfacebook/Rfacebook")
```

```
> library(Rfacebook)
> token <- 'XXXXX'
> me <- getUsers("me", token, private_info=TRUE)
> me
```

	id	name	username	first_name	middle_name	last_name	gender	locale	likes
picture	birthday	location	hometown	relationship_status					
1	796069860405544	Han-Ming Wu	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA				

```
>
> my.friends <- getFriends(token)
Only friends who use the application will be returned
See ?getFriends for more details
```



摘要FB朋友資料

34/44

```
> head(my.friends)
      id      name username first_name middle_name last_name gender locale likes
1  684393172      NA      Day      male    en_US      NA
2 10152206734717861      NA      劉      male    zh_TW      NA
3  778224889      NA     Chiu      male    zh_TW      NA
4  1071791267      NA     Chen      male    zh_TW      NA
5 10204043583107964      NA     Liu      male    zh_TW      NA
6  1458094397
```

```
1 https://scontent.xx.fbcdn.net/v/t1.0-1/c0.33.200.200/18
2 https://lookaside.facebook.com/pla
3 https://scontent.xx.fbcdn.net/v/t1.0-1/p200x200/31790543_101
4 https://scontent.xx.fbcdn.net/v/t1.0-1/p200x200/21731372_102
5 https://lookaside.facebook.com/pla
6 https://scontent.xx.fbcdn.net/v/t1.0-1/p200x200/14469700_102
  birthday      location      hometown relationship_status
1  <NA>      <NA>      <NA>      NA
2  <NA>      <NA>      <NA>      NA
3  03/08 Taipei, Taiwan      Chiayi      NA
4  <NA>      <NA>      <NA>      NA
5  <NA>      <NA>      <NA>      NA
6  <NA> Taipei, Taiwan Taipei, Taiwan      NA
```

朋友權限

▸ 為什麼要停用可取得用戶朋友相關資訊的權限？

▸ 在 **API 2.0** 版中，我無法取得某位已登入應用程式用戶的完整朋友名單，是否有方法可以取得完整朋友名單？

- 使用 Graph API 2.0 版及以上版本呼叫「/me/friends」，只會回傳用戶的朋友中，也使用這個應用程式的用戶。用戶的這些朋友也必須授予「user_friends」權限。如果想要讓用戶可以在應用程式發佈的動態中標註朋友，您可以使用「朋友標註功能 API」（[/docs/graph-api/reference/user/taggable_friends](https://developers.facebook.com/docs/graph-api/reference/user/taggable_friends)）。
- 如果想要邀請用戶使用應用程式，則視所建立的應用程式類型和適用平台而定，我們提供數種解決方案。如需詳細資訊，請參閱「邀請朋友」相關問題（https://developers.facebook.com/docs/apps/faq#invite_to_app）。

```
> table(my.friends$gender)
```

```
female  male
      3    17
```

```
> # my.friends$locale
```

```
> table(substr(my.friends$locale, 1, 2)) # language
```

```
en ja zh
  4  1 15
```

```
> table(substr(my.friends$locale, 4, 5)) # country
```

```
JP TW US
  1 15  4
```

```
> table(my.friends$hometown)
```

```
Chiayi      Hong Kong      Hsichih Kaohsiung, Taiwan      Taichung, Taiwan      Taipei, Taiwan
  1              1              1              1              4
```

