week 3

agenda

RegExp 介紹

Facebook API 介紹

抓取 FB 蘋果即時新聞 內容

資料庫存取

爬蟲自動化

正規表達式

為什麼寫爬蟲會遇到 RegExp ？

聯合新聞網 udn.com

搜尋冥王星

冥王星

即時 要聞 娛樂 運動 全球 社會 產經 股市 健康 生活 文教 評論 地方 兩岸 數位 旅遊 雜誌 閱讀 購物 音樂

「壁咚」 別為浪漫傷肩、肘、頸

·穿緊身牛仔褲久蹲 肌肉、神經受損　·6大錯誤習慣 害你背超痛

線上影音直播 | 鳴人堂 | 快艇夏季補強 | 美女野獸情侶檔 | 壁咚過度傷筋骨

2015/7/25 星期六　新北市 32 °C ~ 36 °C

即時 分享 熱門 影音

- 14:38 5獨木舟翻覆11人落海 1女斷指
- 14:38 藍挺吳思華提告 綠駁抹黑學運老梗
- 14:29 教育部告不告記者？ 一周內有答案
- 14:24 50嵐蜂蜜出包 業者：落實送驗SOP
- 14:19 網友王子變青蛙 陸女吞玻璃輕生
- 14:06 50嵐蜂蜜出包 3飲品暫停售
- 14:01 神秘薄霧環繞 冥王星再給地球人驚喜
- 13:58 蜂蜜出包 50嵐：加強原料把關
- 13:58 基隆獨木舟翻船 四人平安獲救

願景 | 房地產 | 財經 | 基金 | 樂活 | 時尚 | 汽車 | NBA台灣 | 追星 | 部落格 | 讀書吧 | 讀小說 | 知識庫 | 電子報 | 有設計

udn.com/search/result/2/冥王星

聯合新聞網 udn.com

即時 要聞 娛樂 運動 全球 社會 產經 股市 健康 生活 文教 評論 地方 兩岸 數位 旅遊 雜誌 閱讀 購物

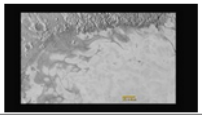
搜尋結果

搜尋 冥王星 的結果：共找到 58筆　約 0.009秒

神秘薄霧環繞 冥王星再給地球人驚喜

即時：2015/07/25

科學家今天表示，美國國家航空暨太空總署「新視野號」太空船上週飛越冰封



關注 udn

udn facebook

Elements Network Sources Timeline Profiles Resources Audits Console

View: Options: Preserve log Disable cache

Filter All XHR Script Style Images Media Fonts Documents WebSockets Other Hide data URLs

Name
Path

analytics.js
www.google-analytics...

beacon.js
b.scorecardresearch.c...

atrk.js
d31qbv1cthcecs.clou...

er.js?1437806725763
rec.udn.com

sdk.js
connect.facebook.net...

accept_var_min.js
/func/js

searchResult4utf8.jsp...
search.udn.com/search

?ids=http://udn.com/...
graph.facebook.com

10 / 51 requests | 8.1 KB / 29...

Headers Preview Response Cookies Timing

```
17 var pageList="";
18 var arr = new Array(rc);
19 var arrList = '';
20 var searchInfo='';
21 //searchInfo='搜尋 <u>' + q + '</u> 的結果：共找到 ' + numFound + '筆 約 ' + runTime ;
22 //searchInfo='搜尋 <u>' + q + '</u> 的結果：共找到 ' + numFound + '筆';
23
24 arr[0] = new Array(7);
25 arr[0][0] = 'http://udn.com/news/story/5/1078576';
26 arr[0][1] = 'http://uc.udn.com.tw/photo/2015/07/25/98/1144486.jpg';
27 arr[0][2] = '神秘薄霧環繞 <u>冥王星</u>再給地球人驚喜';
28 arr[0][3] = '即時';
29 arr[0][4] = '2015/07/25';
30 arr[0][5] = '科學家今天表示，美國國家航空暨太空總署「新視野號」太空船上週飛越冰封星球<u>冥王星</u>時，拍到的清晰輪廓圖可看到一大片大氣薄霧，地表特寫照顯示有氮冰流動。美國國家航空暨';
31 arr[0][6] = '';
32
33 arr[1] = new Array(7);
34 arr[1][0] = 'http://udn.com/news/story/5/1078223';
35 arr[1][1] = 'http://uc.udn.com.tw/photo/2015/07/25/98/1143868.jpg';
36 arr[1][2] = '新視野號揭秘 <u>冥王星</u>可能有氮冰河';
37 arr[1][3] = '即時';
38 arr[1][4] = '2015/07/25';
39 arr[1][5] = '「新視野號」（NEW HORIZONS）太空船回傳的最新影像顯示，<u>冥王星</u>上可能有由氮冰組成的冰河。英國廣播公司（BBC）網站報導，科學家相信他們看到跡象顯示，冥';
40 arr[1][6] = '';
41
```

Data 直接儲存在 javascript code 回傳
需要自己用正規表達截取出來

https://www.stockdog.com.tw/stockdog/index.php?m=overview&sid=1101+台泥

股 查詢總覽 股狗部落格 建議與留言 社群 付費方案 功能特色 行動版 登入 註冊

2015-07-24 台泥(1101) - 上市水泥工業

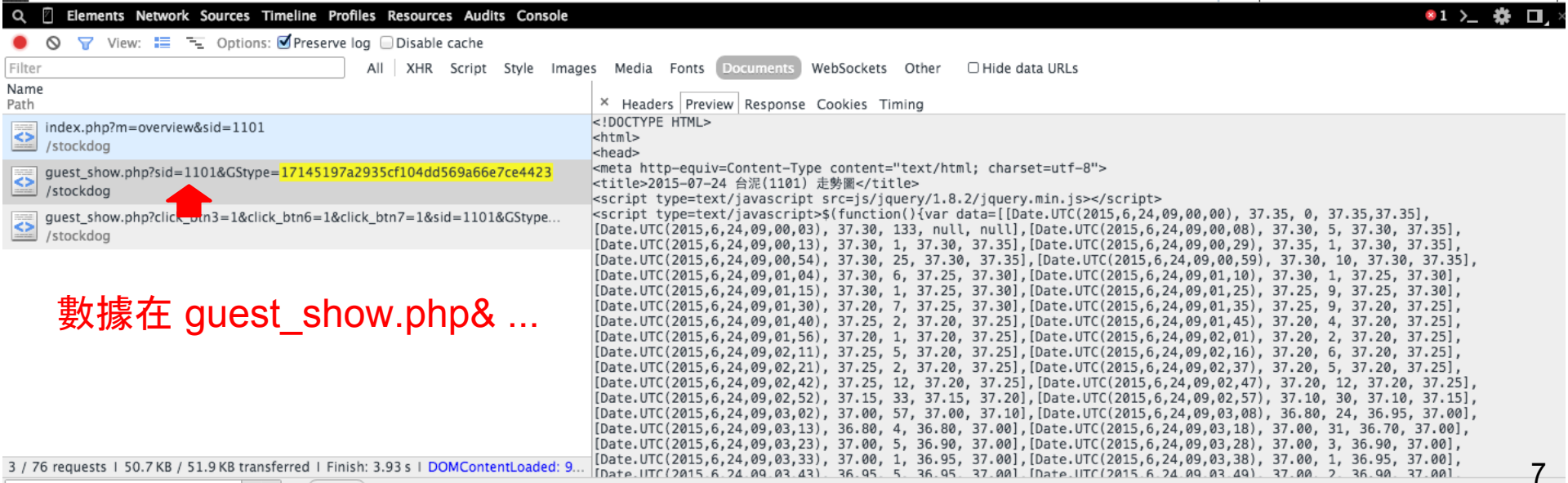
Stock Dog 是一個提供股票資訊的網站

總股數：**3,692,176**張
董監事及大股東持股比：**29.07%**
在外流通張數：**2,618,852**張

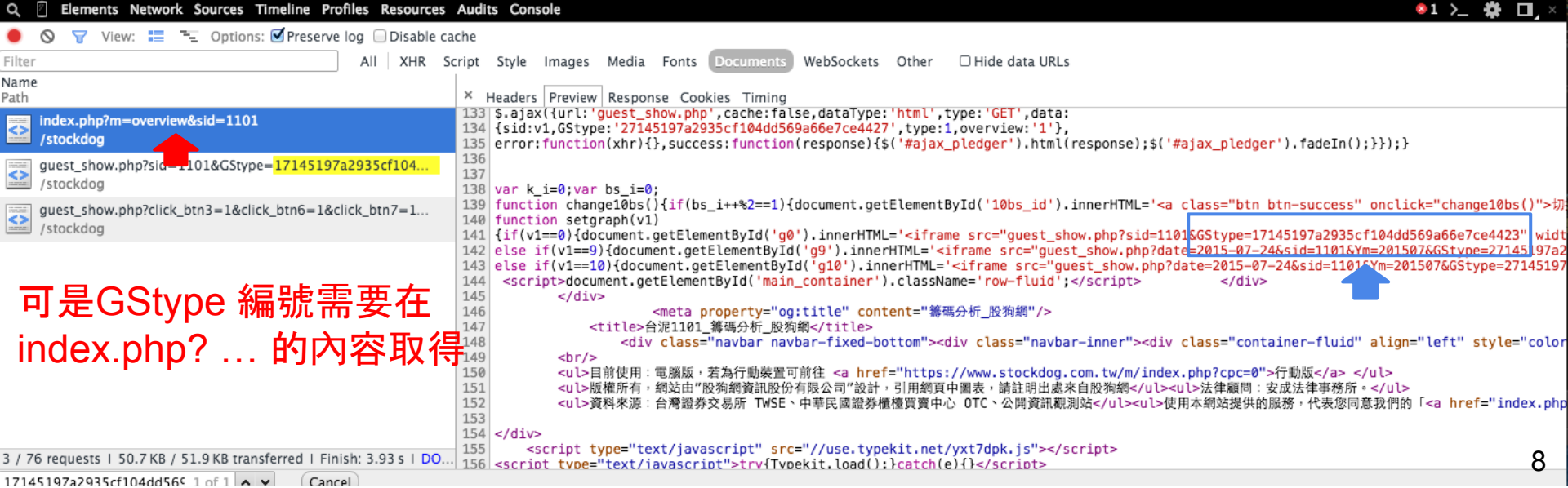
開盤	37.3	漲幅	-1.47%
最高	37.35	振福	1.47%
最低	36.8	漲跌	-0.55
收盤	36.8	昨收	37.35
總量	6719		

委買價	委買量	委賣價	委賣量
36.80	354	36.85	4
36.75	122	36.90	183
36.70	202	36.95	103
36.65	337	37.00	165
36.60	852	37.05	199
委買賣差			1213
委買賣比			2.85

https://www.stockdog.com.tw/stockdog/index.php?m=overview&sid=1101%20%E5%8F%B0%E6%B3%A5

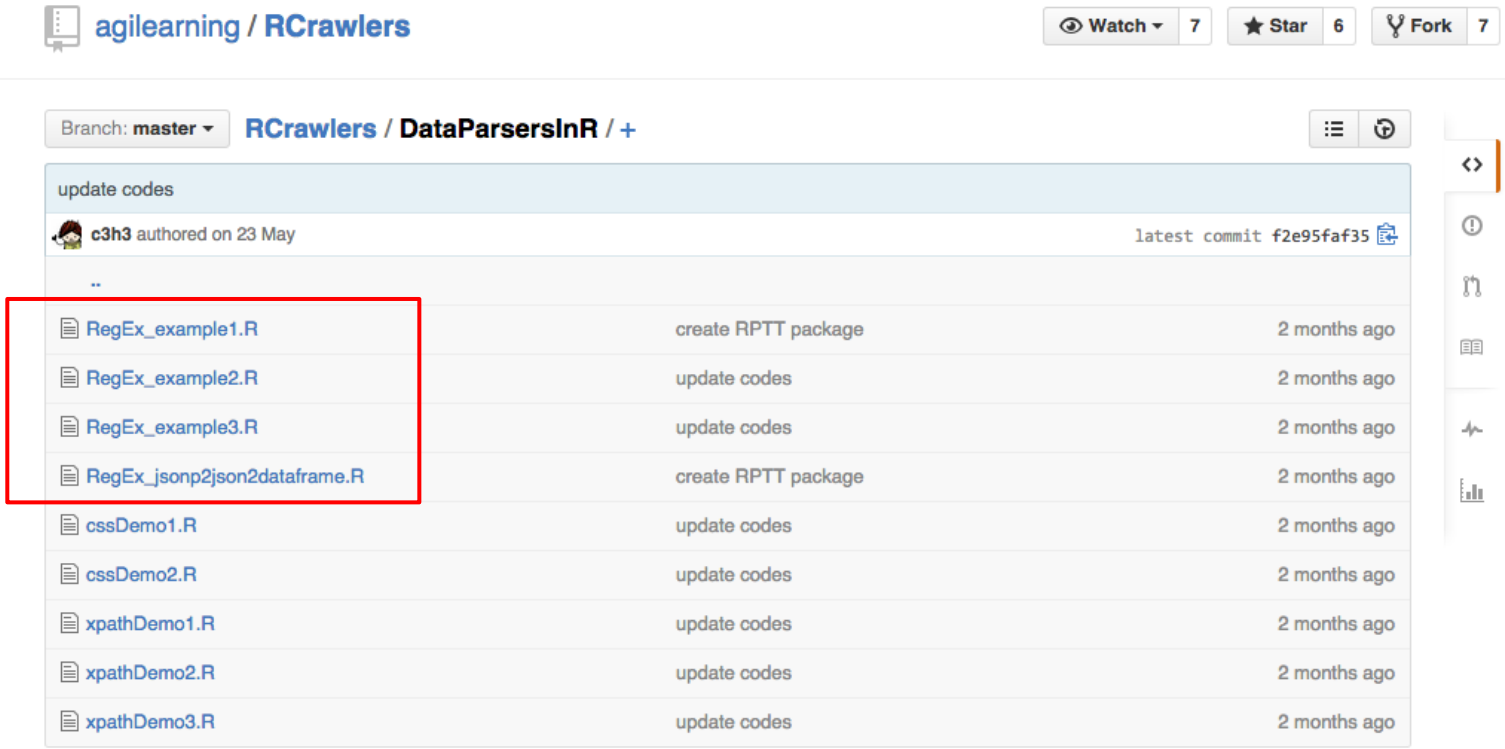


數據在 guest_show.php& ...



