

GO1

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)



CRAN Task View: Web Technologies and Services

Maintainer: Thomas Leeper, Scott Chamberlain, Patrick Mair, Karthik Ram, Christopher Gandrud

Contact: thosjleeper at gmail.com

Version: 2017-02-26

URL: <https://CRAN.R-project.org/view=WebTechnologies>

178 packages

This Task View contains information about how to use R and the world wide web together. The base version of R does not ship with many tools for interacting with the web. Thankfully, there are an increasingly large number of tools for interacting with the web. This task view focuses on packages for obtaining web-based data and information, frameworks for building web-based R applications, and online services that can be accessed from R. A list of available packages and functions is presented below, grouped by the type of activity. The [Open Data Task View](#) provides further discussion of online data sources that can be accessed from R.

If you have any comments or suggestions for additions or improvements for this Task View, go to GitHub and [submit an issue](#), or make some changes and [submit a pull request](#). If you can't contribute on GitHub, . If you have an issue with one of the packages discussed below, please contact the maintainer of that package. If you know of a web service, API, data source, or other online resource that is not yet supported by an R package, consider adding it to [the package development to do list on GitHub](#).

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 Scrapped 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

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)。



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

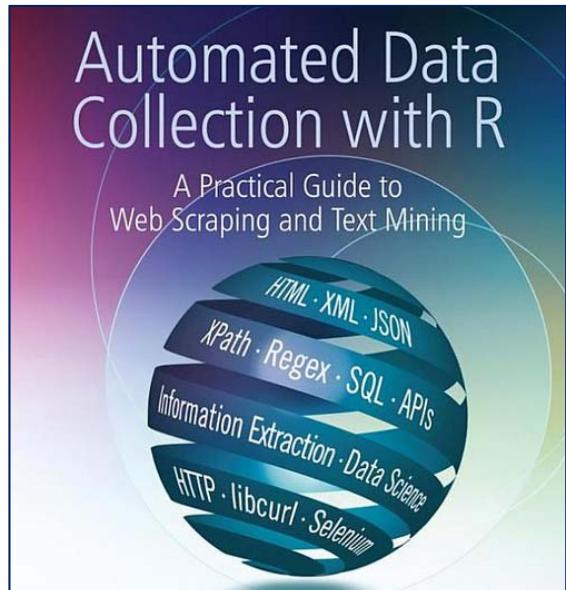
這是郭耀仁在台灣大學資訊工程學系資訊系統訓練班所開設的 R 語言與網站爬蟲課程網頁，課程時數 10 小時。

評分標準

- 50% 出席率
- 50% 隨堂練習或作業

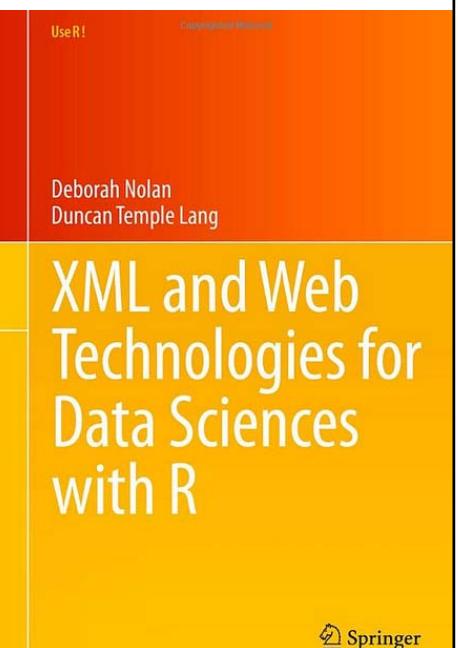
課程大綱

- 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
- `rcurl` 套件
- CSS 與 Xpath 選擇
- 爬蟲練習
- 爬蟲練習 (2)

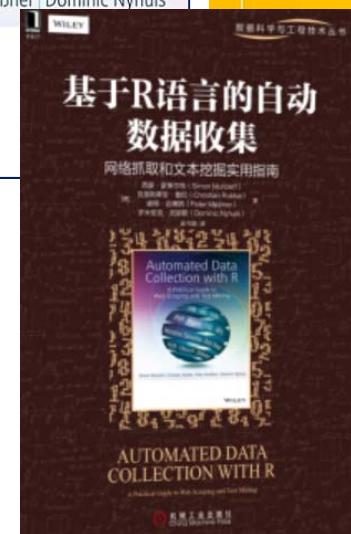


Simon Munzert | Christian Rubba | Peter Meißner | Dominic Nyhuis

Copyrighted Material



Copyrighted Material



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

<https://hmwu.idv.tw>



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()`, `xmlAttrs()`, `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>