```
> my.friends$locale
```

```
[1] "en_US" "zh_TW" "zh_TW" "zh_TW" "zh_TW" "zh_TW" "zh_TW" NA
[8] "en_US" "en_US" "zh_TW" "zh_TW" "zh_TW" "zh_TW" "zh_TW" "zh_TW"
[15] "zh_TW" "en_US" "zh_TW" "zh_TW" "ja_JP" "zh_TW" "zh_TW"
```

取得「活動、社團和粉絲專頁」權限^{35/44}

圖形 API 測試工具

應用程式: [?] Graph API Explorer

存取權杖: 貼到現有的「存取權杖」, 或是點擊「取得用戶存取權杖」

GET → /v3.0 /me?fields=id,name

提交

深入瞭解圖形 API 語法

選擇權限

v3.0

用戶數據權限

<input type="checkbox"/> email	<input type="checkbox"/> user_hometown	<input type="checkbox"/> user_posts
<input type="checkbox"/> user_age_range	<input type="checkbox"/> user_likes	<input type="checkbox"/> user_status
<input type="checkbox"/> user_birthday	<input type="checkbox"/> user_link	<input type="checkbox"/> user_tagged_places
<input type="checkbox"/> user_friends	<input type="checkbox"/> user_location	<input type="checkbox"/> user_videos
<input type="checkbox"/> user_gender	<input type="checkbox"/> user_photos	

活動、社團和粉絲專頁

<input checked="" type="checkbox"/> ads_management	<input checked="" type="checkbox"/> pages_manage_instant_articles	<input checked="" type="checkbox"/> publish_pages
<input checked="" type="checkbox"/> ads_read	<input checked="" type="checkbox"/> pages_messaging	<input checked="" type="checkbox"/> publish_to_groups
<input checked="" type="checkbox"/> business_management	<input checked="" type="checkbox"/> pages_messaging_payments	<input checked="" type="checkbox"/> read_page_messaging
<input checked="" type="checkbox"/> groups_access_member_info	<input checked="" type="checkbox"/> pages_messaging_phone_number	<input checked="" type="checkbox"/> user_events
<input checked="" type="checkbox"/> manage_pages	<input checked="" type="checkbox"/> pages_messaging_subscriptions	<input checked="" type="checkbox"/> user_managed_groups
<input checked="" type="checkbox"/> pages_manage_cta	<input checked="" type="checkbox"/> pages_show_list	

其他

<input type="checkbox"/> instagram_basic	<input type="checkbox"/> instagram_manage_insights	<input type="checkbox"/> read_insights
<input type="checkbox"/> instagram_manage_comments	<input type="checkbox"/> read_audience_network_insights	

預設已包含公開的個人資料。

取得存取權杖

使用 Facebook 帳號登入 - Mozilla Firefox

https://www.facebook.com/v3.0/dialog/oauth?response_type=token

使用 Facebook 帳號登入

Graph API Explorer 想要管理你的廣告、存取你的 Facebook 廣告和相關統計資料、管理你的企業管理平台、進入你管理的社團、管理你的粉絲專頁、管理粉絲專頁上的行動呼籲、管理粉絲專頁的即時文章、為你管理的粉絲專頁在 Messenger 銷售產品和接受付款、從你管理的粉絲專頁傳送訊息、使用用戶電話號碼, 從你管理的粉絲專頁傳送訊息、有了第一次用戶互動之後, 就可以隨時從你管理的粉絲專頁傳送訊息、顯示你管理的粉絲專頁清單、以你管理的粉絲專頁發佈、access your Facebook Pages' messages、publish to your groups on your behalf 和 access your public information on content in your groups。

選擇你要允許的權限

稍後再說 確定

切換存取權杖

36/44

文件 工具 支援 我的應用程式 搜尋 developers.facebook.com

圖形 API 測試工具 應用程式: [?] Graph API Explorer

存取權杖: EAACEdEose0cBAPM9Fu6fugWYzCA2y1FYTQuXHJKFoZCZAZCQj5jtb5YBNSwDfXN4EviORjnaZA5zyfezXaDQ6Xot

GET → /v3.0 /me?fields=id,name

淡江大學 數學系 系友會

- 返回用戶權杖
- 要求 publish_pages
- 解除安裝應用程式

粉絲專頁存取權杖

- Department of Mathematics, Tamkang University
- Hmwu-page