可是GStype 編號需要在 index.php? … 的內容取得

正規表達練習



https://github.com/agilearning/RCrawlers/tree/master/DataParsersInR

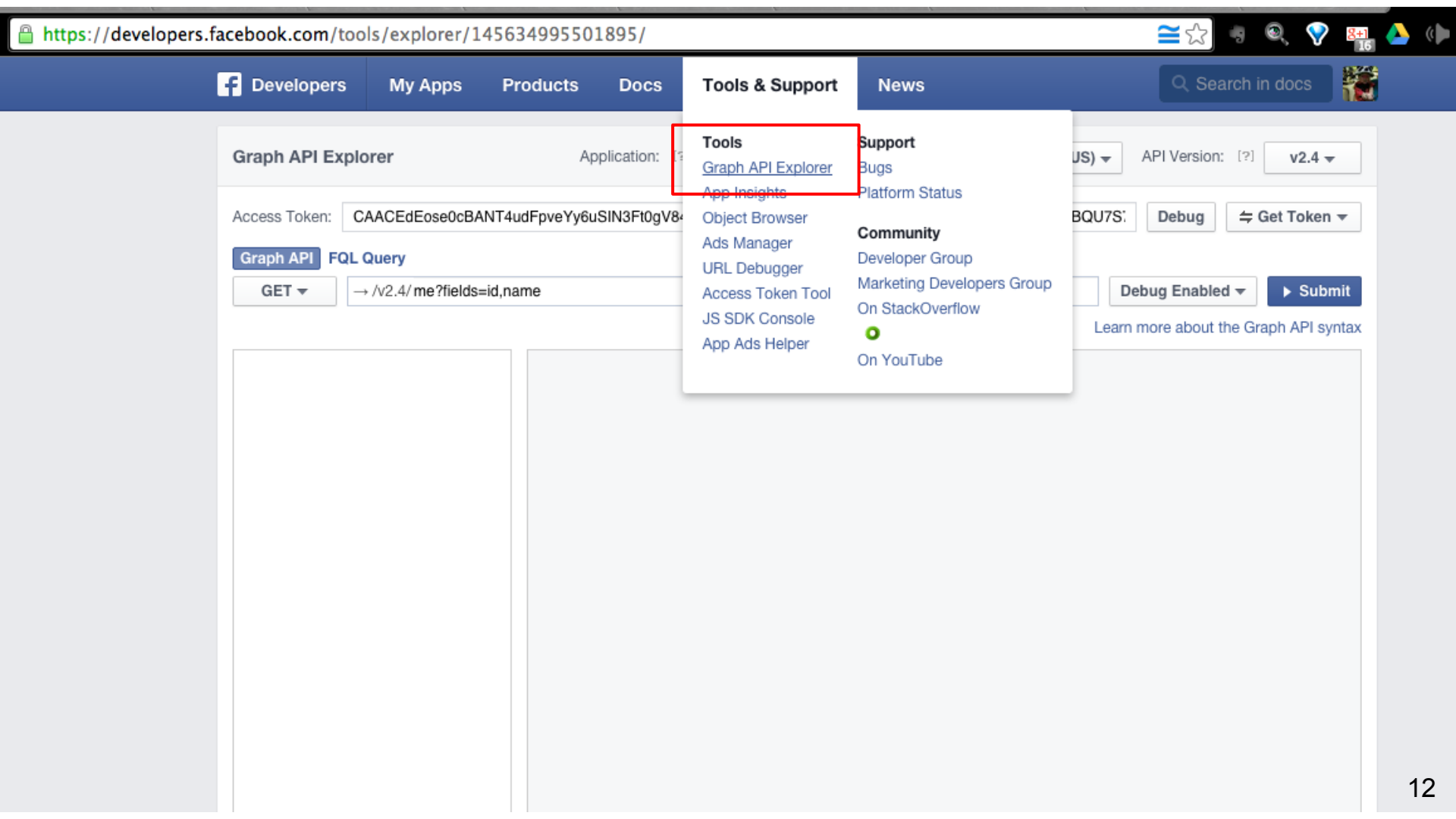
練習 5分鐘

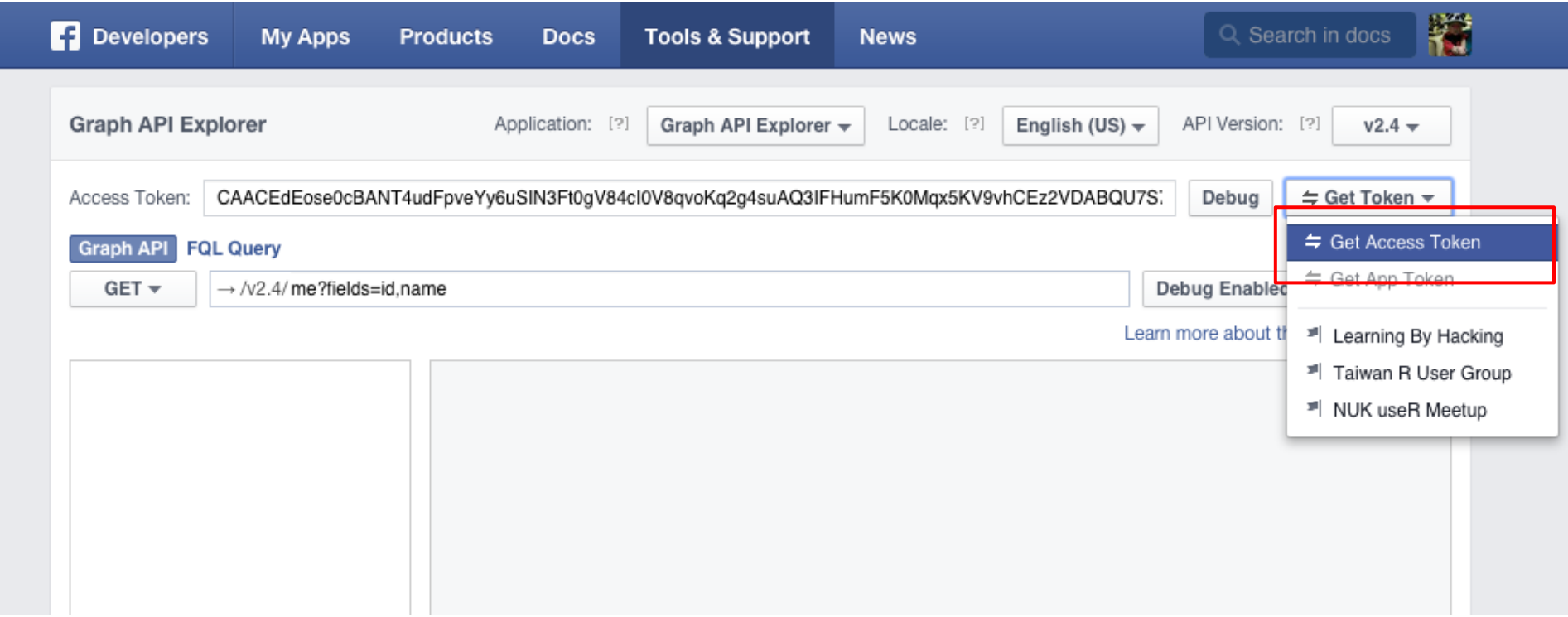
Facebook API

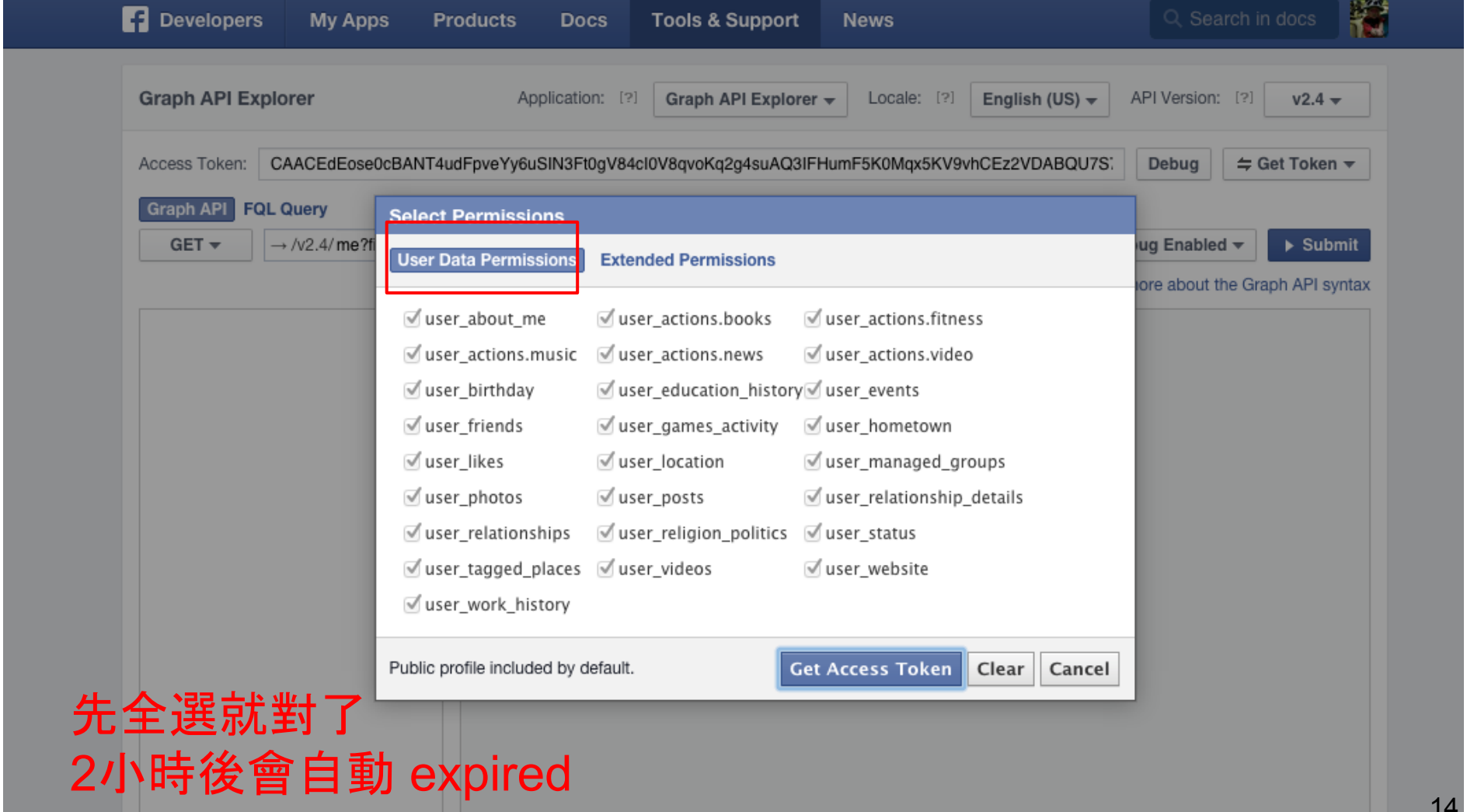
觀察 API 怎麼用

如何透過 API 拿到資料

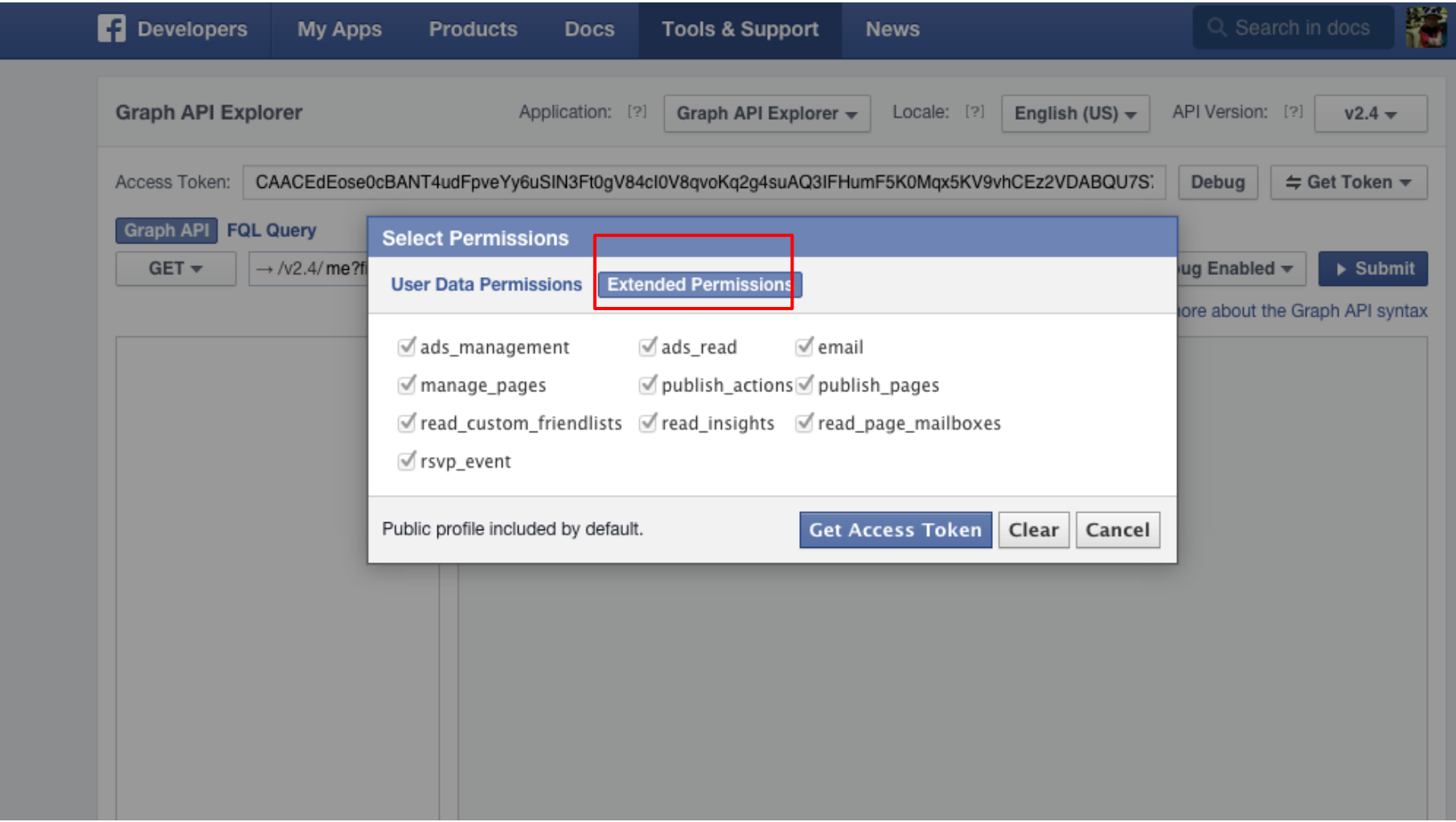
如何使用 Search API

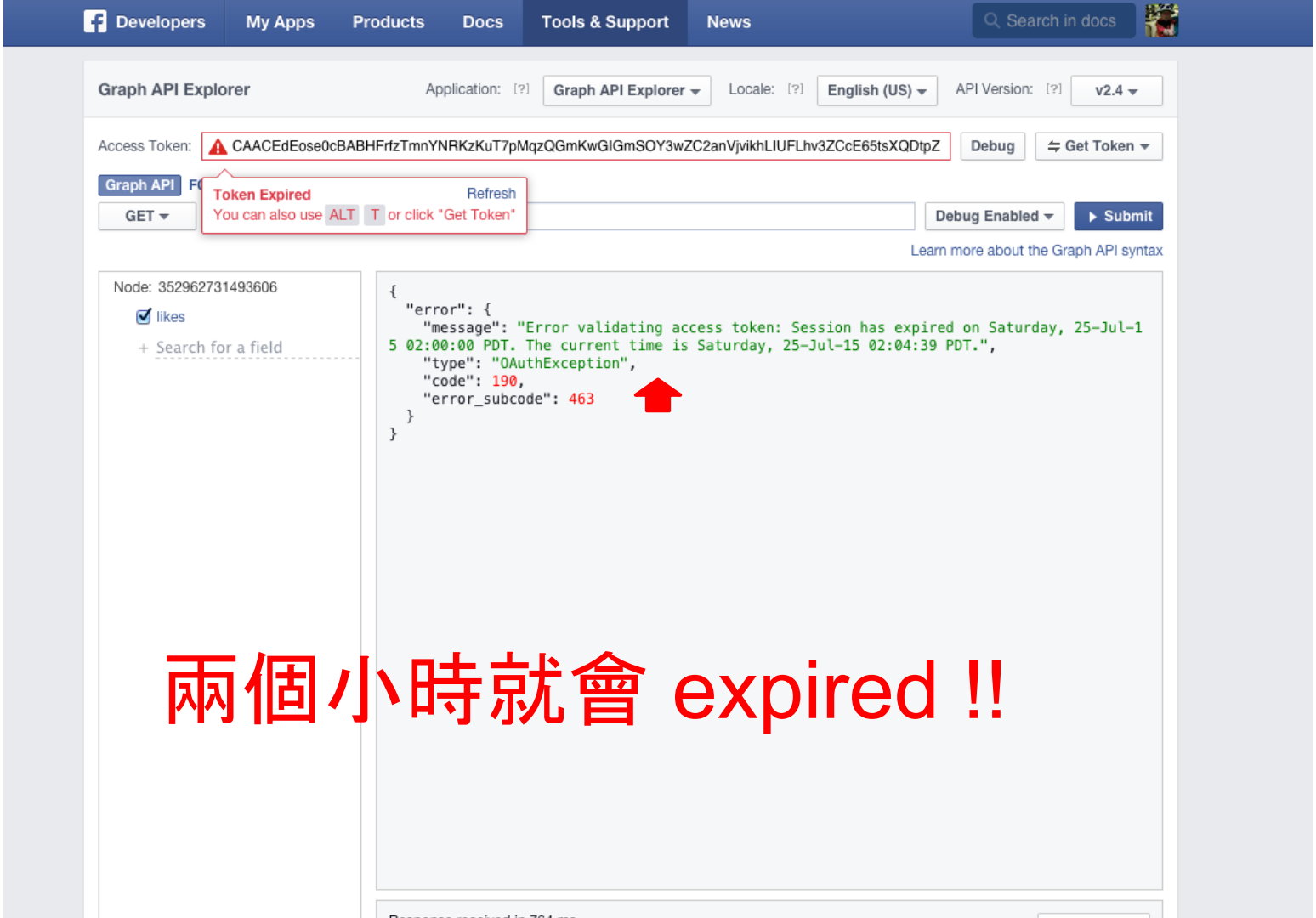

https://developers.facebook.com/tools/explorer/145634995501895/
Developers
My Apps
Products
Docs
Tools & Support
News
Search in docs
Graph API Explorer
Application:
API Version:
v2.4
Access Token:
Debug
Get Token
Graph API
FQL Query
GET
/v2.4/ me?fields=id,name
Debug Enabled
Submit
Learn more about the Graph API syntax
Tools
Graph API Explorer
App Insights
Object Browser
Ads Manager
URL Debugger
Access Token Tool
JS SDK Console
App Ads Helper
Support
Bugs
Platform Status
Community
Developer Group
Marketing Developers Group
On StackOverflow
On YouTube


Developers
My Apps
Products
Docs
Tools & Support
News
Search in docs
Graph API Explorer
Application: [?]
Graph API Explorer
Locale: [?]
English (US)
API Version: [?]
v2.4
Access Token:
CAACEdEose0cBANT4udFpveYy6uSIN3Ft0gV84cI0V8qvoKq2g4suAQ3IFHumF5K0Mqx5KV9vhCEz2VDABQU7S
Debug
Get Token
Graph API
FQL Query
GET
/v2.4/ me?fields=id,name
Debug Enabled
Learn more about t
Get Access Token
Get App Token
Learning By Hacking
Taiwan R User Group
NUK useR Meetup