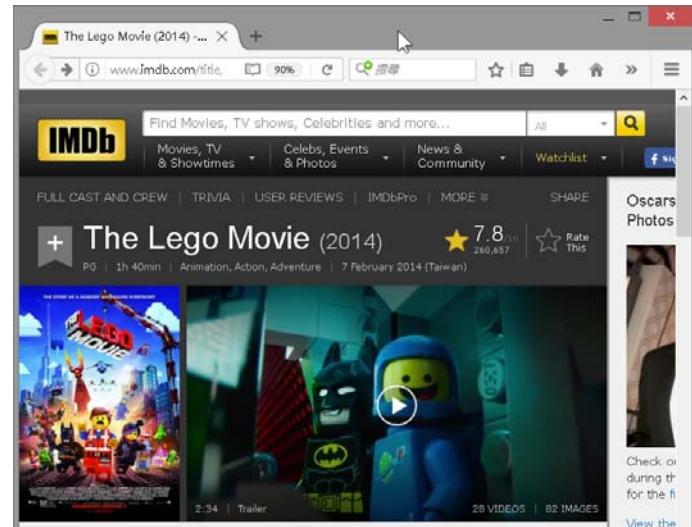
<div id="test"> text style by #test </div>
<div class="test1"> text style by .test1 </div>
```

```
#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



Actor	Character
Will Arnett	Batman / Bruce Wayne (voice)
Elizabeth Banks	Wyldstyle / Lucy (voice)
Craig Berry	Blake / Additional Voices (voice)
Alison Brie	Unikitty (voice)
David Burrows	Ocean Robot / Additional Voices (voice)
Anthony Daniels	C-3PO (voice)
Charlie Day	Ron Swanson (voice)

```
> 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 at <https://cran.r-project.org/web/packages/rvest/vignettes/selectorgadget.html>. It includes sections for Installation (dragging the bookmarklet to the toolbar) and Use (instructions for clicking elements to highlight them). A small screenshot of the SelectorGadget interface on an IMDB page is also shown.

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

The screenshot shows the SelectorGadget interface on the IMDB page for "The Lego Movie (2014)". It highlights several cast members' names in yellow. A red dashed circle highlights the ".itemprop" input field containing "name". Another red dashed circle highlights the "Clear (48)" button. The bottom status bar shows the XPath expression: `www.imdb.com/title/tt1490017/`.

<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

The screenshot shows the SelectorGadget extension running in a browser window for 'The Lego Movie (2014)'. The extension's interface includes a search bar at the top, followed by several dropdown menus: 'Movies, TV & Showtimes', 'Celebs, Events & Photos', 'News & Community', and 'Watchlist'. Below these are sections for 'CELEBS', 'NEWS', and 'LATE'. The main content area displays the movie's title, rating, runtime, and a thumbnail image. A red dashed circle highlights the 'titleCast .itemprop' selector input field, which contains the CSS selector '#titleCast .itemprop'. A red box highlights the 'Animation' category under the 'PHOTOS' section. A green box highlights the 'Action' category under the same section. The bottom of the extension shows 'Clear (30)', 'Toggle Position', 'XPath', 'Help', and 'X' buttons. The browser's address bar shows the URL: 'www.imdb.com/genre/Animation?ref_=tt_ov_inf'.

(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/>

The screenshot shows the Firefox browser window with the URL <http://getfirebug.com/> in the address bar. The Firebug extension page is displayed, featuring a large image of a firebug and the text "Firebug Web Development Evolved". A message on the page states: "The Firebug extension isn't being updated any longer. We invite you to use instead, which ship with Firefox 45 and later." Below this, there's a link to "Migration from Firebug". The Firefox toolbar is visible at the top, and the Firebug UI menu is open on the right side of the screen.

工具 (I) 說明 (H)

下載 (D) Ctrl+J
附加元件 (A) Ctrl+Shift+A
登入至 Sync (Y)...
網頁開發者 (W)
頁面資訊 (I)
選項 (O)

Firebug

開啟 Firebug F12
Firebug UI 位置
以編輯器開啟
顯示 Firebug UI
文字大小
選項
Firebug 線上
顯示面板
自訂快速鍵
關於 ... 2.0.19

網頁工具箱 (I) Ctrl+Shift+I
檢測器 (I) Ctrl+Shift+C
網頁主控台 (W) Ctrl+Shift+K
除錯器 (D) Ctrl+Shift+S
樣式編輯器 (L) Shift+F7
效能 (P) Shift+F5
網路 (N) Ctrl+Shift+Q
DOM (D) Ctrl+Shift+W
開發者工具列 (V) Shift+F2
WebIDE (W) Shift+F8
瀏覽器主控台 (B) Ctrl+Shift+J
適應性設計模式 (R) Ctrl+Shift+M
滴管 (Y)
程式碼速記本 (S) Shift+F4
Service Workers (K)
頁面原始碼 (Q) Ctrl+U
取得更多工具 (M)

The most popular and powerful web development tool

- Inspect HTML and modify style and layout in real-time
- Use the most advanced JavaScript debugger available for any browser
- Accurately analyze network usage and performance
- Extend Firebug and add features to make Firebug even more powerful
- Get the information you need to get it done with Firebug.

More Features »

Introduction to Firebug
Firebug pyroentomologist Rob Campbell gives a quick introduction to Firebug.
[Watch now »](#)

More Screencasts »

Inspect
Pinpoint an element in a webpage with ease and precision.

Log
Send messages to the console direct from your webpage through Javascript.

Profile
Measure your Javascript performance in the Console's Profiler.



如何查詢html_nodes(tag): 方法2: firebug for firefox

10/44



CSS ID

The screenshot shows the Firebug interface over a movie cast list from IMDB. In the DOM panel, a specific span element under the class 'cast_list' is highlighted with a red dashed border. This element contains the name 'Will Arnett'. In the main browser window, the same span element is also highlighted with a red dashed border, and a blue arrow points from the Firebug DOM highlight to the corresponding element in the browser's cast list.

Cast

span.itemprop | 71.2 x 17.6

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)

Related list

... /nm0006969/?ref_=tt_cl_t2" itemprop="url">