- 淡江大學 數學系 系友會

Access to Facebook API via R



Documentation for package 'Rfacebook' version 0.6.15

• [DESCRIPTION file.](#)

Help Pages

Rfacebook-package	Access to Facebook API via R
callAPI	Make an API request
fbOAuth	Create OAuth token to Facebook R session
getCheckins	Extract list of checkins of a Facebook friend
getCommentReplies	Extract replies to comments on page post
getEvents	Extract list of events from a public Facebook page or group
getFQL	Executes a FQL query to the Facebook Graph API
getFriends	Extract list of friends with their information
getGroup	Extract list of posts from a public Facebook group
getInsights	Extract Insights metric from a Facebook page (admin role required)

[getLikes](#)

[getNetwork](#)

[getNewsfeed](#)

[getPage](#)

[getPageToken](#)

[getPost](#)

[getReactions](#)

[getShares](#)

[getUsers](#)

[Rfacebook](#)

[searchFacebook](#)

[searchGroup](#)

[searchPages](#)

[updateStatus](#)

Extract list of likes of a Facebook friend

Extract network of friends of authenticated user

Download recent posts from the authenticated user's newsfeed

Extract list of posts from a public Facebook page

Get a page access token

Extract information about a public Facebook post

Extract total count of reactions to one or more Facebook posts

Extract list of users who publicly shared a public Facebook post

Extract information about one or more Facebook users

Access to Facebook API via R

Search public posts that mention a string

Find Facebook ID of a group

Search pages that mention a string

Update Facebook status from R

建立 Facebook App

37/44

facebook for developers

產品 文件 更多 開始使用

https://developers.facebook.com/

1 註冊 — 2 第一款應用程式 — 3 跟我們介紹一下你自己吧

歡迎使用 Facebook for Developers

1 建立 Facebook for Developers 帳號

下一步

一旦繼續，就代表你同意接收 Facebook 傳送的行銷相關電子通訊，包括新聞、活動、最新消息和促銷電子郵件。你可以隨時到「開發者工具」中的「行銷偏好」進行調整。

文件 工具 支援 我的應用程式 搜尋 developers.facebook.com

2 建立新的應用程式

facebook for developers 文件 工具 支援 我的應用程式 搜尋 developers.facebook.com

RTestApp 應用程式編號：187128678586686 關閉 狀態：調整中 查看分析工具

4

3 建立新的應用程式編號

開始將 Facebook 整合到你的應用程式或網站

顯示名稱

RTestApp

聯絡電子郵件

hanming.wu@gmail.com

一旦繼續，就代表你同意 Facebook 開放平台政策

取消 建立應用程式編號

Account Kit 順利建立帳號，無須使用密碼。 閱讀文件 設定

Facebook 登入 世界第一社交登入產品。 閱讀文件 設定

Analytics 瞭解用戶如何在應用程式、裝置、平台和網站中與你的企業互動。 閱讀文件 設定

Messenger 自訂你和用戶在 Messenger 上的互動方式。 閱讀文件 設定



獲取應用程式編號及應用程式密鑰^{38/44}

oposers 文件 工具 支援 我的應用程式 搜尋 developers.facebook.com

RTestApp 應用程式編號: 364039657416919 關閉 狀態: 調整中 查看分析工具

主控板

設定

基本資料

進階

角色

提示

應用程式審查

產品

應用程式編號

364039657416919

應用程式密鑰

..... 顯示

顯示名稱

RTestApp

命名空間

應用程式網域

聯絡電子郵件

hanming.wu@gmail.com

隱私政策網址

「登入」對話方塊和應用程式詳細資料的隱私政策

服務條款網址

「登入」對話方塊和應用程式詳細資料的服務條款

應用程式圖示 (1024 x 1024)

1024 x 1024

類別

選擇類別

到這裡瞭解更多有關應用程式類別的資訊

Business Use

This app uses Facebook tools or data to

Support my own business

Provide services to other businesses

+ 新增平台



新增平台

39/44

選擇平台



Facebook 網頁遊戲



網站



iOS



Android



Windows 8



網站

立即開始



網站網址

http://localhost:1410/

+ 新增平台

捨棄

儲存變更

連結FB

40/44