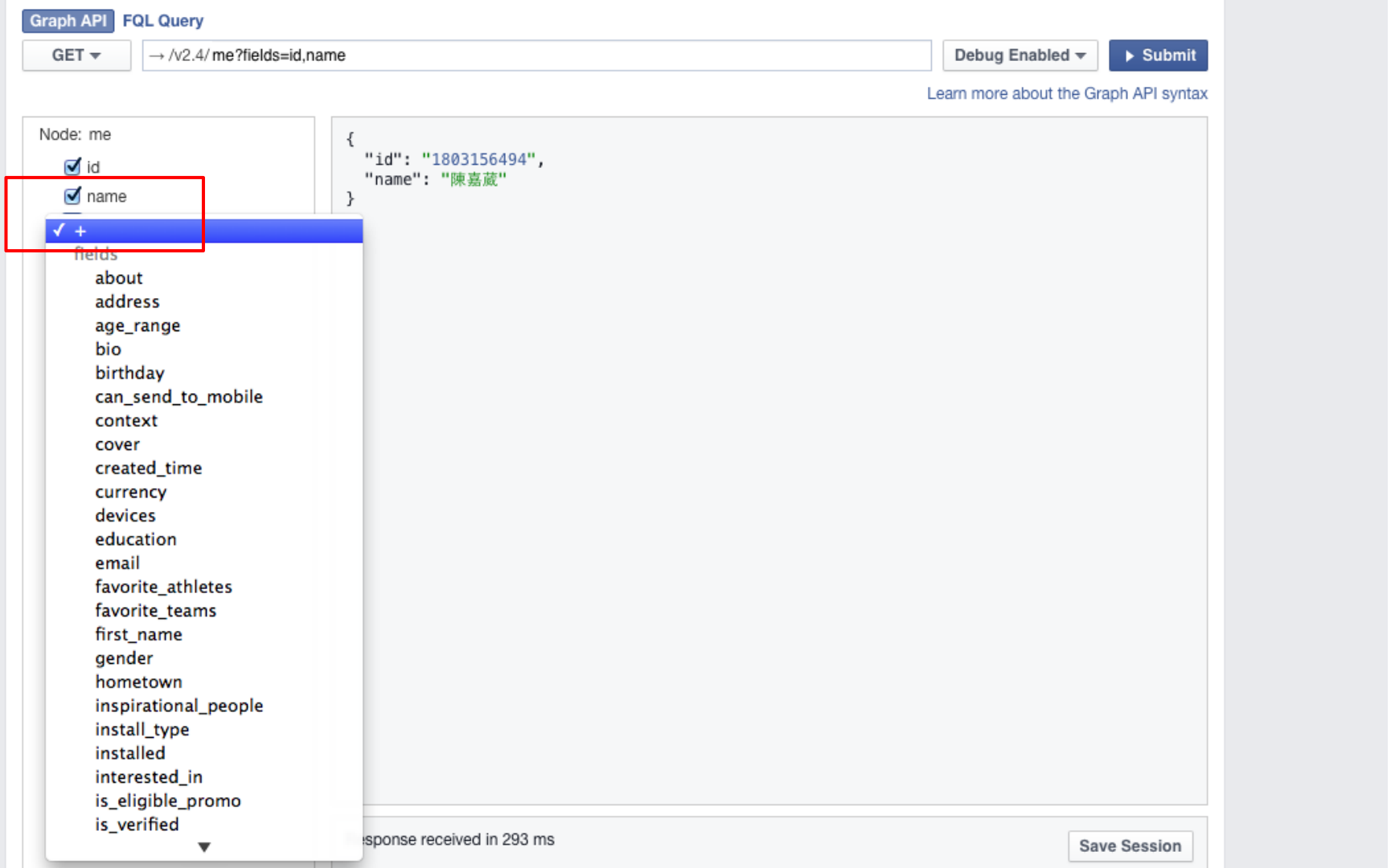


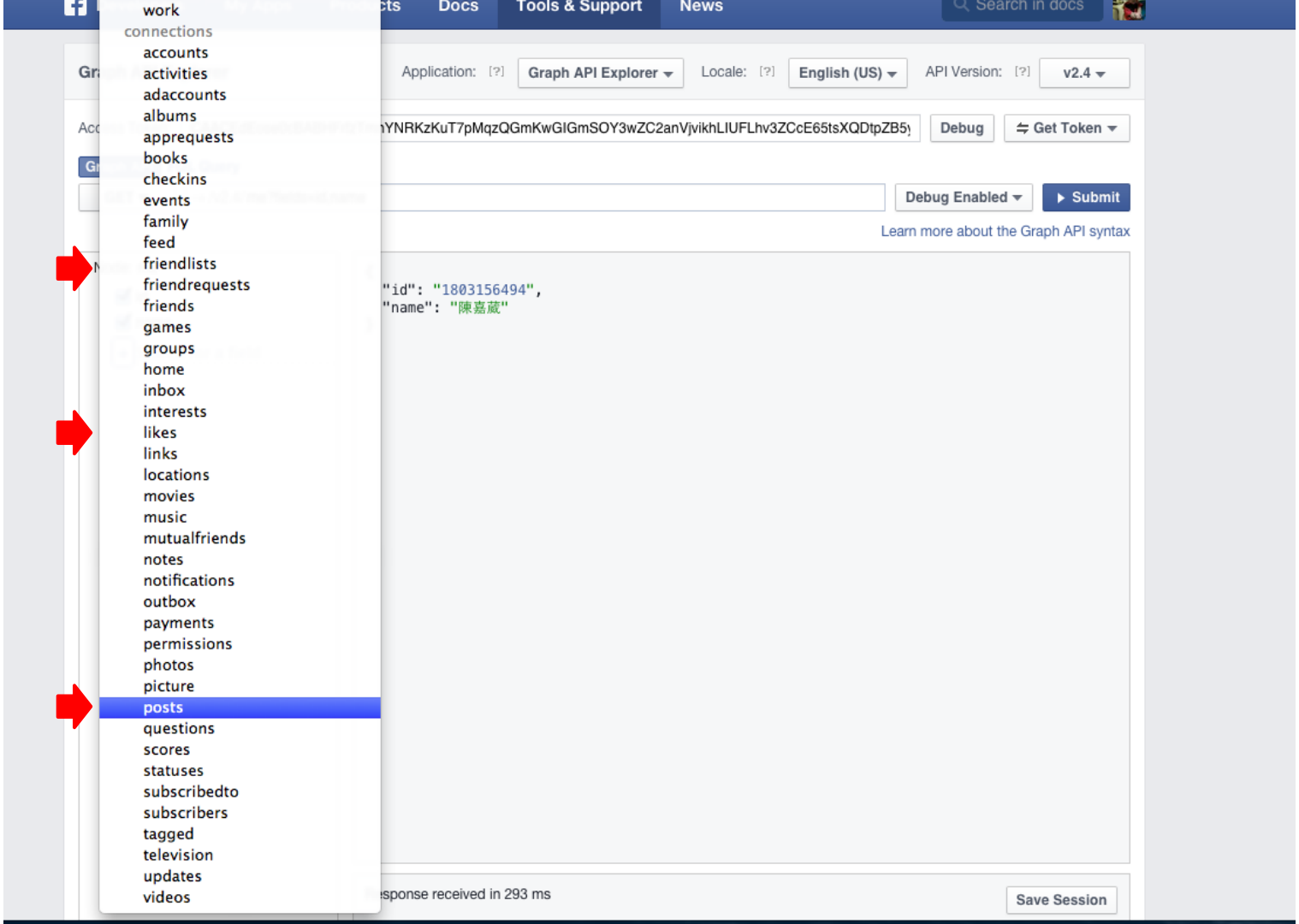
先全選就對了
2小時後會自動 expired


Developers
My Apps
Products
Docs
Tools & Support
News
Search in docs
Graph API Explorer
Application: [?]
Graph API Explorer
Locale: [?]
English (US)
API Version: [?]
v2.4
Access Token:
CAACEdEose0cBANT4udFpveYy6uSIN3Ft0gV84cI0V8qvoKq2g4suAQ3IFHumF5K0Mqx5KV9vhCEz2VDABQU7S
Debug
Get Token
Graph API
FQL Query
GET
→ /v2.4/ me?fi
ug Enabled
Submit
ore about the Graph API syntax
Select Permissions
User Data Permissions
Extended Permissions
ads_management
ads_read
email
manage_pages
publish_actions
publish_pages
read_custom_friendlists
read_insights
read_page_mailboxes
rsvp_event
Public profile included by default.
Get Access Token
Clear
Cancel


Developers
My Apps
Products
Docs
Tools & Support
News
Search in docs
Graph API Explorer
Application: [?]
Graph API Explorer
Locale: [?]
English (US)
API Version: [?]
v2.4
Access Token:
CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65tsXQDtpZ
Debug
Get Token
Graph API
Token Expired
Refresh
You can also use ALT T or click "Get Token"
GET
Debug Enabled
Submit
Learn more about the Graph API syntax
Node: 352962731493606
likes
Search for a field
{
"error": {
"message": "Error validating access token: Session has expired on Saturday, 25-Jul-15 02:00:00 PDT. The current time is Saturday, 25-Jul-15 02:04:39 PDT.",
"type": "OAuthException",
"code": 190,
"error_subcode": 463
}
}
兩個小時就會 expired !!


Developers
My Apps
Products
Docs
Tools & Support
News
Search in docs
Graph API Explorer
Application: [?]
Graph API Explorer
Locale: [?]
English (US)
API Version: [?]
v2.4
Access Token:
CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65tsXQDtpZB5
Debug
Get Token
Graph API
FQL Query
GET
→ /v2.4/ me?fields=id,name
Debug Enabled
Submit
Learn more about the Graph API syntax
Node: me
id
name
+ Search for a field
{
"id": "1803156494",
"name": "陳嘉葳"
}
大家應該會來到這個畫面


Graph API
FQL Query
GET
→ /v2.4/ me?fields=id,name
Debug Enabled
Submit
Learn more about the Graph API syntax
Node: me
id
name
+
fields
about
address
age_range
bio
birthday
can_send_to_mobile
context
cover
created_time
currency
devices
education
email
favorite_athletes
favorite_teams
first_name
gender
hometown
inspirational_people
install_type
installed
interested_in
is_eligible_promo
is_verified
{
"id": "1803156494",
"name": "陳嘉葳"
}
esponse received in 293 ms
Save Session


Docs
Tools & Support
News
Search in docs
work
connections
accounts
activities
adaccounts
albums
apprequests
books
checkins
events
family
feed
friendlists
friendrequests
friends
games
groups
home
inbox
interests
likes
links
locations
movies
music
mutualfriends
notes
notifications
outbox
payments
permissions
photos
picture
posts
questions
scores
statuses
subscribedto
subscribers
tagged
television
updates
videos
Application: [?]
Graph API Explorer
Locale: [?]
English (US)
API Version: [?]
v2.4
Debug
Get Token
Debug Enabled
Submit
Learn more about the Graph API syntax
"id": "1803156494",
"name": "陳嘉葳"
Save Session

Node: me

- id
- name
- posts
  - message
  - + Search for a field
- + Search for a field


```
{
  "id": "1803156494",
  "name": "陳嘉葳",
  "posts": {
    "data": [
      {
        "id": "1803156494_10203280881301813"
      },
      {
        "id": "1803156494_10203270934573151"
      },
      {
        "message": "AI 自己學習互相PK接球遊戲
[code] https://github.com/ugo-nama-kun/DQN-chainer
",
        "id": "1803156494_10203269291052064"
      },
      {
        "message": "今天小育來分享資安方面的議題，好多人一直發問撐超久 XD
林育漢",
        "id": "1803156494_10203266504142393"
      },
      {
        "id": "1803156494_10203266151093567"
      },
      {
        "message": "[CNN] 回來看錄影還是覺得很有趣 XD
感謝 Liang Bo Wang
",
        "id": "1803156494_10203262121032818"
      },
      {
        "message": "scikit-learn 官網從前天掛點到現在，我好怕明天上班怎辦 ...",
        "id": "1803156494_10203261870826563"
```


Response received in 1407 ms

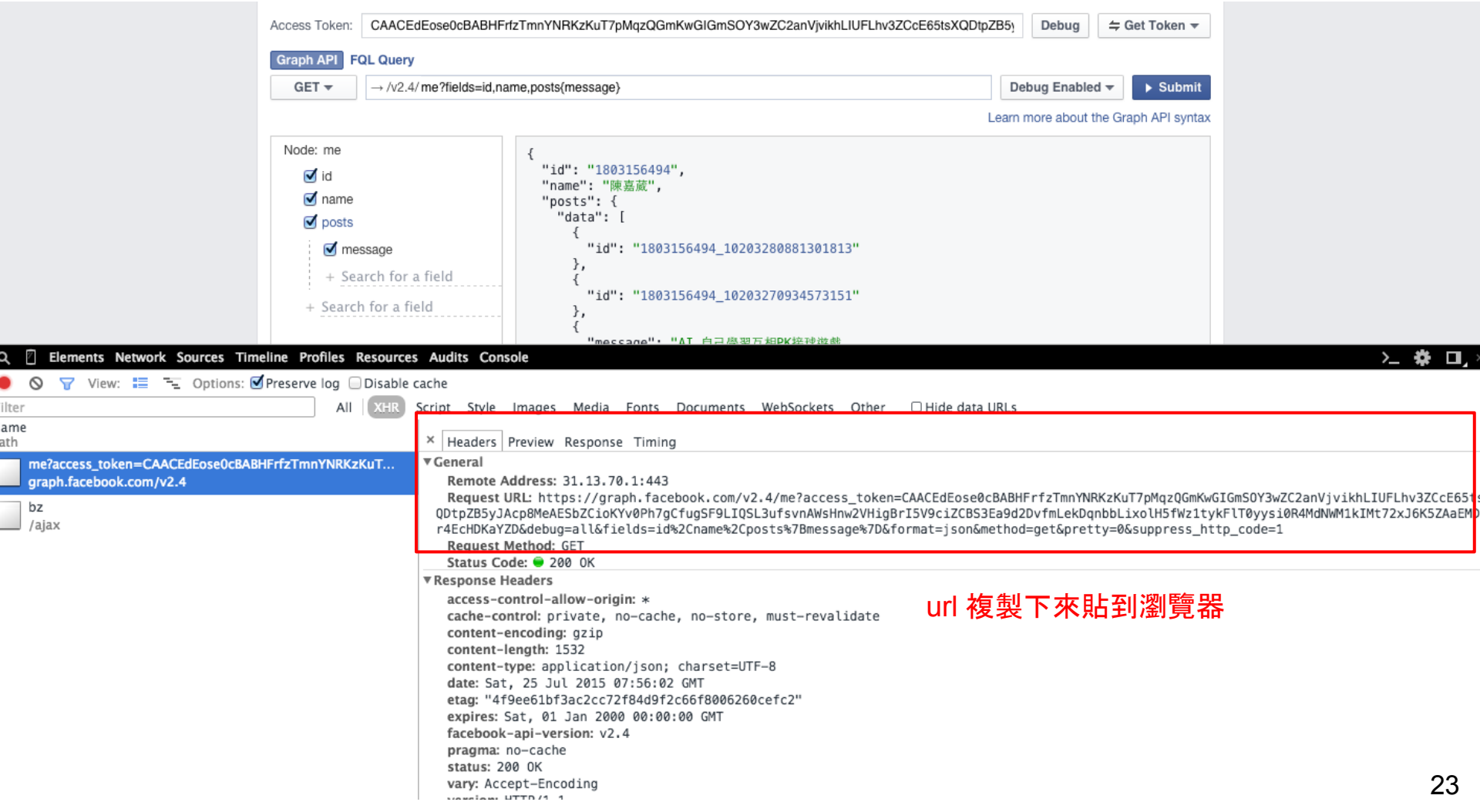
Save Session

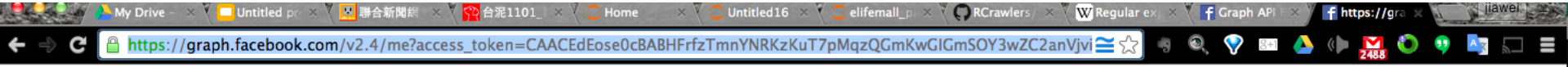
練習 5分鐘

除了 post 還有 message 以外

找找看有沒有其它 你想要的 data :)

要如何用程式拿到
在Graph API Explorer 看到的 data ??


Access Token: CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65tsXQDtpZB5
Debug
Get Token
Graph API
FQL Query
GET
/v2.4/me?fields=id,name,posts{message}
Debug Enabled
Submit
Learn more about the Graph API syntax
Node: me
id
name
posts
message
Search for a field
Search for a field
{
"id": "1803156494",
"name": "陳嘉葳",
"posts": {
"data": [
{
"id": "1803156494_10203280881301813"
},
{
"id": "1803156494_10203270934573151"
},
{
Elements Network Sources Timeline Profiles Resources Audits Console
View: Options: Preserve log Disable cache
All XHR Script Style Images Media Fonts Documents WebSockets Other Hide data URLs
me?access_token=CAACEdEose0cBABHFrfzTmnYNRKzKuT...
graph.facebook.com/v2.4
bz
/ajax
Headers Preview Response Timing
General
Remote Address: 31.13.70.1:443
Request URL: https://graph.facebook.com/v2.4/me?access_token=CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65t
QDtpZB5yJAcp8MeAESbZCioKYv0Ph7gCfugSF9LIQSL3ufsvnAWsHnw2VHigBrI5V9ciZCBS3Ea9d2DvfmLekDqnbbLixolH5fWz1tykFlT0yysi0R4MdNWM1kIMt72xJ6K5ZAaEMD
r4EcHDKaYZD&debug=all&fields=id%2Cname%2Cposts%7Bmessage%7D&format=json&method=get&pretty=0&suppress_http_code=1
Request Method: GET
Status Code: 200 OK
Response Headers
access-control-allow-origin: *
cache-control: private, no-cache, no-store, must-revalidate
content-encoding: gzip
content-length: 1532
content-type: application/json; charset=UTF-8
date: Sat, 25 Jul 2015 07:56:02 GMT
etag: "4f9ee61bf3ac2cc72f84d9f2c66f8006260cefc2"
expires: Sat, 01 Jan 2000 00:00:00 GMT
facebook-api-version: v2.4
pragma: no-cache
status: 200 OK
vary: Accept-Encoding
url 複製下來貼到瀏覽器



{"id":"1803156494","name":"\u9673\u5609\u8473","posts":{"data":[{"id":"1803156494_10203280881301813"},{"id":"1803156494_10203270934573151"},{"message":"AI
\u81ea\u5df1\u5b78\u7fd2\u4e92\u76f8PK\u63a5\u7403\u904a\u6232\n[code] https:\/\/github.com\/ugo-nama-kun\/DQN-chainer\n","id":"1803156494_10203269291052064"},
{"message":"\u4eca\u5929\u5c0f\u80b2\u4f86\u5206\u4eab\u8cc7\u5b89\u65b9\u9762\u7684\u8b70\u984c\uff0c\u597d\u591a\u4eba\u4e00\u76f4\u767c\u554f\u6490\u8d85\u4e45
XD\n\u6797\u80b2\u6f22","id":"1803156494_10203266504142393"},{"id":"1803156494_10203266151093567"},{"message":"[CNN]
\u56de\u4f86\u770b\u9304\u5f71\u9084\u662f\u89ba\u5f97\u5f88\u6709\u8da3 XD\n\u611f\u8b1d Liang Bo Wang\n","id":"1803156494_10203262121032818"},{"message":"scikit-learn
\u5b98\u7db2\u5f9e\u524d\u5929\u639b\u9ede\u5230\u73fe\u5728\uff0c\u6211\u597d\u6015\u660e\u5929\u4e0a\u73ed\u600e\u8fa6 ...","id":"1803156494_10203261870826563"},{"message":"Wilhelm
Fumin \u6628\u5929\u624d\u525b\u804a\u5230\uff0c\u4eca\u5929\u5c31\u770b\u5230\u6709\u5de5\u5177\u53ef\u4ee5\u7528 XD","id":"1803156494_10203261531498080"},
{"id":"1803156494_10203249754323658"},{"id":"1803156494_10203249441395835"},
{"message":"\u90a3\u500b\u767c\u73fe\u51a5\u738b\u661f\u7684\u5e74\u8f15\u4eba","id":"1803156494_10203248470251557"},{"id":"1803156494_10203244761478840"},
{"message":"\u4e09\u5c0f\u6642\u7684\u6700\u4f73\u5316\u6559\u5b78\uff0c\u53ef\u4ee5\u770b\u5f88\u4e45\u4e86 ~~\n","id":"1803156494_10203238922892879"},{"message":"SciPy 2015
\u7684\u5f71\u7247\u5df2\u7d93\u90fd\u653e\u51fa\u4f86\u4e86\uff0c\u597d\u8fc5\u901f\u5594 \uff20\uff20\n","id":"1803156494_10203238907412492"},
{"message":"\u597d\u5f37\u5927\u5594\uff01","id":"1803156494_10203230737368246"},{"message":"Excel power view \u597d\u5f37\u5594\uff01\n\u53ef\u4ee5\u76f4\u63a5\u9023\u7d50 bing map
\u756b3d\u76f4\u65b9\u5716\u548c\u5bc6\u5ea6\u5716\n\u5fae\u8edf\u8d85\u5f37\u7121\u8aa4 \u738b\u4eae\u535a","id":"1803156494_10203230644885934"},
{"id":"1803156494_10203222709087544"},{"message":"http:\/\/icml.cc\/2015\/?page_id=97","id":"1803156494_10203217897607260"},{"id":"1803156494_10203211676451735"},
{"id":"1803156494_10203211265041450"},{"message":"\u4eca\u5e74 COSCUP \u6709\u4e00\u5834\u95dc\u65bc\u795e\u7d93\u5716\u9748\u6a5f \u7684 hands-
on","id":"1803156494_10203208560693843"},{"id":"1803156494_10203208420650342"},{"id":"1803156494_10203206612445138"}],"paging":
{"previous":"https:\/\/graph.facebook.com\/v2.4\/1803156494\/posts?
fields=message&since=1437658261&access_token=CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65tsXQDtpZB5yJAcp8MeAESbZCioKYv0Ph7gCfugSF9LIQSL3ufsvnAWsHnw2VHi
gBrI5V9ciZCBS3Ea9d2DvfmLekDqnbbLixolH5fWz1tykFlT0yysi0R4MdNWM1kIMt72xJ6K5ZAaEMDKr4EcHDKaYZD&limit=25&__paging_token=enc_AdDiCQAR2TDDEcUWnFaOGGGvGmgIIdq4G3QPpvr17HiyJqAcYETxsBUG2pn1Ez
FMXdLE2A8ZBQMfi3ULq33qGF4ZCI&__previous=1","next":"https:\/\/graph.facebook.com\/v2.4\/1803156494\/posts?
fields=message&access_token=CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65tsXQDtpZB5yJAcp8MeAESbZCioKYv0Ph7gCfugSF9LIQSL3ufsvnAWsHnw2VHigBrI5V9ciZCBS3Ea9
d2DvfmLekDqnbbLixolH5fWz1tykFlT0yysi0R4MdNWM1kIMt72xJ6K5ZAaEMDKr4EcHDKaYZD&limit=25&until=1436228395&__paging_token=enc_AdCXGlLw9dZCaKmlTeF4vfMYKd0C6myf8K6tZBtlehGJCC3jV47rf62qvDZCbe
Y8v7uf1SnZALGuWoTIBCPaPnbEVLrZB"}}}

```
In [1]: library(httr)
        library(rjson)

In [2]: url = "https://graph.facebook.com/v2.4/me?access_token=CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIU

In [3]: res = GET(url)

In [4]: content(res)
```