```
< !DOCTYPE html>
<div id="titleCast" class="article">
  <span class="highlightedLink"></span>
  <h2>Cast</h2>
  <table class="cast_list">
    <tbody>
      <tr>
        <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">
```



課堂練習

11/44

```
> 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/MV5BMTg4MDk1ODExN
15BMl5BanBnXkFtZTgwNzIyNjg3MDE
@._V1_UX182_CR0,0,182,268_AL_.jpg"
```

The screenshot shows a browser window displaying the IMDb page for the movie 'The Lego Movie' (2014). The page includes the movie's title, rating (7.8), and a thumbnail image. A red dashed circle highlights the movie's rating of 7.8, which corresponds to the output shown in the R code above.

NOTE:

InfoLite Extension for Google Chrome:

https://chrome.google.com/webstore/detail/infolite/ijpjbadabbpe_diegelkhgpiekdlmfpgal

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

The screenshot illustrates a process for extracting news titles from the LTN website using Firebug and SelectorGadget.

Left Side (Firefox Browser):

- The browser window shows the LTN website (<http://news.ltn.com.tw/list/BreakingNews>) with various news items listed.
- A Firebug tool window is open, showing developer keyboard shortcuts. The "網頁開發者 (W)" tab is selected.
- The "工具 (I)" dropdown menu is open, and the "頁面原始碼 (O)" option is highlighted.
- The address bar shows the URL <http://news.ltn.com.tw/list/BreakingNews>.

Right Side (Chrome Browser):

- The browser window shows the page source code for the news list.
- A red dashed circle highlights the class attribute ".picword" applied to several anchor tags.
- The SelectorGadget extension is active, with its interface visible at the bottom right of the browser window.
- The SelectorGadget interface shows the class ".picword" entered in the search input field.
- Two news items are highlighted with green boxes:
 - "艾瑪史東奧斯卡封后 爽擒獎座戰服跟著秋" (March 1, 2017)
 - "(影音)學霸偶像再+1! Chen成「EXO」第3位碩士" (March 1, 2017)



Example 2: 抓取新聞標題

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

課堂練習: appledaily

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