```
> fb.oauth <- fbOAuth(  
+   app_id="364039657416919",  
+   app_secret="5ala9h...aa618f0056ca82be",  
+   extended_permissions = TRUE)  
Copy and paste into Site URL on Facebook App Settings:  
When done, press any key to continue...  
Waiting for authentication in browser...  
Press Esc/Ctrl + C to abort  
Authentication complete.  
Error in init_oauth2.0(self$endpoint, self$app, scope = "  
  Bad Request (HTTP 400). Failed to get an access token  
此外: Warning message:  
In strsplit(rawToChar(raw), "\r?\n") : 輸入的字串 1 不  
> |
```

錯誤



Insecure Login Blocked: You can't get an access token or log in to this app from an insecure page. Try re-loading the page as https://

確定

檔案 (F) 編輯 (E) 檢視 (V) 歷史 (S) 書籤 (B) 工具 (I) 說明 (H)

localhost:1410/?error_code=1349 X +

← → ↻ 🏠 ⓘ localhost:1410/?error_code=13 ... 🛡️ ☆ 🔍 搜尋 ⬇️ ⏏️ ☰

Authentication complete. Please close this page and return to R.

```
> # save your fb_oauth object and use it for the next time  
> save(fb.oauth, file="fb.oauth")  
> load("fb.oauth")  
> me <- getUsers("me", token=fb.oauth)  
> my.likes <- getLikes(user="me", token=fb.oauth)
```

Rfacebook授權失敗

41/44

RTestApp

應用程式編號：364039657416919

關閉

狀態：調整中

查看分析工具

主控板

設定

角色

提示

應用程式審查

產品 +

Facebook 登入

設定

快速入門

用戶端 OAuth 設定

是

用戶端 OAuth 登入

啟用標準 OAuth 用戶端權杖流程。透過以下選項來鎖定允許哪些權杖重新導向 URI，可保護應用程式安全並防止濫用。如果不使用，則可以全域停用。 [?]

是

網路 OAuth 登入

啟用網頁型用戶端 OAuth 登入。 [?]

是

強制採用 HTTPS

強烈推薦對重新導向 URI 和 JavaScript SDK 強制採用 HTTPS。 [?]

否

強制網路 OAuth 重新驗證

開啟時，會提醒用戶輸入 Facebook 密碼才能登入網頁。 [?]

否

嵌入的瀏覽器 OAuth 登入

是

對重新導向 URI 使用 Strict 模式

只允許使用 Facebook SDK 或完全符合有效 OAuth 重新導向 URI 的 URI。 [?]

有效的 OAuth 重新導向 URI

http://localhost:1410/ x

否

從裝置登入

針對智慧型電視等裝置啟用 OAuth 用戶端登入流程 [?]

捨棄

儲存變更

自由時報

Liberty Times Net

即時新聞 影音 娛樂 汽車 時尚 體育 3C 評論 玩咖

3C科技

智慧手機 實用秘技 電腦應用 網路社群 好攝相機 家電娛樂 科技趣

3C科技 > 網路社群

網路社群

資訊安全

臉書

臉書逾 5千萬用戶個資遭盜 歐盟要查 股價重挫

即時新聞 / 綜合報導 / 2018-03-20 09:48



外媒驚爆臉書 (Facebook) 逾5000萬用戶的個資遭分析公司不當擷取，而該分析公司是由2016美國總統川普競選團隊所聘的，值得注意的是，臉書的股價19日開盤後更因此大跌逾5%；另外，歐盟司法事務執委喬羅瓦 (Vera Jourova) 表示，她預計會在本週出訪美國期間，要求臉書公司說明。

2016美國總統川普競選團隊所聘的分析公司「劍橋分析」 (Cambridge Analytica)，被媒體爆出不當擷取逾5000萬臉書用戶的個資，以用來開發、設計來預測或影響選民投票選擇的軟體。

FB 粉絲專頁

42/44



查詢粉絲專頁的FB ID：
<https://lookup-id.com>