Out[4]:

$id
'1803156494'
$name
'陳嘉葳'
$posts
$data

1. **$id** = '1803156494_10203280881301813'
2. **$id** = '1803156494_10203270934573151'
3. **$message**
   'AI 自己學習互相PK接球遊戲 [code] https://github.com/ugo-nama-kun/DQN-chainer '
   $id
   '1803156494_10203269291052064'
4. **$message**
   '今天小育來分享資安方面的議題，好多人一直發問撐超久 XD 林育漢'
   $id
   '1803156494_10203266504142393'
5. **$id** = '1803156494_10203266151093567'
6. **$message**
   '[CNN] 回來看錄影還是覺得很有趣 XD 感謝 Liang Bo Wang '
   $id
   '1803156494_10203262121032818'

練習 5分鐘

試試看抓其他你想要的 data

如果我想抓 蘋果新聞
facebook 裡面的 data ...

https://www.facebook.com/apple.realtimenews?fref=ts

蘋果日報即時新聞

嘉葳 Home

蘋果日報即時新聞

Media/News/Publishing

Contact Us | Like | Message | •••

Timeline | About | Photos | Likes | More

668k people like this
Joseph Jiang and 6 other friends

Invite friends to like this Page

ABOUT

關掉電視、丟掉報紙！蘋果即時新聞最快最好看

http://www.appledaily.com.tw/realtimenews

Post | Photo / Video

Write something...

Post

蘋果日報即時新聞
2 mins ·

「我十七歲的時候，雖然國文成績優異，不可能寫錯字，但是，我十七歲幹嘛呢？當時的我，有勇氣去質疑所讀的教科書嗎？」

「知識不是只有犀利與正確。知識也有它溫柔感性的一面。知識的溫柔，我們稱為智慧。」

只限於語言表面材料的細節是否正確，也必須通盤整體考量，這一段語文內容所指涉的意涵，從微觀到巨觀，學生經歷了什麼樣成長與補強的過程。

這句話：學生寫這一句話的動機是什麼？這一句話所傳達的意涵又是什

Create Page

Recent
2015
2014
2013

Sponsored

GQ Taiwan
S.S.22 夏日男POPUP Sh...
我們都認真工作，但絕不想因此而放棄貪玩愛自由的Surfing精神。 GQ和LEXUS聯手日本最強選物店BEAMS所打造的 S.S.22 夏日男POPUP Shop, 將成為台灣首...
Thursday, July 23 at 1:00pm

Join Event · 1,260 people are going

要先找出蘋果即時新聞的 facebook ID

可以用 facebook search API 來找 :)

Searching

You can search over many public objects in the social graph with the `/search` endpoint. The format for searches is:

```
GET graph.facebook.com
  /search?
    q={your-query}&
    [type={object-type}](#searchtypes)
```

All Graph API search queries require an access token included in the request. The type of access token you need depends on the type of search you're executing.

- Searches across Page and Place objects requires an app access token.
- All other endpoints require a user access token.

Available Search Types

We support search for the following types:

Type	Description	`q` value
user	Search for a person (if they allow their name to be searched for).	Name.
page	Search for a Page.	Name.
event	Search for an event.	Name.
group	Search for a Group.	Name.
place	Search for a place. You can narrow your search to a specific location by adding the center parameter (with latitude and longitude) and an optional distance parameter:	Name.

Marketing API
Messenger
Pages
Payments for Games
Sharing
Social Plugins
App Development

APIs and SDKs

Graph API
- Using the Graph API
- Reference
- Common Scenarios
- Other APIs
- Advanced

SDK for iOS
SDK for Android
SDK for JavaScript
SDK for PHP
SDK for Unity

We support search for the following types:

Type	Description	`q` value
`user`	Search for a person (if they allow their name to be searched for).	Name.
`page`	Search for a Page.	Name.
`event`	Search for an event.	Name.
`group`	Search for a Group.	Name.
`place`	Search for a place. You can narrow your search to a specific location and distance by adding the `center` parameter (with latitude and longitude) and an optional `distance` parameter:	Name.
`placetopic`	Returns a list of possible place Page topics and their IDs. Use with `topic_filter=all` parameter to get the full list.	None.
`ad_*`	A collection of different search options that can be used to find targeting options.	See Targeting Options docs

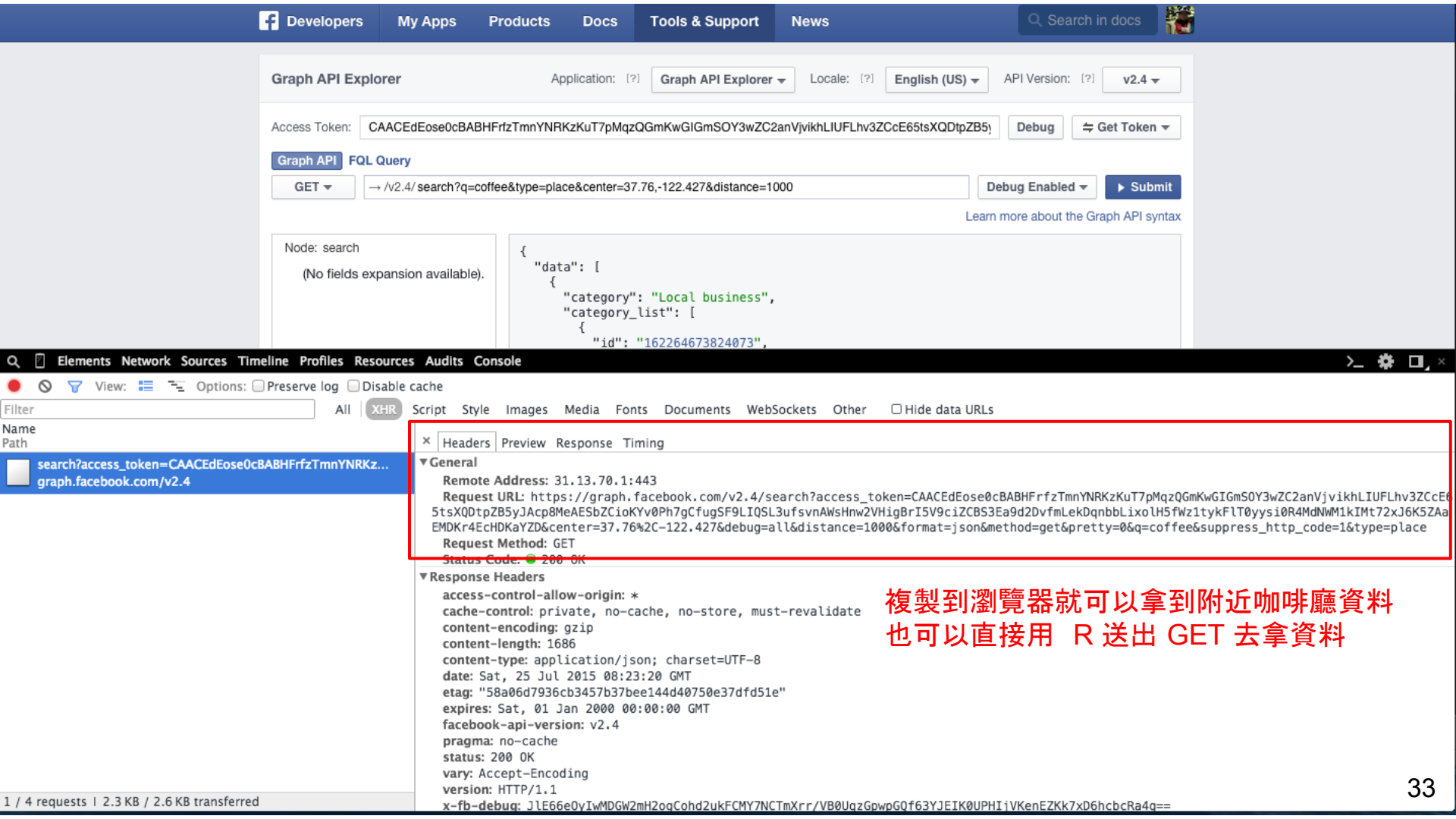
官方文件給的範例,
查詢在這個經緯度周圍一公里 內的咖啡廳

Example:

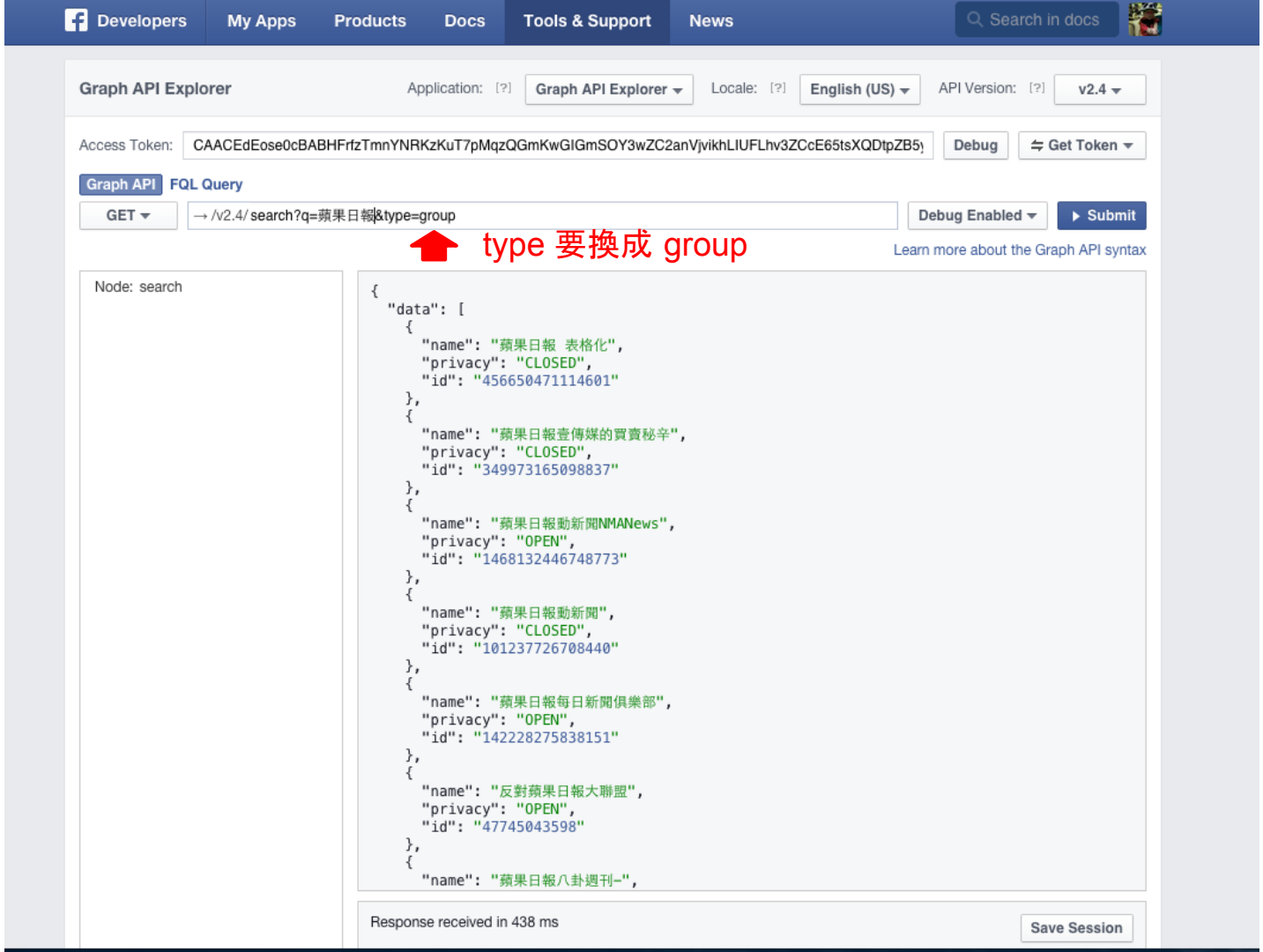
```
GET graph.facebook.com
  /search?
    q=coffee&
    type=place&
    center=37.76,-122.427&
    distance=1000
```

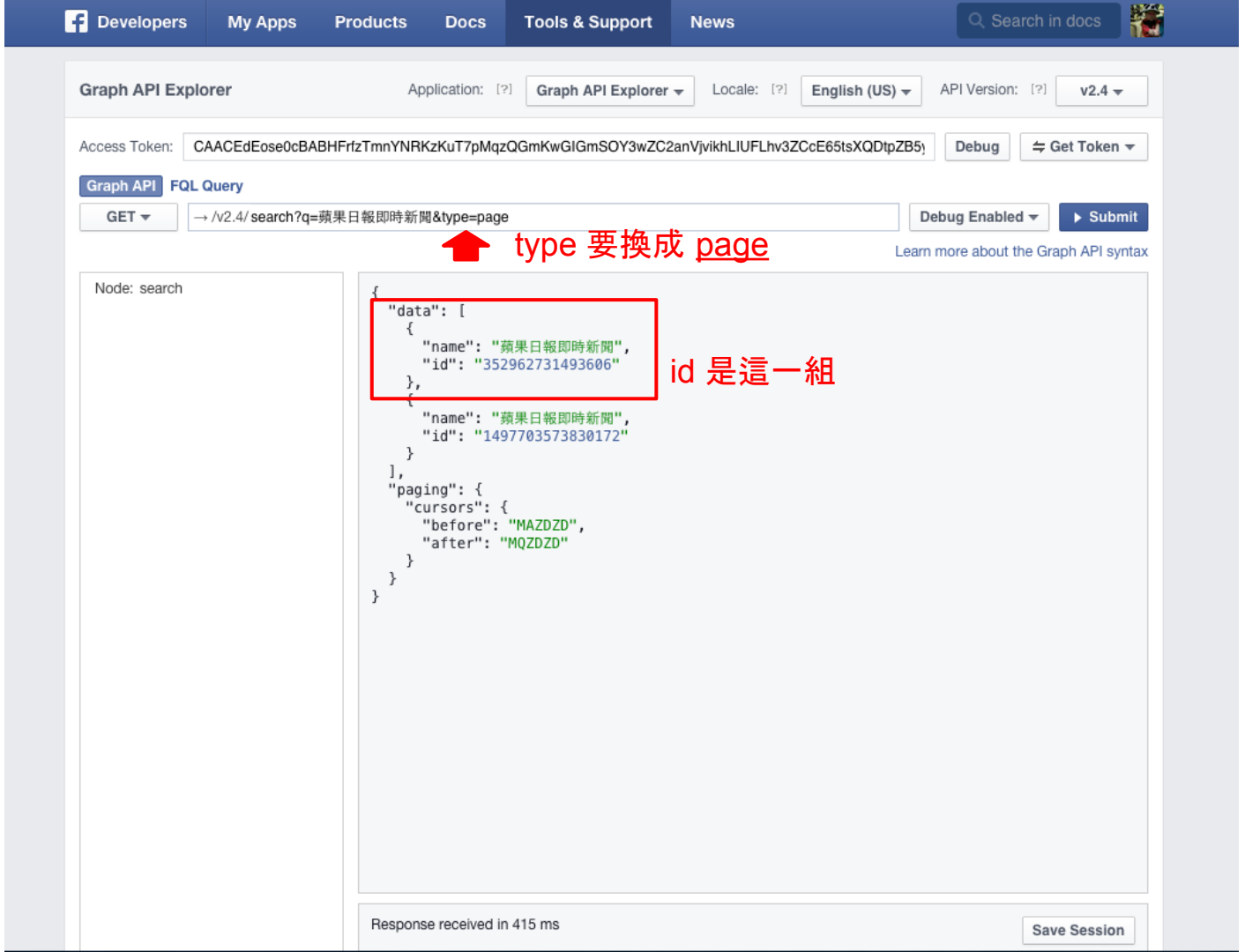
Basics
Reading
Publishing
Updating
Deleting
Searching
Available Search Types
Handling Errors
Debugging API Requests