```

> 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, '.rtddt')
> # news.time <- html_text(html_nodes(news.rddt, 'time'))
>
> news.time <- html_text(html_nodes(apple.news, '.rtddt time'))
> news.title <- html_text(html_nodes(apple.news, '.rtddt h1'))
> news.category <- html_text(html_nodes(apple.news, '.rtddt h2'))
> news.url <- html_attr(html_nodes(apple.news, '.rtddt 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	女星變臉成癟 山根隆起鼻型再進化(21597)	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」 美女教援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

```



課堂練習: IMDb



Find Movies, TV shows, Celebrities and more... All

Movies, TV & Showtimes Celebs, Events & Photos News & Community Watchlist

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

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

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	MPAA Ratings	Runtime	User Ratings	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
[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



台北历史天气查询_台北历史天气

lishi.tianqi.com/taipei/index.html

天气网 www.tianqi.com

台北历史天气

城市选择：台湾 台北 台北 设为默认

首页 全国天气 国际天气 历史天气 pm2.5查询 视频播报 天气新闻 天气

当前位置：天气网 > 历史天气查询 > 台北历史天气

台北历史天气统计 (2011-01-01到2017-02-01) 历史天气详情

1750

台北历史天气查询_台北历史天气

lishi.tianqi.com/taipei/index.html

台北历史天气详情

2017年01月天气

2016年12月天气 2016年11月天气 2016年10月天气 2016年09月天气 2016年08月天气

2016年07月天气 2016年06月天气 2016年05月天气 2016年04月天气 2016年03月天气

台北1月份天气|台北1月份天气

lishi.tianqi.com/taipei/201701.html

台北2017年1月份天气详情

日期	最高气温	最低气温	天气	风向	风力
2017-01-01	17	0	晴	微风	小于3级
2017-01-02	18	0	小雨	微风	小于3级
2017-01-03	18	0	小雨	微风	小于3级
2017-01-04	19	0	晴	微风	小于3级
2017-01-05	18	0	多云	微风	小于3级

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

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

台北1月份天气|台北1月份天气

lishi.tianqi.com/taipei/201701.html

台北2017年1月份天气详情

日期	最高气温	最低气温	天气	风向	风力
2017-01-01	17	0	晴	微风	小于3级
2017-01-02	18	0	小雨	微风	小于3级
2017-01-03	18	0	小雨	微风	小于3级
2017-01-04	19	0	晴	微风	小于3级
2017-01-05	18	0	多云	微风	小于3级

工具栏：检测器 主控台 除错器 模式编辑器 效能 跟踪 网路 日 书签 搜索 HTML

HTML 结构分析：

```

</ul>
  <div id="tool_site" class="box-base m7">
    <div class="box-hd">
      <h3 class="box-t-1">台北2017年1月份天气详情</h3>
    </div>
    <div class="tq tongji2">
      <ul class="t1">
        <li>日期</li>
        <li>最高气温</li>
        <li>最低气温</li>
        <li>天气</li>
        <li>风向</li>
        <li>风力</li>
      </ul>
    </div>
  </div>

```



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="taibei", 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/taibei/", year, month, ".html")
    url.list <- c(url.list, url)
  }

  tianqi.table <- NULL
  for(x in url.list){
    # x <- "http://lishi.tianqi.com/taibei/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
}

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

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

library(stringr)

	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



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

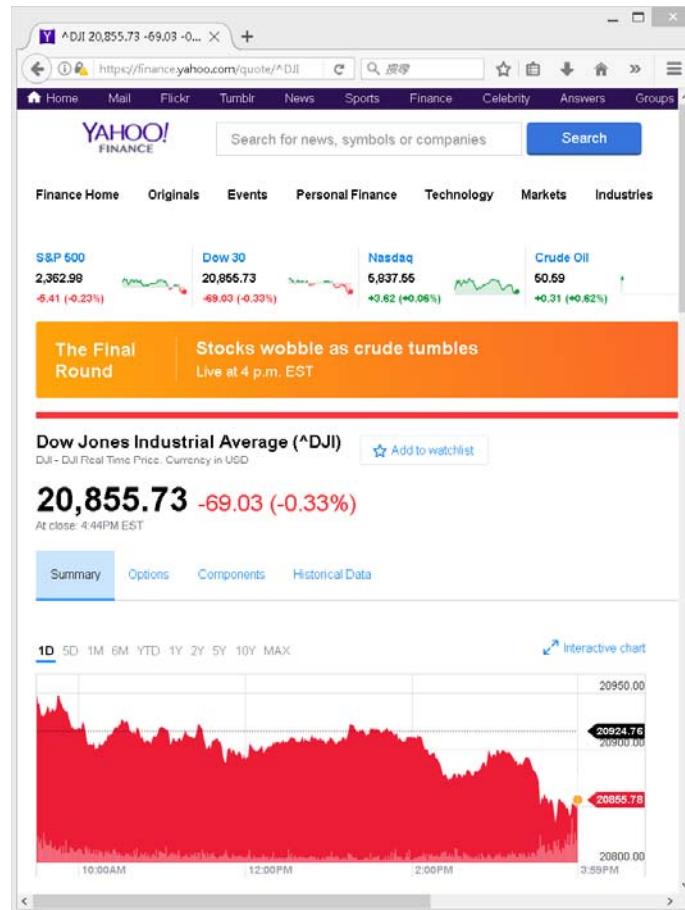
FACEBOOK

- <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



- **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)
```