```
str(fb.pchome24)
'data.frame':      100
obs. of 10 variables:
 $ from_id      : chr
 $ from_name    : chr
 $ message      : chr
 $ created_time : chr
 $ type         : chr
 $ link         : chr
 $ id           : chr
 $ likes_count  : num
 $ comments_count : num
 $ shares_count : num
```

```
> fb.pchome24 <- getPage(page="pchome24h", token=fb.oauth, n=100)
> # fb.pchome24 <- getPage(page="203507002993446", token=fb.oauth,
    since='2016/01/01', until='2016/01/05')

> fb.pchome24
> write.csv(fb.pchome24, "fb.pchome24.csv")
```

查詢粉絲專頁的FB ID：

43/44

facebook for developers 文件 工具 支援 我的應用程式 搜尋 developers.facebook.com

RTestApp 應用程式編號：3640396

主控板 設定 基本資料 進階 角色 提示

應用程式編號 364039657416919

顯示名稱 RTestApp

圖形 API 測試工具
測試、建立和驗證 API 呼叫，並為回應除錯。

存取權杖工具
產生用戶和應用程式存取權杖以進行測試。

分享偵錯工具
預覽內容在 Facebook 分享時的外觀。

查看所有工具

關閉 狀態：調整中

facebook for developers 文件 工具 支援 我的應用程式 搜尋 developers.facebook.com

圖形 API 測試工具 應用程式：[?] Graph API Explorer

存取權杖：EAACEdEose0cBAKiNgB0XQJ3mTXdBNCcmWURWIDYNzvn161q7nMbFCrfMPEhDoCbkaJfKv02iEHImuQObbEaHjXl7tArNATJJ6m(取得權杖

GET → /v3.0 / pchome24h?fields=id,name 提交

Node: pchome24h

☒ id
☒ name
+ 搜尋欄位

```
{
  "id": "203507002993446",
  "name": "PChome 24h購物"
}
```

深入瞭解圖形 API 語法

商用工具

44/44


Microsoft | Power BI 登入 免費註冊

開始使用 Power BI Desktop

更少資料準備，更多資料讀嘆

使用自由形式拖放畫布、現代化資料視覺效果和簡單的報表撰寫，視覺化地探索資料。

[免費下載 >](#)



開始共用您的資料視覺效果和深入資訊

POWER BI 雲端共同作業與共用

POWER BI 報表伺服器 內部部署報表部署



dcplus 數位行銷實戰家 主題課程 行銷知識

首頁 / 市集 / 找工具 / 工具市集 / Fanpage Karma – 一步完整分析的 Facebook 粉絲團經營工具

fanpage karma

Fanpage Karma – 一步完整分析的 Facebook 粉絲團經營工具

Fanpage Karma，能分析Facebook 粉絲團，Twitter、Instagram等，操作簡單，並提供完善的分析報告，各時段貼文表現、粉絲喜好、廣告價值、競品分析，還有最新功能，結合Pinterest、Google Analytics，連IKEA都在用，你還不用嗎？絕對讓你徹底掌握粉絲團喔！

fanpage karma ANALYTICS ENGAGE PUBLISH MONITORING TEAMBOARD PRICING ELON EXTRAS LOGIN

The Allround-Tool for strong Social-Media Management.

Analyze and improve your Social Media Profiles

With Fanpage Karma, you can not only analyze your profiles and those of the competition, but also manage all your customer conversations in one tool - even composing, planning and publishing posts. Try it now and make customers and fans happy even faster!

[TRY IT NOW](#) [LEARN MORE](#)