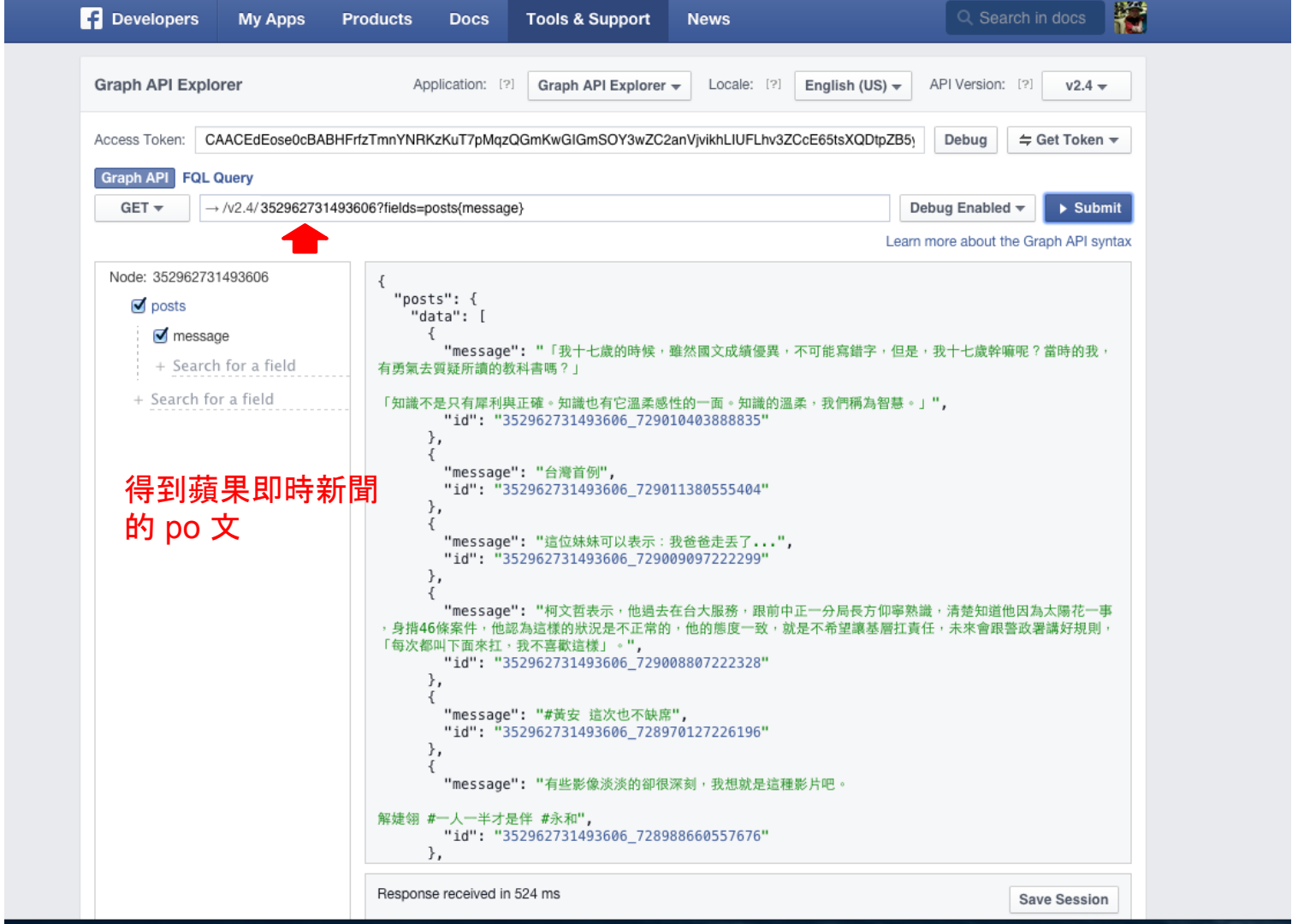

Access Token: CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65tsXQDtpZB5
Debug
Get Token
Graph API
FQL Query
GET
→ /v2.4/ search?q=coffee&type=place&center=37.76,-122.427&distance=1000
Debug Enabled
Submit
Learn more about the Graph API syntax
Node: search
(No fields expansion available).
{
"data": [
{
"category": "Local business",
"category_list": [
{
"id": "162264673824073",
"name": "Specialty Grocery Store"
},
{
"id": "128673187201735",
"name": "Coffee Shop"
}
],
"location": {
"street": "4023 18th St",
"city": "San Francisco",
"state": "CA",
"country": "United States",
"zip": "94114-2501",
"latitude": 37.7607613,
"longitude": -122.4332886
},
"name": "Philz Coffee - Castro",
"id": "151116474914629"
},
{
"category": "Local business",
"category_list": [
{
"id": "197871390225897",
"name": "Cafe"
}
],
第一間
第二間
...
上一頁的範例
在 Graph API Explorer 的用法
第一間咖啡廳的類別名稱
(屬於兩種類別)
第一間咖啡廳的所在國家，街道，經緯度等資訊
第一間咖啡廳的名稱 與 id
Response received in 988 ms
Save Session



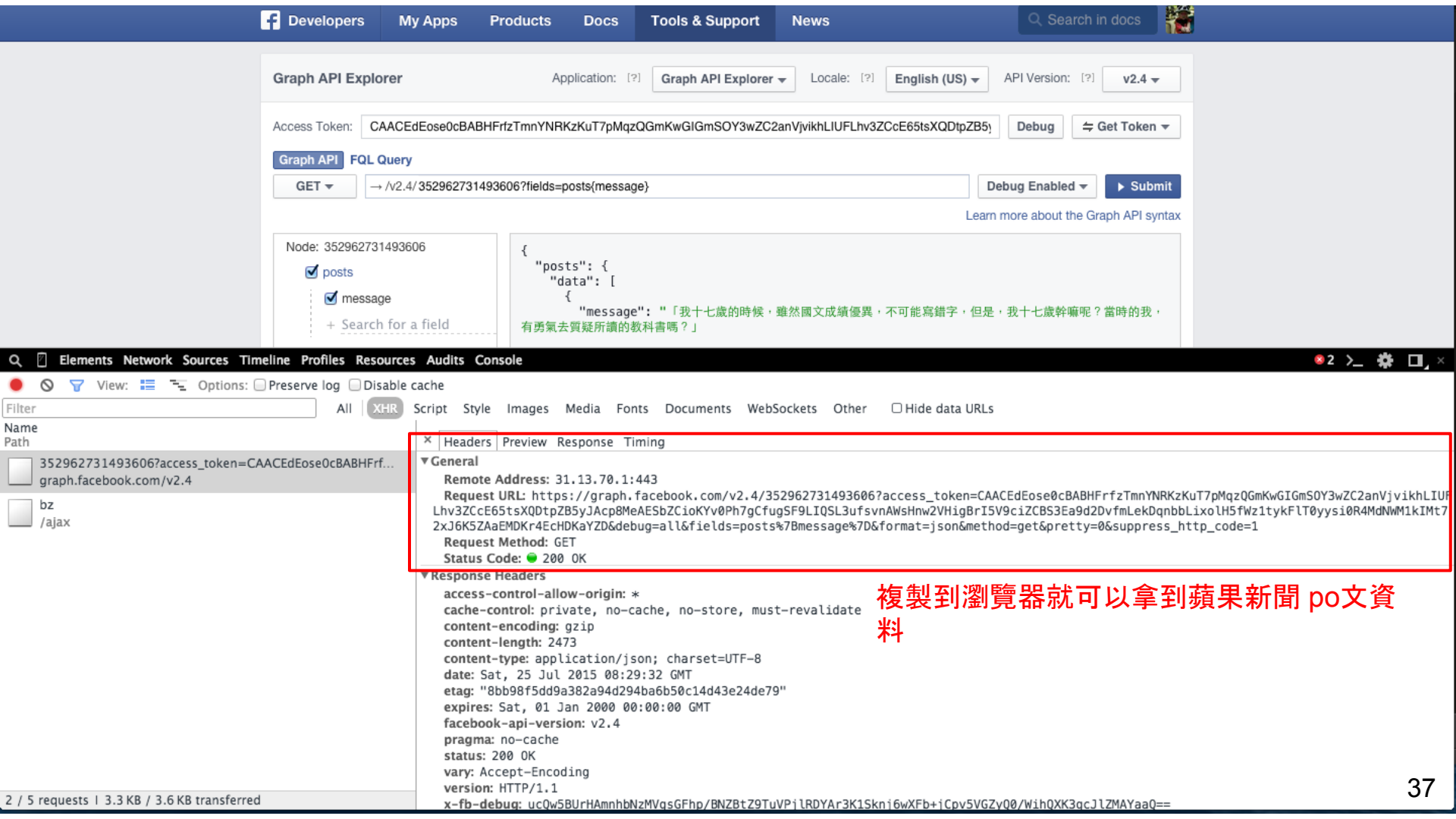
複製到瀏覽器就可以拿到附近咖啡廳資料
也可以直接用 R 送出 GET 去拿資料


Developers
My Apps
Products
Docs
Tools & Support
News
Search in docs
Graph API Explorer
Application: [?] Graph API Explorer
Locale: [?] English (US)
API Version: [?] v2.4
Access Token: CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGlGmSOY3wZC2anVjvikhLlUFLhv3ZCcE65tsXQDtpZB5
Debug
Get Token
Graph API
FQL Query
GET
→ /v2.4/ search?q=蘋果日報&type=group
Debug Enabled
Submit
type 要換成 group
Learn more about the Graph API syntax
Node: search
{
"data": [
{
"name": "蘋果日報 表格化",
"privacy": "CLOSED",
"id": "456650471114601"
},
{
"name": "蘋果日報壹傳媒的買賣秘辛",
"privacy": "CLOSED",
"id": "349973165098837"
},
{
"name": "蘋果日報動新聞NMANews",
"privacy": "OPEN",
"id": "1468132446748773"
},
{
"name": "蘋果日報動新聞",
"privacy": "CLOSED",
"id": "101237726708440"
},
{
"name": "蘋果日報每日新聞俱樂部",
"privacy": "OPEN",
"id": "142228275838151"
},
{
"name": "反對蘋果日報大聯盟",
"privacy": "OPEN",
"id": "47745043598"
},
{
"name": "蘋果日報八卦週刊-",
Response received in 438 ms
Save Session


Developers
My Apps
Products
Docs
Tools & Support
News
Search in docs
Graph API Explorer
Application: [?] Graph API Explorer
Locale: [?] English (US)
API Version: [?] v2.4
Access Token: CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65tsXQDtpZB5
Debug
Get Token
Graph API
FQL Query
GET
→ /v2.4/ search?q=蘋果日報即時新聞&type=page
Debug Enabled
Submit
type 要換成 page
Learn more about the Graph API syntax
Node: search
{
"data": [
{
"name": "蘋果日報即時新聞",
"id": "352962731493606"
},
{
"name": "蘋果日報即時新聞",
"id": "1497703573830172"
}
],
"paging": {
"cursors": {
"before": "MAZDZD",
"after": "MQZDZD"
}
}
}
id 是這一組
Response received in 415 ms
Save Session



得到蘋果即時新聞的 po 文


Developers
My Apps
Products
Docs
Tools & Support
News
Search in docs
Graph API Explorer
Application: [?] Graph API Explorer
Locale: [?] English (US)
API Version: [?] v2.4
Access Token: CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65tsXQDtpZB5
Debug
Get Token
Graph API
FQL Query
GET
/v2.4/352962731493606?fields=posts{message}
Debug Enabled
Submit
Learn more about the Graph API syntax
Node: 352962731493606
posts
message
Search for a field
{
"posts": {
"data": [
{
"message": "「我十七歲的時候，雖然國文成績優異，不可能寫錯字，但是，我十七歲幹嘛呢？當時的我，有勇氣去質疑所讀的教科書嗎？」
Elements
Network
Sources
Timeline
Profiles
Resources
Audits
Console
View:
Options:
Preserve log
Disable cache
Filter
All
XHR
Script
Style
Images
Media
Fonts
Documents
WebSockets
Other
Hide data URLs
Name
Path
352962731493606?access_token=CAACEdEose0cBABHFrf...
graph.facebook.com/v2.4
bz
/ajax
Headers
Preview
Response
Timing
General
Remote Address: 31.13.70.1:443
Request URL: https://graph.facebook.com/v2.4/352962731493606?access_token=CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUF
Lhv3ZCcE65tsXQDtpZB5yJAcp8MeAESbZCioKYv0Ph7gCfugSF9LIQSL3ufsvnAWsHnw2VHigBrI5V9ciZCBS3Ea9d2DvfmLekDqnbbLixolH5fWz1tykFlT0yysi0R4MdNWM1kIMt7
2xJ6K5ZAaEMDKr4EcHDKaYZD&debug=all&fields=posts%7Bmessage%7D&format=json&method=get&pretty=0&suppress_http_code=1
Request Method: GET
Status Code: 200 OK
Response Headers
access-control-allow-origin: *
cache-control: private, no-cache, no-store, must-revalidate
content-encoding: gzip
content-length: 2473
content-type: application/json; charset=UTF-8
date: Sat, 25 Jul 2015 08:29:32 GMT
etag: "8bb98f5dd9a382a94d294ba6b50c14d43e24de79"
expires: Sat, 01 Jan 2000 00:00:00 GMT
facebook-api-version: v2.4
pragma: no-cache
status: 200 OK
vary: Accept-Encoding
version: HTTP/1.1
x-fb-debug: ucQw5BUrHAmnhbNzMVqsGFhp/BNZBtZ9TuVPjlRDYAr3K1Skni6wXFb+jCpv5VGZyQ0/WihQXK3qcJlZMAYaaQ==
2 / 5 requests | 3.3 KB / 3.6 KB transferred
複製到瀏覽器就可以拿到蘋果新聞 po文資料