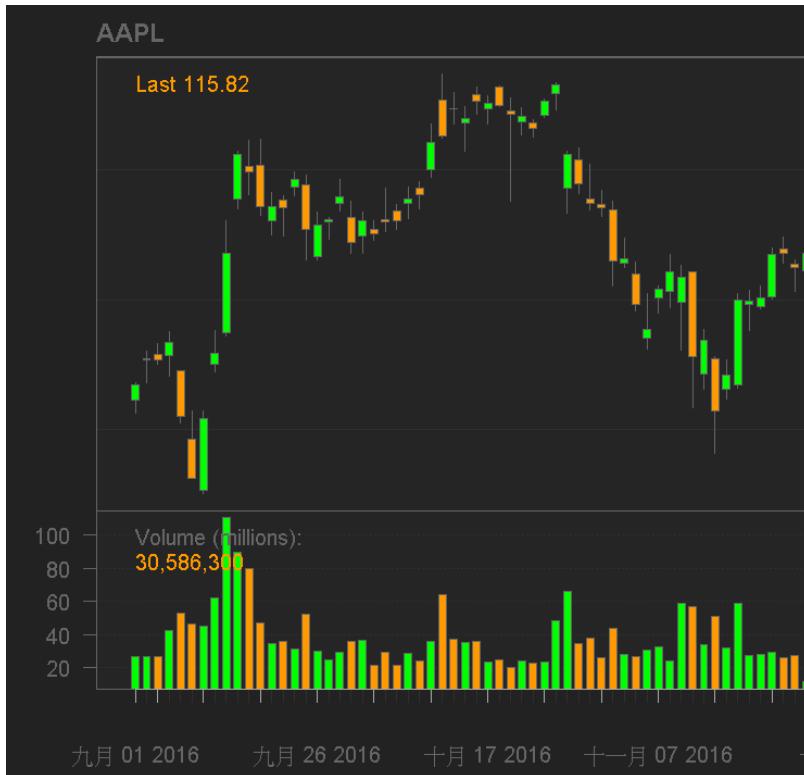
```
> AAPL.df <- as.data.frame(AAPL)
> class(AAPL.df)
[1] "data.frame"
```



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://teacher cannon.weebly.com/uploads/6/8/3/8/6838442/sm_g_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
1 $117.23M      NA   Finance
2 $89.16M       2015 Basic Industries
3 $116.6M       NA      <NA>
4 $81.69M       1988 Capital Goods Industrial Machinery/Components
5 $15.98M       NA Technology Diversified Commercial Services
6 $15.81M       NA Public Utilities
                                         Industry Exchange
                                         Real Estate      AMEX
                                         Precious Metals  AMEX
                                         <NA>            AMEX
                                         AMEX
                                         AMEX
                                         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

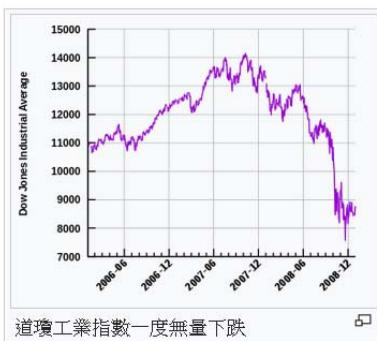
chartSeries(TSMC)



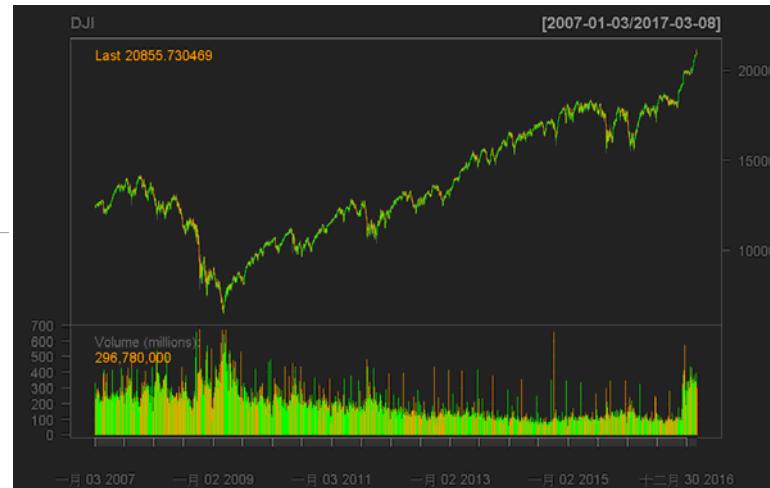
chartSeries(UMC)



維基百科
自由的百科全書



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>

This tool lets you make calls to Facebook's Graph API.
Learn more about Graph API.

Log into Facebook to use this tool.

Log In

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。

圖形 API 測試工具

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

應用程式: [?] Graph API Explorer ▾

取得權杖 ▾

- 取得用戶存取權杖
- 取得應用程式權杖
- 取得粉絲專頁存取權杖

<https://hmwu.idv.tw>



選擇要開放的權限

32/44

選擇權限

v3.0 ▾ ×

用戶數據權限

email user_hometown user_posts
 user_age_range user_likes user_status
 user_birthday user_link user_tagged_places
 user_friends user_location user_videos
 user_gender user_photos

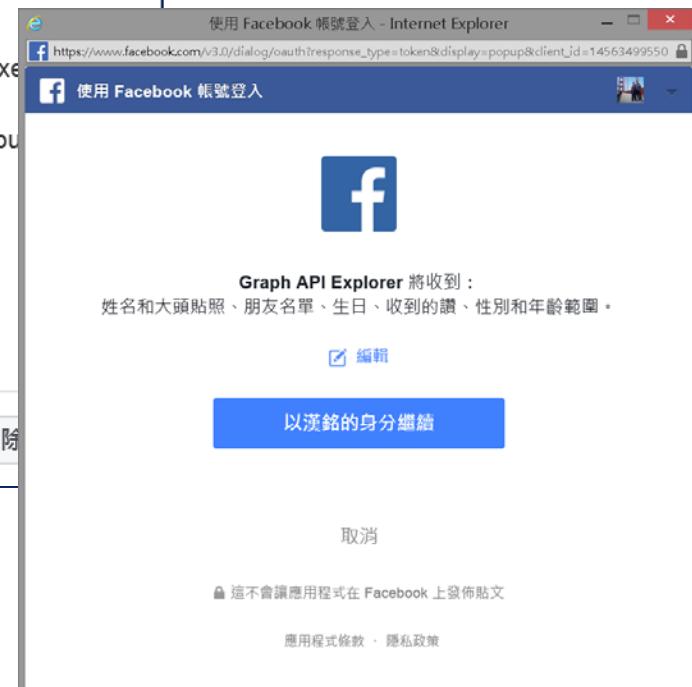
活動、社團和粉絲專頁

ads_management pages_manage_instant_articles publish_pages
 ads_read pages.messaging publish_to_groups
 business_management pages.messaging_payments read_page_mailbox
 groups.access_member_info pages.messaging_phone_number user_events
 manage_pages pages.messaging_subscriptions user_managed_group
 pages.manage_cta pages.show_list

其他

instagram_basic instagram_manage_insights read_insights
 instagram_manage_comments read_audience_network_insights