https://graph.facebook.com/v2.4/352962731493606?access_token=CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGI...


```
{"posts":{"data":
[{"message":"\u654d\u5229\u4e9e\u3001\u5229\u6bd4\u4e9e\u53ca\u5384\u5229\u5782\u4e9e\u653f\u5c40\u52d5\u76ea\uff0c\u88fd\u9020\u51fa\u5927\u91cf\u8a66\u5716\u5077\u6e21\u5f80\u6b50\
u6d32\u7684\u96e3\u6c11...\u4f46\u537b\u4e0d\u662f\u6bcf\u500b\u4eba\u90fd\u6709\u80fd\u6d3b\u4e0b\u4f86\u7684\u597d\u904b\u6c23...","id":"352962731493606_729009783888897"},
{"message":"\u300c\u6211\u5341\u4e03\u6b72\u7684\u6642\u5019\uff0c\u96d6\u7136\u570b\u6587\u6210\u7e3e\u512a\u7570\uff0c\u4e0d\u53ef\u80fd\u5beb\u932f\u5b57\uff0c\u4f46\u662f\uff0c\u
6211\u5341\u4e03\u6b72\u5e79\u561b\u5462\uff1f\u7576\u6642\u7684\u6211\uff0c\u6709\u52c7\u6c23\u53bb\u8cea\u7591\u6240\u8b80\u7684\u6559\u79d1\u66f8\u55ce\uff1f\u300d\n\n\u300c\u77e5
\u8b58\u4e0d\u662f\u53ea\u6709\u7280\u5229\u8207\u6b63\u78ba\u3002\u77e5\u8b58\u4e5f\u6709\u5b83\u6eab\u67d4\u611f\u6027\u7684\u4e00\u9762\u3002\u77e5\u8b58\u7684\u6eab\u67d4\uff0c\u
6211\u5011\u7a31\u70ba\u667a\u6167\u3002\u300d","id":"352962731493606_729010403888835"},{"message":"\u53f0\u7063\u9996\u4f8b","id":"352962731493606_729011380555404"},
{"message":"\u9019\u4f4d\u59b9\u59b9\u53ef\u4ee5\u8868\u793a\uff1a\u6211\u7238\u7238\u8d70\u4e1f\u4e86...","id":"352962731493606_729009097222299"},
{"message":"\u67ef\u6587\u54f2\u8868\u793a\uff0c\u4ed6\u904e\u53bb\u5728\u53f0\u5927\u670d\u52d9\uff0c\u8ddf\u524d\u4e2d\u6b63\u4e00\u5206\u5c40\u9577\u65b9\u4ef0\u5be7\u719f\u8b58\u
ff0c\u6e05\u695a\u77e5\u9053\u4ed6\u56e0\u70ba\u592a\u967d\u82b1\u4e00\u4e8b\uff0c\u8eab\u63f946\u689d\u6848\u4ef6\uff0c\u4ed6\u8a8d\u70ba\u9019\u6a23\u7684\u72c0\u6cc1\u662f\u4e0d\u
6b63\u5e38\u7684\uff0c\u4ed6\u7684\u614b\u5ea6\u4e00\u81f4\uff0c\u5c31\u662f\u4e0d\u5e0c\u671b\u8b93\u57fa\u5c64\u625b\u8cac\u4efb\uff0c\u672a\u4f86\u6703\u8ddf\u8b66\u653f\u7f72\u8b
1b\u597d\u898f\u5247\uff0c\u300c\u6bcf\u6b21\u90fd\u53eb\u4e0b\u9762\u4f86\u625b\uff0c\u6211\u4e0d\u559c\u6b61\u9019\u6a23\u300d\u3002","id":"352962731493606_729008807222328"},
{"message":"#\u9ec3\u5b89 \u9019\u6b21\u4e5f\u4e0d\u7f3a\u5e2d","id":"352962731493606_728970127226196"},
{"message":"\u6709\u4e9b\u5f71\u50cf\u6de1\u6de1\u7684\u537b\u5f88\u6df1\u523b\uff0c\u6211\u60f3\u5c31\u662f\u9019\u7a2e\u5f71\u7247\u5427\u3002\n\n\u89e3\u5a55\u7fce
#\u4e00\u4eba\u4e00\u534a\u624d\u662f\u4f34 #\u6c38\u548c","id":"352962731493606_728988660557676"},{"message":"\u8b9a\n\n#\u53f0\u9435","id":"352962731493606_728967820559760"},
{"message":"#\u4e2d\u80af\u6587\n\n#\u4e8b\u60c5\u4e0d\u662f\u4f60\u60f3\u7684\u90a3\u9ebc\u7c21\u55ae","id":"352962731493606_728950990561443"},
{"message":"\u597d\u96e2\u8b5c\u7684\u5927\u8eca\n\n#\u5927\u8eca\u5c31\u662f\u4efb\u6027","id":"352962731493606_728950230561519"},
{"message":"\u5616\u5616\u5616\uff0c\u4f60\u77e5\u9053\u7684\u592a\u591a\u4e86\n\n#\u5f97\u7f6a\u65b9\u4e08\u9084\u60f3\u8d70","id":"352962731493606_728906400565902"},
{"message":"\u5c0d\u767c\u7968\u7684\u5feb\u6a02\u65e5\u5b50\u53c8\u4f86\u56c9\uff0c\u795d\u5927\u5bb6\u4e2d\u734e\n\n#\u7d71\u4e00\u767c\u7968","id":"352962731493606_728947397228469
"},
{"message":"\u53d7\u50b7\u6700\u91cd\u7684\u662f\u9a0e\u58eb\uff0c\u6700\u7121\u8f9c\u7684\u4e5f\u662f\u9a0e\u58eb\n\n#\u8def\u908a\u958b\u8eca\u9580","id":"352962731493606_728905737
232635"},{"message":"\u9a0e\u58eb\u4e00\u53bb\u4e0d\u56de\u982d\u5440\n\n#\u963f\u65af\u62c9 #\u98a8\u898b\u96bc\u4eba
#\u4eca\u5929\u7684\u6211\u6c92\u6709\u6975\u9650","id":"352962731493606_728905600565982"},
{"message":"\u4f54\u9818\u90e8\u9577\u5ba4\u672a\u6210\u5e74\u5b78\u751f\u8072\u660e\n\n#\u6700\u65b0","id":"352962731493606_728908663899009"},
{"message":"\u597d\u73cd\u8cb4\u7684\u756b\u9762\n\n#\u8b93\u7238\u5abd\u770b\u4e00\u4e0b #\u8a18\u5f97\u7576\u6642\u5e74\u7d00\u5c0f","id":"352962731493606_728904960566046"},
{"message":"\u883b\u53ef\u611b\u7684\u5440\n\n#\u5fc3\u610f\u6700\u91cd\u8981 #\u4e0d\u6536\u53ef\u80fd\u66f4\u6050\u6016","id":"352962731493606_728865293903346"},
{"message":"\u8e39\u5171\u5566\uff01","id":"352962731493606_728865147236694"},
{"message":"\u5927\u5bb6\u9ede\u9032\u65b0\u805e\u5f8c\uff0c\u61c9\u8a72\u8ddf\u5c0f\u7de8\u6709\u4e00\u6a23\u7684\u611f\u89ba\n\n#\u8b66\u5bdf #\u5c0f\u96bb\u99ac
\n\n\n\n\n\n\n\n\n\n\n\n\n\n\n\u8aaa\u597d\u7684\u5c0f\u96bb\u99ac\u5973\u8b66\u82b1\u54a7\uff01","id":"352962731493606_728890053900870"},
{"message":"50\u5d50\u53c8\u51fa\u5305\u4e86.\n\n#50\u5d50  #\u8702\u871c","id":"352962731493606_728864973903378"},
{"message":"\u800d\u5e25\u8acb\u5728\u5bb6\n\n#\u99d5\u7167\u53ef\u4ee5\u91cd\u8003\u55ce #\u4e0a\u8def\u8acb\u5c0f\u5fc3
#\u53f0\u4e2d\u5e02","id":"352962731493606_728843170572225"},{"message":"\u54c8\u54c8\u54c8\u54c8\u54c8\u54c8\u54c8\n\n#\u7db2\u53cb\u5275\u610f\u7121\u9650 #41J\u4e5f\u662f\u65bd
#\u4f86\u5e6b\u5e6b\u5979\u5427","id":"352962731493606_728842973905578"},
{"message":"\u8001\u8a71\u4e00\u53e5\n\n#\u5b69\u5b50\u7684\u6559\u80b2\u4e0d\u80fd\u7b49","id":"352962731493606_728818247241384"},
{"message":"#\u6700\u65b0\n\n#\u4eca\u5929\u4e0d\u80fd\u8dd1\u5c71 #\u6ff1\u6d77\u516c\u8def","id":"352962731493606_728842057239003"},
{"message":"\u662f\u50b3\u8aaa\u4e2d\u624d\u6703\u51fa\u73fe\u7684\u56db\u8173\u795e\u7378\u55ce\uff01","id":"352962731493606_728818180574724"}],"paging":
{"previous":"https:\/\/graph.facebook.com\/v2.4\/352962731493606\/posts?
fields=message&since=1437813000&access_token=CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65tsXQDtpZB5yJAcp8MeAESbZCioKYv0Ph7gCfugSF9LIQSL3ufsvnAWsHnw2VHi
gBrI5V9ciZCBS3Ea9d2DvfmLekDqnbbLixolH5fWz1tykFlT0yysi0R4MdNWM1kIMt72xJ6K5ZAaEMDKr4EcHDKaYZD&limit=25&__paging_token=enc_AdCXDC4b5xuasUc2aHfM4UJCaFCNK34EkkVfL9b5gvOeZAFSxSYxiUc30JaKgi
es9nkKH49cRA6u6RaWVB61spiHUMuGm8sXjtZAF8JI019xQZBfgZDZD&__previous=1","next":"https:\/\/graph.facebook.com\/v2.4\/352962731493606\/posts?
fields=message&access_token=CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZC2anVjvikhLIUFLhv3ZCcE65tsXQDtpZB5yJAcp8MeAESbZCioKYv0Ph7gCfugSF9LIQSL3ufsvnAWsHnw2VHigBrI5V9ciZCBS3Ea9
d2DvfmLekDqnbbLixolH5fWz1tykFlT0yysi0R4MdNWM1kIMt72xJ6K5ZAaEMDKr4EcHDKaYZD&limit=25&until=1437784201&__paging_token=enc_AdAe9puRu7K2s8CwxuFB4TGIbQsYgFMAtJahEo8YKDkfUSSv41DGFqfIM1QptB
ZBzUQZAf2Vg7D9ZBCu4elfbHNNCiHnRuqDh2QT4iz5ESw7walJQZDZD"}},"id":"352962731493606"}
```

```
In [1]: library(httr)
        library(rjson)

In [14]: url = "https://graph.facebook.com/v2.4/352962731493606?access_token=CAACEdEose0cBABHFrfzTmnYNRKzKuT7pMqzQGmKwGIGmSOY3wZ(

In [15]: res = content(GET(url))

In [16]: res
```

Out[16]: **$posts**
$data

1. **$message**
   '敘利亞、利比亞及厄利垂亞政局動盪，製造出大量試圖偷渡往歐洲的難民...但卻不是每個人都有能活下來的好運氣...'
   $id
   '352962731493606_729009783888897'

2. **$message**
   '「我十七歲的時候，雖然國文成績優異，不可能寫錯字，但是，我十七歲幹嘛呢？當時的我，有勇氣去質疑所讀的教科書嗎？」「知識不是只有犀利與正確。知識也有它溫柔感性的一面。知識的溫柔，我們稱為智慧。」'
   $id
   '352962731493606_729010403888835'

3. **$message**
   '台灣首例'
   $id
   '352962731493606_729011380555404'

4. **$message**
   '這位妹妹可以表示：我爸爸走丟了...'
   $id
   '352962731493606_729009097222299'

5. **$message**

透過 R 程式拿到蘋果新聞 po文資料

練習 5分鐘

res$ … 底下還有 post, paging …

請觀察一下這些是在做什麼 :）

把抓下來的 data 進行整理
變成表格的樣子 ！

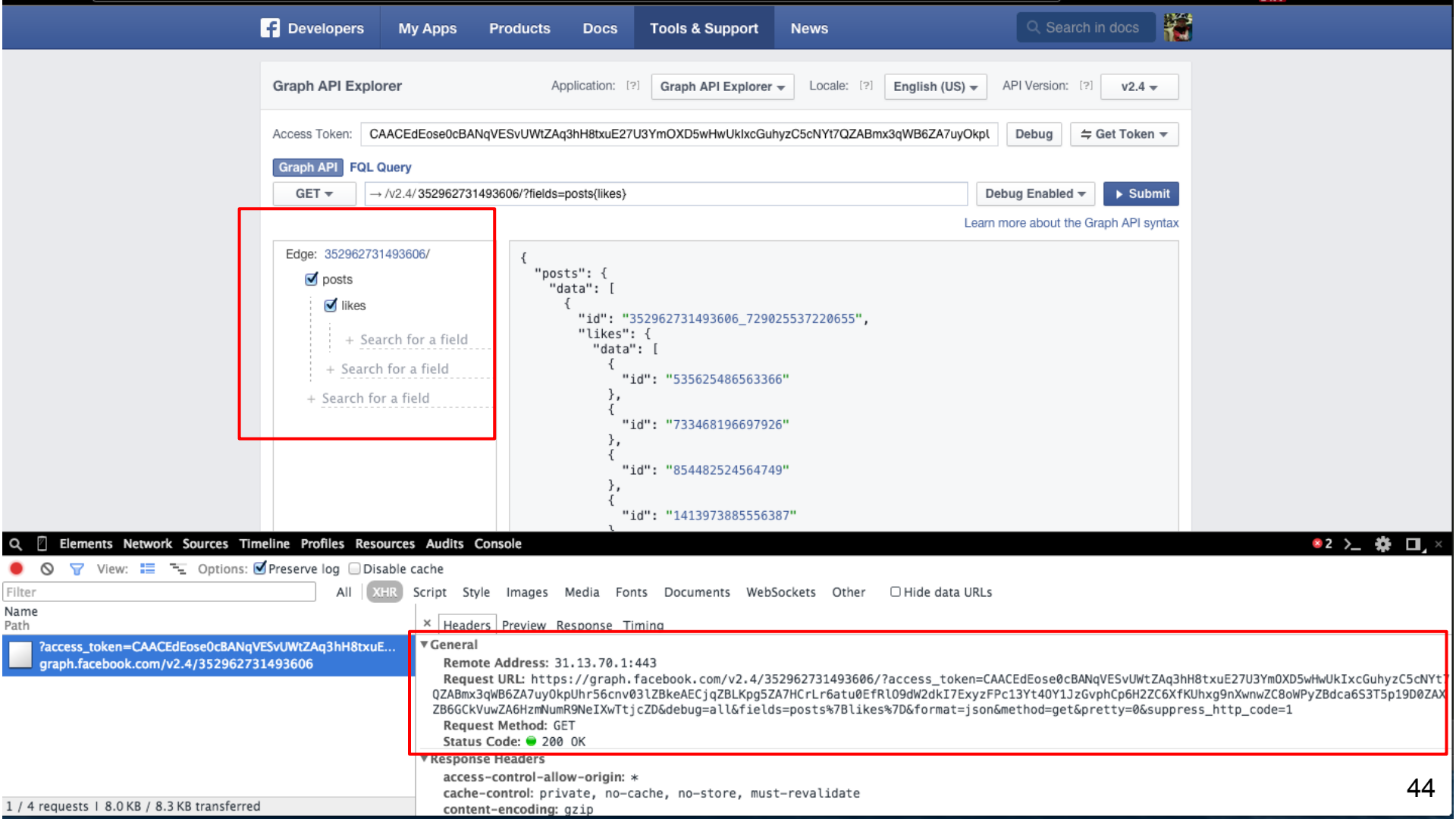
In [18]:
```
do.call(rbind, res$posts$data)
```

Out[18]:

message	id
敘利亞、利比亞及厄利垂亞政局動盪，製造出大量試圖偷渡往歐洲的難民...但卻不是每個人都有能活下來的好運氣...	352962731493606_729009783888897
「我十七歲的時候，雖然國文成績優異，不可能寫錯字，但是，我十七歲幹嘛呢？當時的我，有勇氣去質疑所讀的教科書嗎？」「知識不是只有犀利與正確。知識也有它溫柔感性的一面。知識的溫柔，我們稱為智慧。」	352962731493606_729010403888835
台灣首例	352962731493606_729011380555404
這位妹妹可以表示：我爸爸走丟了...	352962731493606_729009097222299
柯文哲表示，他過去在台大服務，跟前中正一分局長方仰寧熟識，清楚知道他因為太陽花一事，身揹46條案件，他認為這樣的狀況是不正常的，他的態度一致，就是不希望讓基層扛責任，未來會跟警政署講好規則，「每次都叫下面來扛，我不喜歡這樣」。	352962731493606_729008807222328
#黃安 這次也不缺席	352962731493606_728970127226196
有些影像淡淡的卻很深刻，我想就是這種影片吧。 解婕翎 #一人一半才是伴 #永和	352962731493606_728988660557676
讚 #台鐵	352962731493606_728967820559760
#中肯文 #事情不是你想的那麼簡單	352962731493606_728950990561443
好離譜的大車 #大車就是任性	352962731493606_728950230561519
嘖嘖嘖，你知道的太多了 #得罪方丈還想走	352962731493606_728906400565902
對發票的快樂日子又來囉，祝大家中獎 #統一發票	352962731493606_728947397228469
受傷最重的是騎士，最無辜的也是騎士 #路邊開車門	352962731493606_728905737232635
騎士一去不回頭呀 #阿斯拉 #風見隼人 #今天的我沒有極限	352962731493606_728905600565982
佔領部長室未成年學生聲明 #最新	352962731493606_728908663899009
好珍貴的畫面 #讓爸媽看一下 #記得當時年紀小	352962731493606_728904960566046
蠻可愛的呀 #心意最重要 #不收可能更恐怖	352962731493606_728865293903346
蹋共啦！	352962731493606_728865147236694

常聽說 ...
要抓 facebook 的按👍
要自己一直往下捲很麻煩 :(

(要做社群分析, 或觀點挖掘)


Developers
My Apps
Products
Docs
Tools & Support
News
Search in docs
Graph API Explorer
Application: [?] Graph API Explorer
Locale: [?] English (US)
API Version: [?] v2.4
Access Token: CAACEdEose0cBANqVESvUWtZAq3hH8txuE27U3YmOXD5wHwUkIxcGuhyzC5cNYt7QZABmx3qWB6ZA7uyOkpl
Debug
Get Token
Graph API
FQL Query
GET
→ /v2.4/352962731493606/?fields=posts{likes}
Debug Enabled
Submit
Learn more about the Graph API syntax
Edge: 352962731493606/
posts
likes
+ Search for a field
+ Search for a field
+ Search for a field
{
"posts": {
"data": [
{
"id": "352962731493606_729025537220655",
"likes": {
"data": [
{
"id": "535625486563366"
},
{
"id": "733468196697926"
},
{
"id": "854482524564749"
},
{
"id": "1413973885556387"
Elements
Network
Sources
Timeline
Profiles
Resources
Audits
Console
View:
Options:
Preserve log
Disable cache
Filter
All
XHR
Script
Style
Images
Media
Fonts
Documents
WebSockets
Other
Hide data URLs
Name
Path
?access_token=CAACEdEose0cBANqVESvUWtZAq3hH8txuE...
graph.facebook.com/v2.4/352962731493606
Headers
Preview
Response
Timing
General
Remote Address: 31.13.70.1:443
Request URL: https://graph.facebook.com/v2.4/352962731493606/?access_token=CAACEdEose0cBANqVESvUWtZAq3hH8txuE27U3YmOXD5wHwUkIxcGuhyzC5cNYt7QZABmx3qWB6ZA7uyOkpUhr56cnv03lZBkeAECjqZBLKpg5ZA7HCrLr6atu0EfRlO9dW2dkI7ExyzFPc13Yt4OY1JzGvphCp6H2ZC6XfKUhxg9nXwnwZC8oWPyZBdca6S3T5p19D0ZAXZB6GCkVuwZA6HzmNumR9NeIXwTtjcZD&debug=all&fields=posts%7Blikes%7D&format=json&method=get&pretty=0&suppress_http_code=1
Request Method: GET
Status Code: 200 OK
Response Headers
access-control-allow-origin: *
cache-control: private, no-cache, no-store, must-revalidate
content-encoding: gzip
1 / 4 requests | 8.0 KB / 8.3 KB transferred

```
In [21]: url = "https://graph.facebook.com/v2.4/352962731493606/?access_token=CAACEdEose0cBANqVESvUWtZAq3hH8txuE27U3YmOXD5wHwUkI
```

```
In [22]: res = content(GET(url))
```

```
In [23]: res$posts$data
```

```
Out[23]: 1. $id
         '352962731493606_729025537220655'
         $likes
         $data
           A. $id = '256515907884819'
           B. $id = '813815345305329'
           C. $id = '614920171928586'
           D. $id = '1488454388101539'
           E. $id = '649265348520420'
           F. $id = '667586220013890'
           G. $id = '737110892979155'
           H. $id = '253410108192365'
            I. $id = '1444147985909755'
           J. $id = '1519188568354590'
           K. $id = '1423990601202092'
           L. $id = '719911508103768'
           M. $id = '1377375812678572'
```

```
In [ ]:
```

```
In [ ]:
```

```
In [26]: res$posts$data[[1]]
```

```
Out[26]: $id
'352962731493606_729025537220655'
$likes
$data
  1. $id = '256515907884819'
  2. $id = '813815345305329'
  3. $id = '614920171928586'
  4. $id = '1488454388101539'
  5. $id = '649265348520420'
  6. $id = '667586220013890'
  7. $id = '737110892979155'
  8. $id = '253410108192365'
  9. $id = '1444147985909755'
 10. $id = '1519188568354590'
 11. $id = '1423990601202092'
 12. $id = '719911508103768'
 13. $id = '1377375812578572'
 14. $id = '763108580380072'
 15. $id = '864220310261151'
 16. $id = '668151756588804'
 17. $id = '225784700956218'
 18. $id = '103922183276595'
 19. $id = '774755135879740'
 20. $id = '715041565258389'
 21. $id = '1008525422494853'
 22. $id = '850074498425143'
 23. $id = '1474845946130861'
 24. $id = '1587771148155712'
 25. $id = '1428255687495149'
$paging
$cursors
```