預設已包含公開的個人檔案。





取得FB朋友資料

33/44

facebook for developers

文件 工具 支援 C

應用程式: [?] Graph API Explorer ▾

圖形 API 測試工具

存取權杖: EAACEdEose0cBAE2LxoRtYvgZBY9RrzG7ncncFZCK3... 取得權杖 ▾

GET → /v3.0/me?fields=id,name 提交 深入瞭解圖形 API 語法



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

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



```
> library(Rfacebook)
> token <- 'XXXXXX'
> me <- getUsers("me", token, private_info=TRUE)
> me
      id      name 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
>
> 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）。
 - 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                  1                  4
```



取得「活動、社團和粉絲專頁」權限

25/44

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」 取得權杖 ▾

GET → /v3.0/me?fields=id,name

選擇權限 v3.0 深入瞭解圖形 API 語法

用戶數據權限

email user_hometown user_posts
 user_age_range user_likes user_status
 user_birthday user_link user_tagged_places
 user_friends user_location user_videos
 user_gender user_photos

活動、社團和粉絲專頁

ads_management pages_manage_instant_articles publish_page
 ads_read pages.messaging publish_to_group
 business_management pages.messaging_payments read_page_insights
 groups_access_member_info pages.messaging_phone_number user_events
 manage_pages pages.messaging_subscriptions user_manage_groups
 pages_manage_cta pages_show_list

其他

instagram_basic instagram_manage_insights read_insight
 instagram_manage_comments read_audience_network_insights

預設已包含公開的個人檔案。 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&client_id=...&redirect_uri=...&state=...&scope=...

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

<https://hmwu.idv.tw>



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: EAACED... / me?fields=id,name

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 Extract list of likes of a Facebook friend
getNetwork Extract network of friends of authenticated user
getNewsfeed Download recent posts from the authenticated user's newsfeed
getPage Extract list of posts from a public Facebook page
getPageToken Get a page access token
getPost Extract information about a public Facebook post
getReactions Extract total count of reactions to one or more Facebook posts
getShares Extract list of users who publicly shared a public Facebook post
getUsers Extract information about one or more Facebook users
Rfacebook Access to Facebook API via R
searchFacebook Search public posts that mention a string
searchGroup Find Facebook ID of a group
searchPages Search pages that mention a string
updateStatus Update Facebook status from R

<https://hmwu.idv.tw>

建立 Facebook App

<https://developers.facebook.com/>

The screenshot shows the Facebook for Developers interface. On the left, there's a sidebar with navigation items like '主控板', '設定', '角色', '提示', and '應用程式審查'. The main area displays a welcome message: '歡迎使用 Facebook for Developers' with a photo of two people at the Golden Gate Bridge. A red circle labeled '1' highlights the '建立 Facebook for Developers 帳號' button. Below it, a note says: '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「開發人員設定」取消訂閱。' To the right, a search bar has a red circle labeled '2' and a button labeled '建立新的應用程式' with a hand cursor icon. At the bottom, a modal window is open with a red circle labeled '3' around its title '建立新的應用程式編號'. It contains fields for '顯示名稱' (RTestApp) and '聯絡電子郵件' (hanming.wu@gmail.com). At the bottom right of the modal is a '建立應用程式編號' button. The URL 'https://hmwu.idv.tw' is visible at the bottom of the page.

facebook for developers

產品 文件 更多 ▾ 開始使用 (

① 註冊 —— ② 第一款應用程式 —— ③ 跟我們介紹一下你自己吧

20

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「開發人員設定」取消訂閱。

facebook for developers

應用程式編號：187128678586686

狀態：調整中

RTestApp

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

4

Account Kit

Facebook 登入

Analytics

Messenger

閱讀文件 設定 閱讀文件 設定 閱讀文件 設定 閱讀文件 設定

<https://hmwu.idv.tw>



獲取應用程式編號及應用程式密鑰

38/44

opers

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 狀態：調整中 查看分析工具

主控板

設定

基本資料

顯示名稱 RTestApp

應用程式編號 364039657416919

應用程式密鑰 ⋮ 显示

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)