一次給 25筆,
需要自己不斷去 GET 下一頁

(蘋果新聞平均一篇文章超過 500 👍)

```
In [27]: res$posts$data[[2]]
```

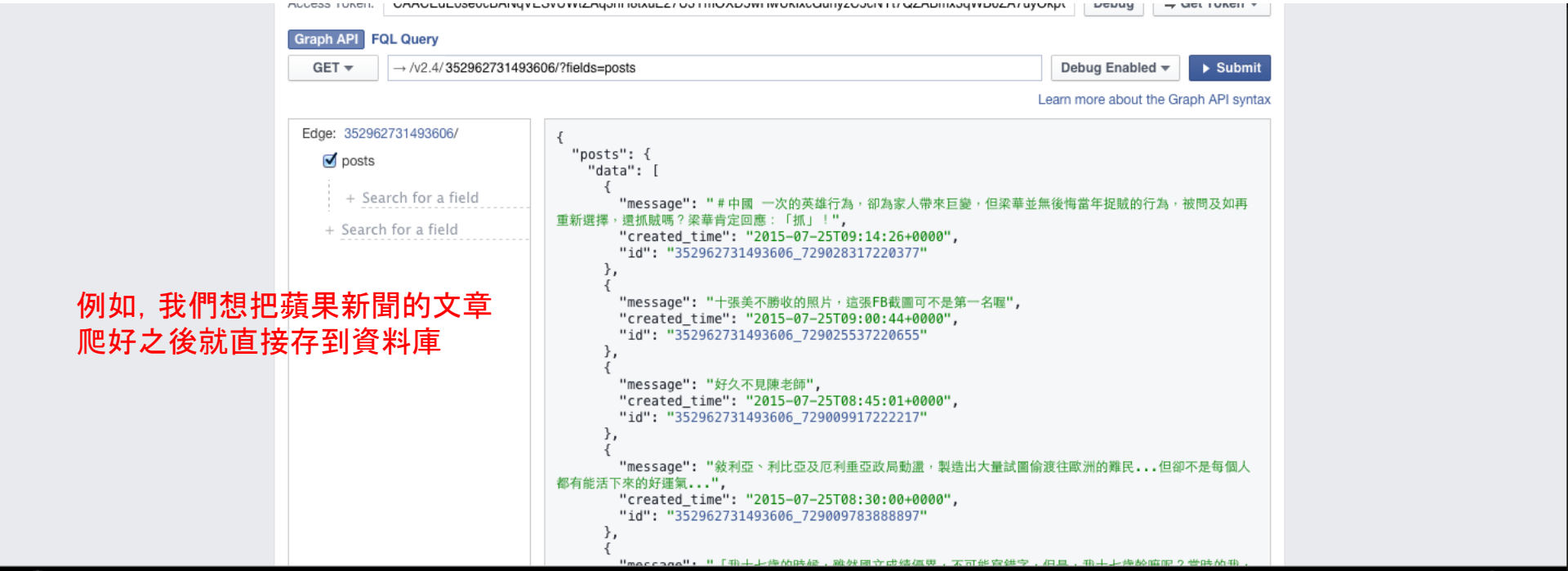
```
Out[27]: $id
'352962731493606_729009917222217'
$likes
$data
  1. $id = '642486195894961'
  2. $id = '806449262700338'
  3. $id = '865610850122726'
  4. $id = '238986379639869'
  5. $id = '1568492620042765'
  6. $id = '10202824355027728'
  7. $id = '710321349032346'
  8. $id = '861015497249010'
  9. $id = '1381075902193874'
 10. $id = '810318048989326'
 11. $id = '1554420924779466'
 12. $id = '110992655904806'
 13. $id = '918824734803259'
 14. $id = '234064736791495'
 15. $id = '809347762412102'
 16. $id = '532642473514928'
 17. $id = '386630278143094'
 18. $id = '577653822332246'
 19. $id = '695334383841746'
 20. $id = '835034753184482'
 21. $id = '843649072318968'
 22. $id = '1431661330428316'
 23. $id = '1519188568354590'
 24. $id = '853438874681914'
 25. $id = '253319124854701'
$paging
$cursors
$after
'MjUzMzE5MTI0ODU0NzAx'
```

不斷 for loop 就對了
太瑣碎就交給同學們自己嘗試

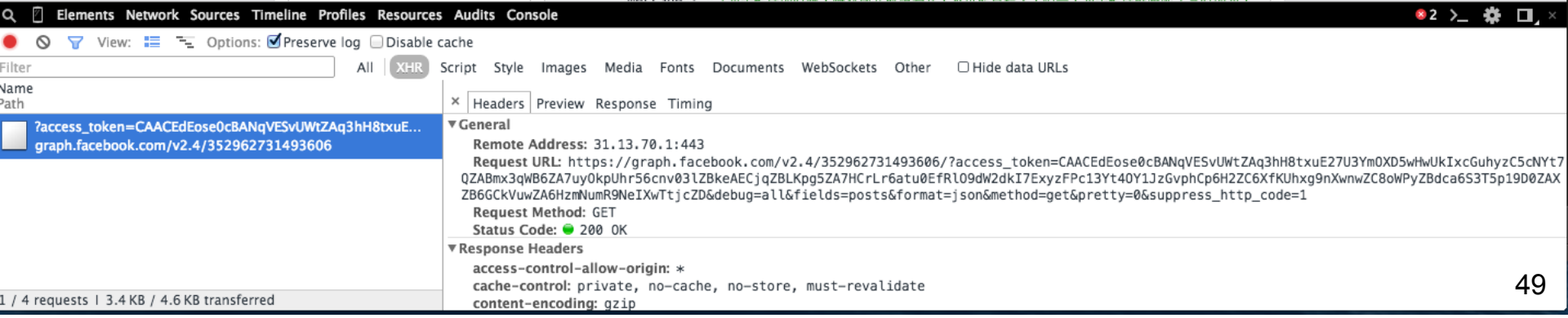
順便出個 A Crawler A Day 題目

大家可以做一下 ”觀點分析”

抓好的 data 可以直接
存到資料庫嗎？



例如，我們想把蘋果新聞的文章
爬好之後就直接存到資料庫



```
In [2]: library(httr)
        library(stringr)
        library(dplyr)
```

```
In [3]: url = "https://graph.facebook.com/v2.4/352962731493606/?access_token=CAACEdEose0cBAPtikGArlT6lXkEgynRnsZCofv87ioaB6F2Jx
```

```
In [4]: res = content(GET(url))
```

```
In [5]: res
```

```
Out[5]: $posts
        $data
          1. $message
             '#最新'
             $created_time
             '2015-07-26T03:56:53+0000'
             $id
             '352962731493606_729357620520780'

          2. $message
             '用影片讓人更了解過去'
             $created_time
             '2015-07-26T03:50:00+0000'
             $id
             '352962731493606_729335113856364'

          3. $message
```

先把 蘋果新聞的 po文抓下來

```
In [6]: tbl = lapply(res$posts$data, data.frame, stringsAsFactors=FALSE)
```

```
In [7]: tbl
```

Out[7]:

1.

	message	created_time	id
1	#最新	2015-07-26T03:56:53+0000	352962731493606_729357620520780

2.

	message	created_time	id
1	用影片讓人更了解過去	2015-07-26T03:50:00+0000	352962731493606_729335113856364

3.

	message	created_time	id
1	曹錦輝下放3A	2015-07-26T03:40:11+0000	352962731493606_729354070521135

4.

	message	created_time	id
1	好奇拍攝出來會是什麼樣子	2015-07-26T03:20:01+0000	352962731493606_729334647189744

5.

```
In [8]: tbl = rbind_all(tbl)
```

```
In [9]: head(tbl)
```

整理成表格

Out[9]:

	message	created_time	id
1	#最新	2015-07-26T03:56:53+0000	352962731493606_729357620520780
2	用影片讓人更了解過去	2015-07-26T03:50:00+0000	352962731493606_729335113856364
3	曹錦輝下放3A	2015-07-26T03:40:11+0000	352962731493606_729354070521135
4	好奇拍攝出來會是什麼樣子	2015-07-26T03:20:01+0000	352962731493606_729334647189744
5	該架小型飛機事發時載著5個人，機上兩人以及地面一名居民可能走避不及....	2015-07-26T03:05:36+0000	352962731493606_729340873855788

時間最好要切開來儲存 …
以後要 Query 比較方便

ex. 20150725 , 130513

可以回想前面的正規表達,
我們使用 stringr 套件的 str_replace function

把 created_time 欄位內的 - : +0000 給拿掉

```
In [10]: tbl$created_time = str_replace_all(tbl$created_time, "-|:|[+]0000", "")
```

```
In [11]: tbl$created_time[1:10]
```

```
Out[11]: '20150726T035653' '20150726T035000' '20150726T034011' '20150726T032001' '20150726T030536' '20150726T025000'
         '20150726T024401' '20150726T022000' '20150726T015001' '20150726T005001'
```

```
In [12]: str_split(tbl$created_time, "T")[1:10]
```

```
Out[12]:
1.  '20150726'  '035653'
2.  '20150726'  '035000'
3.  '20150726'  '034011'
4.  '20150726'  '032001'
5.  '20150726'  '030536'
6.  '20150726'  '025000'
7.  '20150726'  '024401'
8.  '20150726'  '022000'
9.  '20150726'  '015001'
10. '20150726'  '005001'
```

利用 T 這個字，把日期時間給切開

```
In [13]: time_tbl = do.call(rbind, str_split(tbl$created_time, "T"))

In [14]: time_tbl = as.data.frame(time_tbl, stringsAsFactors = FALSE)

In [15]: names(time_tbl) = c('date', 'time')

In [16]: head(time_tbl)
```

Out[16]:

	date	time
1	20150726	035653
2	20150726	035000
3	20150726	034011
4	20150726	032001
5	20150726	030536
6	20150726	025000

把切開的時間日期, 合併成一張表格

In [17]:
```
tbl = cbind(tbl, time_tbl )[-2]
```

把 date 和 time 合併到原本的 tbl 表格
[-2] 是指原本 tbl 的第二欄 create_time 我不要了

In [18]:
```
head(tbl)
```

Out[18]:

	message	id	date	time
1	#最新	352962731493606_729357620520780	20150726	035653
2	用影片讓人更了解過去	352962731493606_729335113856364	20150726	035000
3	曹錦輝下放3A	352962731493606_729354070521135	20150726	034011
4	好奇拍攝出來會是什麼樣子	352962731493606_729334647189744	20150726	032001
5	該架小型飛機事發時載著5個人，機上兩人以及地面一名居民可能走避不及....	352962731493606_729340873855788	20150726	030536
6	嗯哼~BJ4 #曹錦輝	352962731493606_729317440524798	20150726	025000

In [19]:
```
tbl$id = str_replace_all(tbl$id, "352962731493606_", "")
```

順便把 id 多餘部分拿掉

In [20]:
```
head(tbl)
```

Out[20]:

	message	id	date	time
1	#最新	729357620520780	20150726	035653
2	用影片讓人更了解過去	729335113856364	20150726	035000
3	曹錦輝下放3A	729354070521135	20150726	034011
4	好奇拍攝出來會是什麼樣子	729334647189744	20150726	032001
5	該架小型飛機事發時載著5個人，機上兩人以及地面一名居民可能走避不及....	729340873855788	20150726	030536
6	嗯哼~BJ4 #曹錦輝	729317440524798	20150726	025000

蘋果新聞半夜也po文
好認真喔 + _ +

資料庫操作

```
In [21]: library(RSQLite)
         library(dplyr)
```

```
Loading required package: DBI
```

```
In [22]: setwd('/tmp//RCrawler')
```

```
In [23]: getwd()
```

```
Out[23]: '/tmp/RCrawler'
```

在 /tmp/RCrawler 目錄下，創造一個資料庫檔案 my_fb.db

```
In [24]: my_fb <- src_sqlite("my_fb.db", create = TRUE)
```

```
In [27]: dir("/tmp/RCrawler")
```

```
Out[27]: 'my_fb.db'
```

```
In [25]: dim(tbl)
Out[25]:    25  4
```

tbl 裡面目前有 25筆料

```
In [26]: tbl[1:5, ]
```

Out[26]:

	message	id	date	time
1	駕照是怎麼考到的？	729354443854431	20150726	042000
2	黃國昌也要投入立委選戰了	729364717186737	20150726	041859
3	（頂眼鏡）	729360497187159	20150726	040755
4	#最新	729357620520780	20150726	035653
5	用影片讓人更了解過去	729335113856364	20150726	035000

```
In [27]: dbWriteTable(my_fb$con, "my_fb_table", tbl[1:5, ])
Out[27]: TRUE
```

把前五筆資料寫入到資料庫裡面,
表格名稱為 my_fb_table

```
In [28]: #dbRemoveTable(my_fb$con, "my_fb_table")
```

```
In [29]: tbl(my_fb, sql("SELECT * FROM my_fb_table"))
```

查詢看看有沒有資料

```
Out[29]: Source: sqlite 3.8.6 [my_fb.db]
From: <derived table> [?? x 4]

                    message              id     date   time
1          駕照是怎麼考到的？ 729354443854431 20150726 042000
2  黃國昌也要投入立委選戰了 729364717186737 20150726 041859
3                （頂眼鏡） 729360497187159 20150726 040755
4                   #最新 729357620520780 20150726 035653
5        用影片讓人更了解過去 729335113856364 20150726 035000
..                     ...             ...      ...    ...
```

```
In [30]: dbWriteTable(my_fb$con, "my_fb_table", tbl[6:10, ], append=TRUE)
Out[30]: TRUE
```

再把第六到第十筆資料寫入到資料庫裡面

```
In [32]: tbl[6:10, ]
```

Out[32]:

	message	id	date	time
6	曹錦輝下放3A	729354070521135	20150726	034011
7	好奇拍攝出來會是什麼樣子	729334647189744	20150726	032001
8	該架小型飛機事發時載著5個人，機上兩人以及地面一名居民可能走避不及....	729340873855788	20150726	030536
9	嗯哼~BJ4 #曹錦輝	729317440524798	20150726	025000
10	#最新	729331047190104	20150726	024401

```
In [33]: tbl(my_fb, sql("SELECT * FROM my_fb_table WHERE rowid >= 6"))
Out[33]: Source: sqlite 3.8.6 [my_fb.db]
         From: <derived table> [?? x 4]
```

透過rowid 來查詢第六筆以後的資料

```
                                                                  message
1                                                                 曹錦輝下放3A
2                                                               好奇拍攝出來會是什麼樣子
3  該架小型飛機事發時載著5個人，機上兩人以及地面一名居民可能走避不及....
4                                                                 嗯哼~BJ4\n\n#曹錦輝
5                                                                     #最新
..                                                                      ...
Variables not shown: id (chr), date (chr), time (chr)
```

如果我想另外開一個 R 來連結剛才的資料庫
並對資料庫進行關鍵字或時間查詢

In [2]:
```
library(RSQLite)
library(dplyr)
```

In [3]:
```
my_fb <- src_sqlite("/tmp/RCrawler/my_fb.db")
```

In [4]:
```
res = tbl(my_fb, sql("SELECT * FROM my_fb_table WHERE message LIKE '%曹錦輝%'"))
```

In [5]:
```
collect(res)
```

Out[5]:

	message	id	date	time
1	曹錦輝下放3A	729354070521135	20150726	034011
2	嗯哼~BJ4 #曹錦輝	729317440524798	20150726	025000

In [6]:
```
res = tbl(my_fb, sql("SELECT * FROM my_fb_table WHERE time LIKE '02%'"))
```

In [7]:
```
collect(res)
```

Out[7]:

	message	id	date	time
1	嗯哼~BJ4 #曹錦輝	729317440524798	20150726	025000
2	#最新	729331047190104	20150726	024401

練習 5分鐘

自己寫入資料庫
查詢自己有興趣的關鍵字

Data Wrangling with dplyr and tidyr Cheat Sheet

RStudio

Tidy Data - A foundation for wrangling in R

In a tidy data set:

Each **variable** is saved in its own **column**

&

Each **observation** is saved in its own **row**

Tidy data complements R's **vectorized operations**. R will automatically preserve observations as you manipulate variables. No other format works as intuitively with R.

M * A

Syntax - Helpful conventions for wrangling

dplyr::tbl_df(iris)

Converts data to tbl class. tbl's are easier to examine than data frames. R displays only the data that fits onscreen:

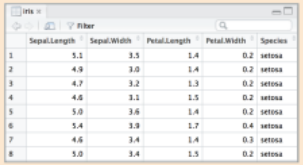
```
Source: local data frame [150 x 5]

   Sepal.Length Sepal.Width Petal.Length
1           5.1         3.5          1.4
2           4.9         3.0          1.4
3           4.7         3.2          1.3
4           4.6         3.1          1.5
5           5.0         3.6          1.4
..          ...         ...          ...
Variables not shown: Petal.Width (dbl),
  Species (fctr)
```

dplyr::glimpse(iris)

Information dense summary of tbl data.

utils::View(iris)

View data set in spreadsheet-like display (note capital V).



dplyr::%>%

Passes object on left hand side as first argument (or . argument) of function on righthand side.

```
x %>% f(y)  is the same as  f(x, y)
y %>% f(x, ., z)  is the same as  f(x, y, z )
```

"Piping" with %>% makes code more readable, e.g.

```
iris %>%
  group_by(Species) %>%
  summarise(avg = mean(Sepal.Width)) %>%
  arrange(avg)
```

Reshaping Data - Change the layout of a data set

tidyr::gather(cases, "year", "n", 2:4)

Gather columns into rows.

tidyr::spread(pollution, size, amount)

Spread rows into columns.

tidyr::separate(storms, date, c("y", "m", "d"))

Separate one column into several.

tidyr::unite(data, col, ..., sep)

Unite several columns into one.

dplyr::data_frame(a = 1:3, b = 4:6)

Combine vectors into data frame (optimized).

dplyr::arrange(mtcars, mpg)

Order rows by values of a column (low to high).

dplyr::arrange(mtcars, desc(mpg))

Order rows by values of a column (high to low).

dplyr::rename(tb, y = year)

Rename the columns of a data frame.

Subset Observations (Rows)

dplyr::filter(iris, Sepal.Length > 7)

Extract rows that meet logical criteria.

dplyr::distinct(iris)

Remove duplicate rows.

dplyr::sample_frac(iris, 0.5, replace = TRUE)

Randomly select fraction of rows.

dplyr::sample_n(iris, 10, replace = TRUE)

Randomly select n rows.

dplyr::slice(iris, 10:15)

Select rows by position.

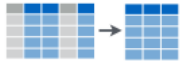
dplyr::top_n(storms, 2, date)

Select and order top n entries (by group if grouped data).

Logic in R - ?Comparison, ?base::Logic

<	Less than	!=	Not equal to
>	Greater than	%in%	Group membership
==	Equal to	is.na	Is NA
<=	Less than or equal to	!is.na	Is not NA
>=	Greater than or equal to	&, |, !, xor, any, all	Boolean operators

Subset Variables (Columns)



dplyr::select(iris, Sepal.Width, Petal.Length, Species)

Select columns by name or helper function.