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 電腦



網站

立即開始



網站網址

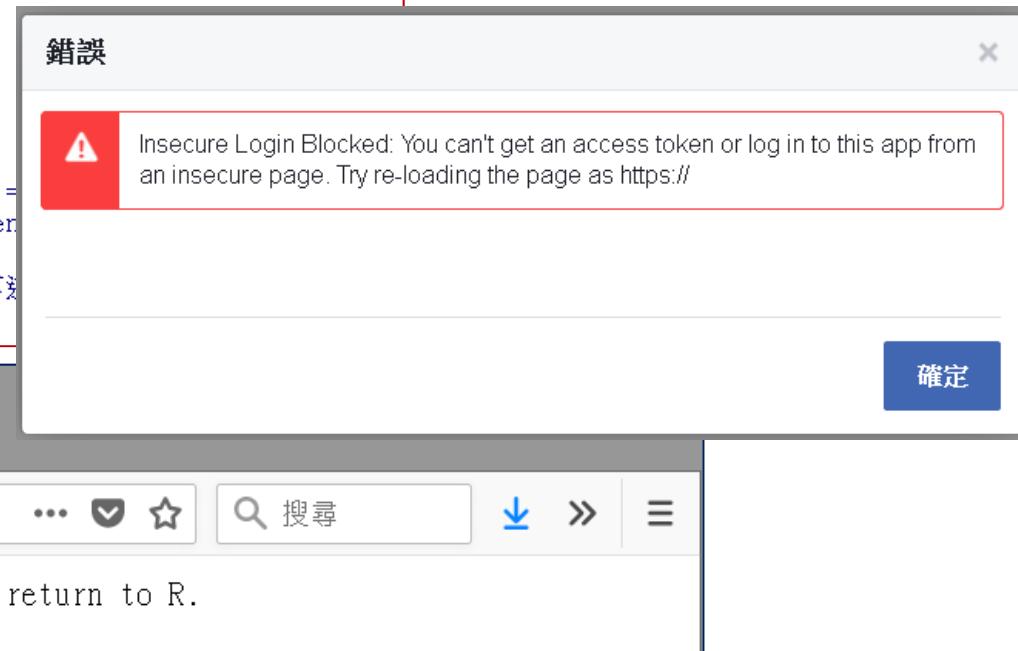
+ 新增平台

捨棄

儲存變更

連結FB

```
> fb.oauth <- fbOAuth(  
+   app_id="364039657416919",  
+   app_secret="5ala9b[REDACTED]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 不足  
> |
```



```
> # 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



RTestApp

應用程式編號：364039657416919



狀態：調整中

用戶端 OAuth 設定

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。 [?]

網路 OAuth 登入
啟用網頁型用戶端 OAuth 登入。 [?]

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。 [?]

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。 [?]

有效的 OAuth 重新導向 URI

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 [?]

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

嵌入的瀏覽器 OAuth 登入

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日開盤後因此大跌逾5%；另外，歐盟司法事務執委喬羅瓦 (Vera Jourová) 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。

捨棄**儲存變更**

FB 粉絲專頁



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

PChome 24h購物 - 首頁 X +

← → G H

i g https://www.facebook.com/pchome24h/

f PChome 24h購物 S

PChome 24h購物 ✓
@pchome24h

首頁
貼文
相片

Like 讚 Share 追蹤 Share ...

```
> 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：
<https://lookup-id.com>

Lookup-ID.com

Facebook ID Facebook Card FB Search Extract Members Directory

Looking for your Facebook profile ID / Group ID / Page ID ...

Type your Facebook profile URL

https://www.facebook.com/pchome24h/ Lookup

Success! If Facebook name is Unknown, then we found your numeric ID:

203507002993446

```
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
```

查詢粉絲專頁的FB ID：



The screenshot shows the Facebook for Developers dashboard. The 'Tools' menu item is highlighted with a hand cursor icon. On the left, there's a sidebar with options like 'Main Dashboard', 'Settings', 'Basic Information', 'Advanced', 'Roles', and 'Notifications'. The main area displays application details for 'RTestApp' (App ID: 364039657416919) and a summary of available tools: Graph API Testing Tool, Access Token Tool, and Share Debugger.



The screenshot shows the Graph API Explorer tool. It has a URL input field containing 'GET /v3.0/pchome24h?fields=id,name'. The 'Node' section shows 'Node: pchome24h' with checkboxes for 'id' and 'name', both of which are checked. The results pane displays a JSON response:

```

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

```

商用工具



Microsoft Power BI

登入 免費註冊

開始使用 Power BI Desktop

更少資料準備，更多資料讚嘆
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

<https://hmwu.idv.tw>

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！

The Allround-Tool for strong Social-Media Management.

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](#)