Helper functions for select - ?select

select(iris, contains("."))
Select columns whose name contains a character string.

select(iris, ends_with("Length"))
Select columns whose name ends with a character string.

select(iris, everything())
Select every column.

select(iris, matches(".t."))
Select columns whose name matches a regular expression.

select(iris, num_range("x", 1:5))
Select columns named x1, x2, x3, x4, x5.

select(iris, one_of(c("Species", "Genus")))
Select columns whose names are in a group of names.

select(iris, starts_with("Sepal"))
Select columns whose name starts with a character string.

select(iris, Sepal.Length:Petal.Width)
Select all columns between Sepal.Length and Petal.Width (inclusive).

select(iris, -Species)
Select all columns except Species.

來源網址

https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf

Summarise Data

dplyr::**summarise(iris, avg = mean(Sepal.Length))**
Summarise data into single row of values.
dplyr::**summarise_each(iris, funs(mean))**
Apply summary function to each column.
dplyr::**count(iris, Species, wt = Sepal.Length)**
Count number of rows with each unique value of variable (with or without weights).

Summarise uses **summary functions**, functions that take a vector of values and return a single value, such as:

dplyr::**first**
First value of a vector.
dplyr::**last**
Last value of a vector.
dplyr::**nth**
Nth value of a vector.
dplyr::**n**
# of values in a vector.
dplyr::**n_distinct**
# of distinct values in a vector.
IQR
IQR of a vector.
min
Minimum value in a vector.
max
Maximum value in a vector.
mean
Mean value of a vector.
median
Median value of a vector.
var
Variance of a vector.
sd
Standard deviation of a vector.

Group Data

dplyr::**group_by(iris, Species)**
Group data into rows with the same value of Species.
dplyr::**ungroup(iris)**
Remove grouping information from data frame.

iris %>% group_by(Species) %>% summarise(...)
Compute separate summary row for each group.

Make New Variables

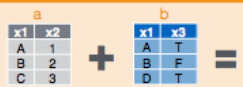
dplyr::**mutate(iris, sepal = Sepal.Length + Sepal. Width)**
Compute and append one or more new columns.
dplyr::**mutate_each(iris, funs(min_rank))**
Apply window function to each column.
dplyr::**transmute(iris, sepal = Sepal.Length + Sepal. Width)**
Compute one or more new columns. Drop original columns.

Mutate uses **window functions**, functions that take a vector of values and return another vector of values, such as:

dplyr::**lead**
Copy with values shifted by 1.
dplyr::**lag**
Copy with values lagged by 1.
dplyr::**dense_rank**
Ranks with no gaps.
dplyr::**min_rank**
Ranks. Ties get min rank.
dplyr::**percent_rank**
Ranks rescaled to [0, 1].
dplyr::**row_number**
Ranks. Ties got to first value.
dplyr::**ntile**
Bin vector into n buckets.
dplyr::**between**
Are values between a and b?
dplyr::**cume_dist**
Cumulative distribution.
dplyr::**cumall**
Cumulative `all`
dplyr::**cumany**
Cumulative `any`
dplyr::**cummean**
Cumulative `mean`
cumsum
Cumulative `sum`
cummax
Cumulative `max`
cummin
Cumulative `min`
cumprod
Cumulative `prod`
pmax
Element-wise `max`
pmin
Element-wise `min`

iris %>% group_by(Species) %>% mutate(...)
Compute new variables by group.

Combine Data Sets



Mutating Joins

dplyr::**left_join(a, b, by = "x1")**
Join matching rows from b to a.

dplyr::**right_join(a, b, by = "x1")**
Join matching rows from a to b.

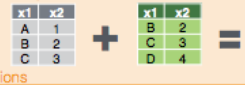
dplyr::**inner_join(a, b, by = "x1")**
Join data. Retain only rows in both sets.

dplyr::**full_join(a, b, by = "x1")**
Join data. Retain all values, all rows.

Filtering Joins

dplyr::**semi_join(a, b, by = "x1")**
All rows in a that have a match in b.

dplyr::**anti_join(a, b, by = "x1")**
All rows in a that do not have a match in b.



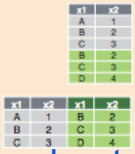
Set Operations

dplyr::**intersect(y, z)**
Rows that appear in both y and z.

dplyr::**union(y, z)**
Rows that appear in either or both y and z.

dplyr::**setdiff(y, z)**
Rows that appear in y but not z.

Binding



dplyr::**bind_rows(y, z)**
Append z to y as new rows.

dplyr::**bind_cols(y, z)**
Append z to y as new columns.
Caution: matches rows by position.

來源網址
https://www.rstudio.com/wp-content/uploads/2015/02/data-wrangling-cheatsheet.pdf

我的爬蟲想自動化 ...

可以每天每小時就去抓一次嗎 (￣▽￣)~*

fb_bot.R

```r
library(httr)
library(dplyr)
library(logging)

basicConfig()
addHandler(writeToFile, file="/tmp/RCrawler/fb_bot.log")

url = "https://graph.facebook.com/v2.4/352962731493606/?access_token=CAACEdEose0cBAHXtl3FUZCC86TRWhaTuOa4BE6dNIOePCPH6PQFcN9jzXiRiUZARjeVvJeqfY5YW9TnTr8M
res = content(GET(url))
tbl = lapply(res$posts$data, data.frame, stringsAsFactors=FALSE)
tbl = rbind_all(tbl)
print(head(tbl))

loginfo("finished %s", "352962731493606")
```

apple — root@localhost: /tmp/RCrawler — ssh — 80×29

root@localhost: /tmp/RCrawler

```
root@localhost:/tmp/RCrawler# Rscript fb_bot.R
```

可以在 command line
利用 Rscript 執行 R 程式

fb_bot.R

```r
library(httr)
library(dplyr)
library(logging)

basicConfig()
addHandler(writeToFile, file="/tmp/RCrawler/fb_bot.log")

url = "https://graph.facebook.com/v2.4/352962731493606/?access_token=CAACEdEose0cBAHXtl3FUZCC86TRWhaTu0a4BE6dNIOePCPH6PQFcN9jzXiRiUZARjeVvJeqfY5YW9TnTr8Mw2
res = content(GET(url))
tbl = lapply(res$posts$data, data.frame, stringsAsFactors=FALSE)
tbl = rbind_all(tbl)
print(head(tbl))

loginfo("finished %s", "352962731493606")
```

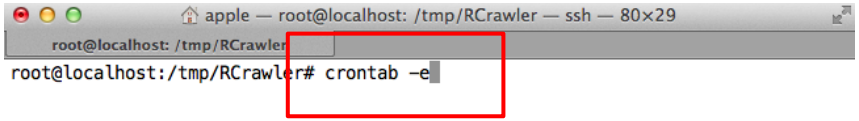
apple — root@localhost: /tmp/RCrawler — ssh — 80×29

root@localhost: /tmp/RCrawler

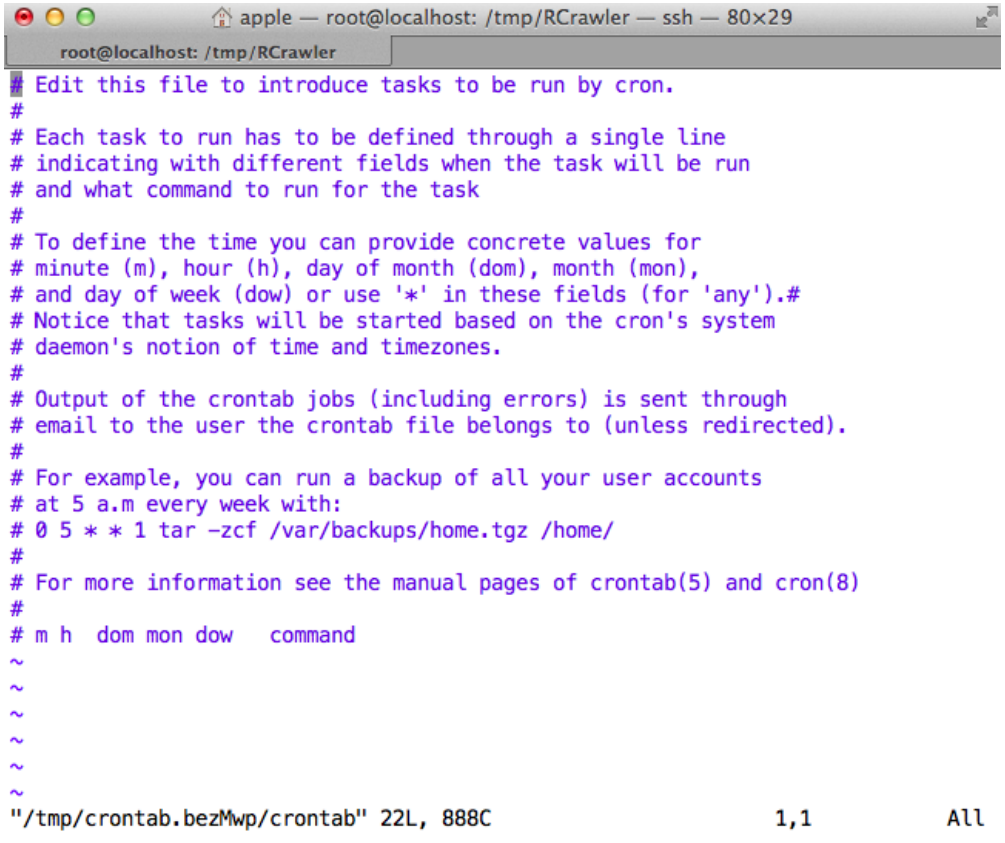
```
The following objects are masked from ‘package:stats’:

    filter, lag

The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union

Loading required package: methods
Source: local data frame [6 x 3]

                                                                message
1
                                                         小編...小編還沒老...
2 一名30歲女子與兒子一起乘扶手梯上樓，快抵達樓層時腳下踏板突然鬆脫，眼看就要被捲
進手扶梯時，母親在最後一刻舉起兒子，自己卻不幸被捲入手扶梯下方的機械間…
3
                                         小知識，希望在意外發生時可以幫上大家
4
                                                                   賀！
5
                                                     末段感言相當值得回味...
6
                                                         請看完內文再下評論
Variables not shown: created_time (chr), id (chr)
2015-07-26 13:09:58 INFO::finished 352962731493606
root@localhost:/tmp/RCrawler#
```

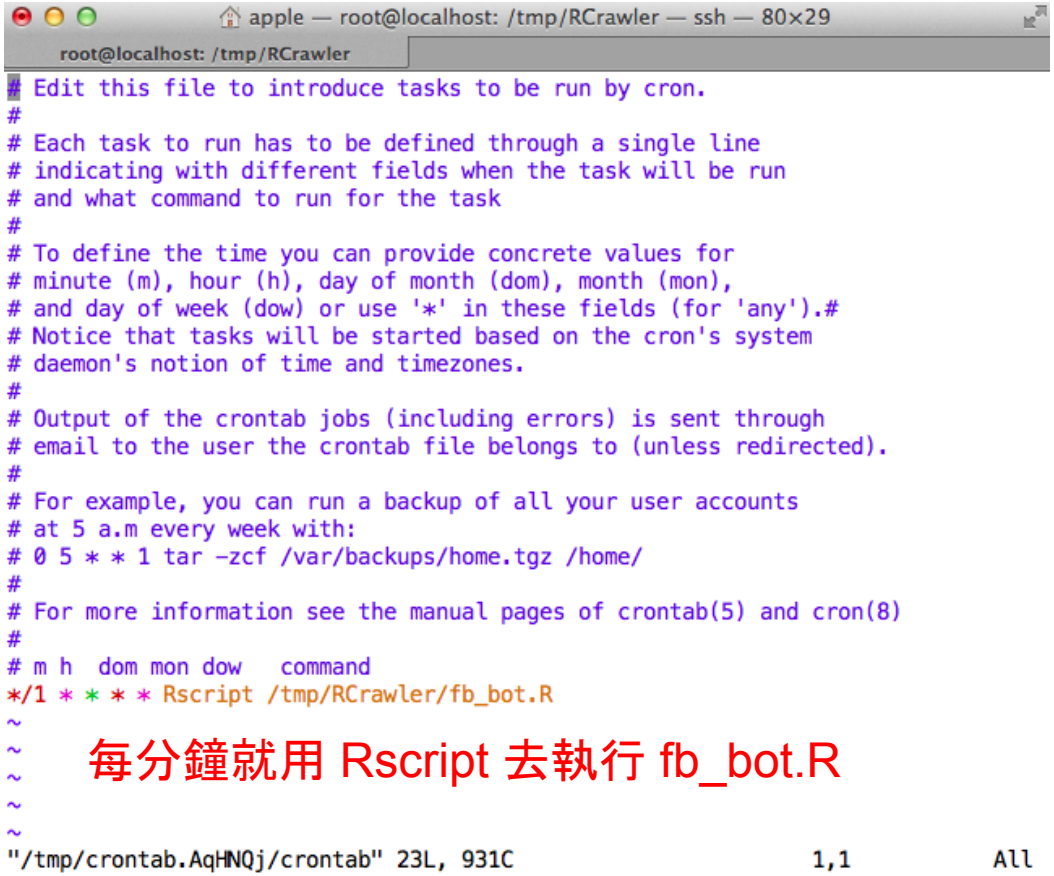
執行結果

接著來設定一下爬蟲的自動化排程



按下 Enter 之後會出現





http://www.puritys.me/docs-blog/article-20-cron-jobs-crontab-%E6%8E%92%E7%A8%8B%E6%95%99%E5%AD%B8.html

fb_bot.R

```r
library(httr)
library(dplyr)
library(logging)

basicConfig()
addHandler(writeToFile, file="/tmp/RCrawler/fb_bot.log")

url = "https://graph.facebook.com/v2.4/352962731493606/?access_token=CAACEdEose0cBAHXtl3FUZCC86TRWhaTuOa4BE6dNIOePCPH6PQFcN9jzXiRiUZARjeVvJeqfY5YW9TnTr8M
res = content(GET(url))
tbl = lapply(res$posts$data, data.frame, stringsAsFactors=FALSE)
tbl = rbind_all(tbl)
print(head(tbl))

loginfo("finished %s", "352962731493606")
```

apple — root@localhost: /tmp/RCrawler — ssh — 80×29

root@localhost: /tmp/RCrawler

```
root@localhost:/tmp/RCrawler# cat fb_bot.log
2015-07-26 13:09:02 INFO::finished 352962731493606
2015-07-26 13:09:47 INFO::finished 352962731493606
2015-07-26 13:09:58 INFO::finished 352962731493606
2015-07-26 13:10:03 INFO::finished 352962731493606
2015-07-26 13:11:02 INFO::finished 352962731493606
2015-07-26 13:12:02 INFO::finished 352962731493606
2015-07-26 13:13:02 INFO::finished 352962731493606
2015-07-26 13:14:02 INFO::finished 352962731493606
2015-07-26 13:15:02 INFO::finished 352962731493606
root@localhost:/tmp/RCrawler#
```

每分鐘就會產生一筆 log 記錄

練習 5分鐘

觀察看看有沒有 log 自動產生

可以不只抓蘋果新聞嗎？ 我還想抓其他粉絲團 ...

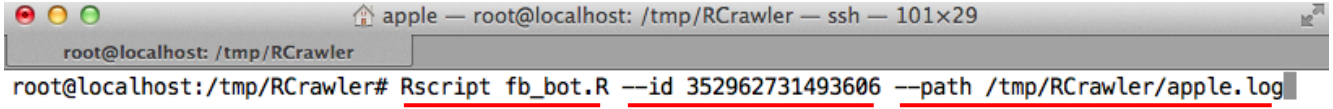
R 語言可以吃參數，利用參數可以達到靈活性

ex. 直接在 command line 執行

Rscript fb_bot.R --id xxxxx --path xxxxx

fb_bot.R

```r
library(httr)
library(dplyr)
library(logging)
library(optparse)


option_list = list(

  make_option(c("-i", "--id"), action="store", default="me", type='character',
              help="please enter the facebook id"),

  make_option(c("-p", "--path"), action="store", default="/tmp/RCrawler/fb_bot.log", type='character',
              help="please enter the log path")
)

opt = parse_args(OptionParser(option_list=option_list))

basicConfig()
addHandler(writeToFile, file=opt$path)

TOKEN = "CAACEdEose0cBAHXtl3FUZCC86TRWhaTuOa4BE6dNIOePCPH6PQFcN9jzXiRiUZARjeVvJeqfY5YW9TnTr8Mw2ZBy4Q1TONK57MoCD6tLiXMKzquv89D2wlNeRg8xssBwKWO5OCTEQwJ1bP22A

url = paste0("https://graph.facebook.com/v2.4/", opt$id, "/?access_token=", TOKEN, "&debug=all&fields=posts&format=json&method=get&pretty=0&suppress_http_c
res = content(GET(url))
tbl = lapply(res$posts$data, data.frame, stringsAsFactors=FALSE)
tbl = rbind_all(tbl)
print(head(tbl))

loginfo("finished %s", opt$id)
```



這樣就能帶參數執行 R 程式
(其實是我平常跑實驗很愛用 XD)

記得順便到 crontab -e 裡面去做修改 :)

那如果程式跑到一半，網路不穩或連線異常斷掉
我可以讓程式自動重跑嗎 ???

ex. 發現它斷線了，自動重新連線，最多重連3次

fb_bot.R

```r
library(httr)
library(dplyr)
library(logging)
library(optparse)


option_list = list(

  make_option(c("-i", "--id"), action="store", default="me", type='character',
              help="please enter the facebook id"),

  make_option(c("-p", "--path"), action="store", default="/tmp/RCrawler/fb_bot.log", type='character',
              help="please enter the log path")
)

opt = parse_args(OptionParser(option_list=option_list))

basicConfig()
addHandler(writeToFile, file=opt$path)

TOKEN = "CAACEdEose0cBAHXtl3FUZCC86TRWhaTuOa4BE6dNIOePCPH6PQFcN9jzXiRiUZARjeVvJeqfY5YW9TnTr8Mw2ZBy4Q1TONK57MoCD6tLiXMKzquv89D2wlNeRg8xssBwKWO5OCTEQwJ1bP2

url = paste0("https://graph.facebook.com/v2.4/", opt$id, "/?access_token=", TOKEN, "&debug=all&fields=posts&format=json&method=get&pretty=0&suppress_http

for(i in 1:3){

    tryCatch({
        res = content(GET(url))
        break

    }, error = function(e){
        logerror("url :%s", url)
        logerror(e)
        Sys.sleep(5)
    })
}

tbl = lapply(res$posts$data, data.frame, stringsAsFactors=FALSE)
tbl = rbind_all(tbl)
print(head(tbl))

loginfo("finished %s", opt$id)
```