

心理與神經資訊學

(Psychoinformatics & Neuroinformatics)

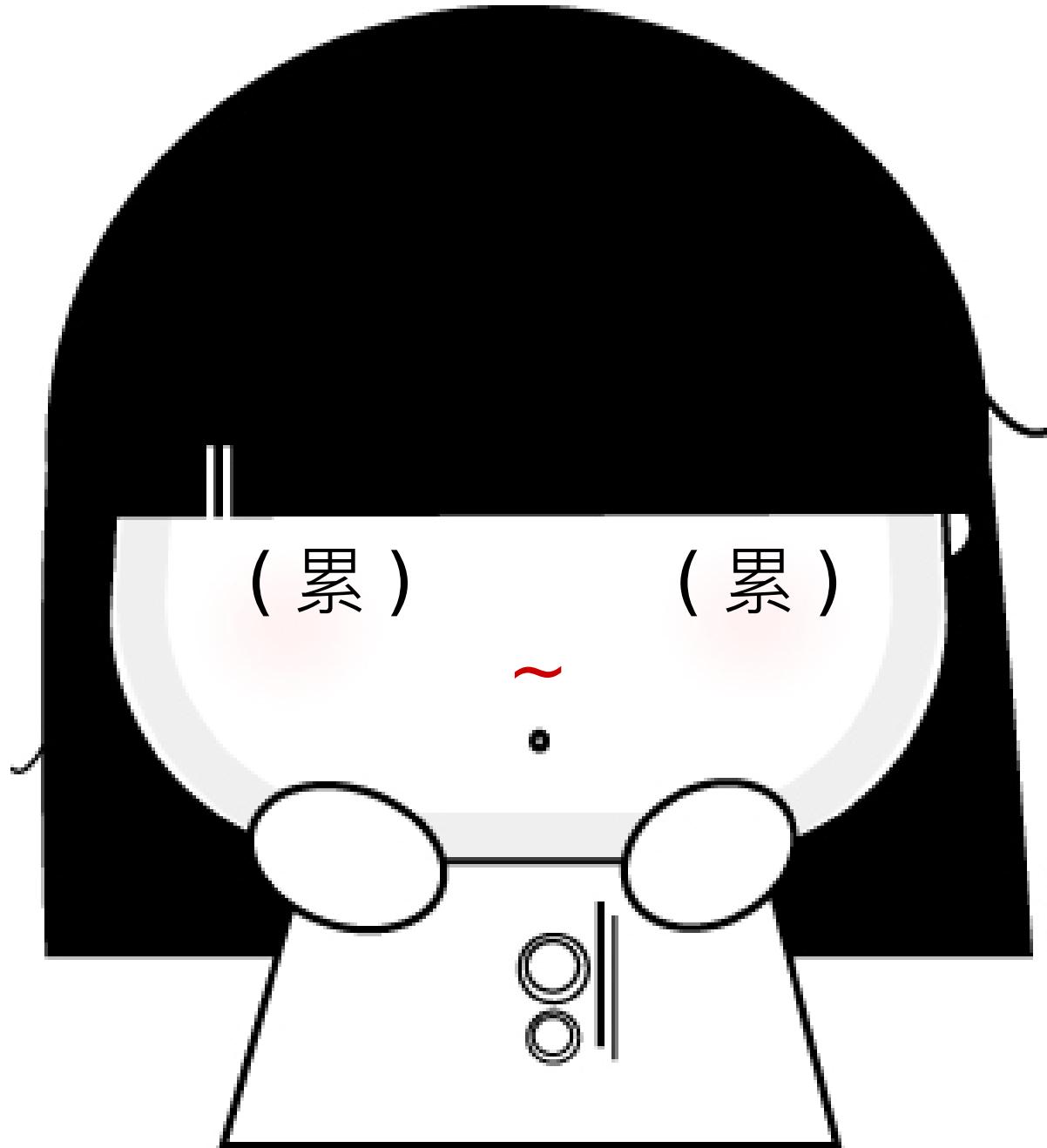
課號：Psy1007

識別碼：20710100

教室：博雅 101

時間：四 234





網路資料的搜集 (Web APIs)

今天需要安裝的模組

請在 Jupyter Notebook 下輸入：

!pip install dcard-spider facebook-sdk tweepy
(以上分別是處理 Dcard, FB, & Twitter 需要的模組)

!pip install oauth2 google-api-python-client
(以上分別是處理 Dcard, FB, & Twitter 的模組)

!pip install --upgrade oauth2client
(以上是進一步更新處理 Google 服務的模組)



What are API and SDK?

API=Application Programming Interface



若把 API 看成函數，則類似 Python Module 的 SDK (Software Development Kit) 則是一群 APIs 的集合

幻想文天堂 Dcard

<https://www.dcard.tw/f/funny/p/226048661>

 搜尋有趣板、@人物 Q 註冊 登入 ▾

全部

▼ 分類看板

- 汽機車
- 語言
- 感情
- 女孩
- 美妝
- 穿搭
- 有趣

▶ 校園看板

套房隔音太差，常常聽到鄰居的...

3月25日 12:32

我租的套房雖然是水泥隔間
但因為屋頂是輕鋼架所以隔音還是不太好
沒聽過隔壁貓叫春
只有一個不知道是隔壁還是樓下的女人很愛半夜唱歌

某天被她吵得不行
一怒之下把wifi熱點的名字改成

別他媽半夜唱歌啊！

• 逐條顯示 4G 03:35 100%  名稱

別他媽半夜唱歌啊！

 29284  228 



神秘的 Dcard APIs

端看看 https://www.dcard.tw/_api/posts/226048661

看板資訊	<code>/_api/forums</code>
文章資訊	<code>/_api/forums/</code> 看板名稱 <code>/posts</code> (照時間排序 : <code>popular=false</code>) (照發文前後 : <code>before/after=</code> 文章編號)
文章內文	<code>/_api/posts/</code> 文章編號 (照時間排序 : <code>popular=false</code>)
文章內連結	<code>/_api/posts/</code> 文章編號 <code>/links</code>
文章內留言	<code>/_api/posts/</code> 文章編號 <code>/comments</code> (照留言前後 : <code>before/after=</code> 樓層編號)

用 Python 呼叫 Dcard APIs



Web APIs 大部分回傳的是 json 格式

```
import urllib.request,json
u='https://www.dcard.tw/_api/posts/226048661'
r=urllib.request.Request(u,headers={'User-Agent':''})
data=urllib.request.urlopen(r).read()
j_data=json.loads(data.decode('utf-8'))
print(j_data['media'][0]['url'])
for key in j_data.keys():
    print(key,':',j_data[key])
```

Dict → json.dumps → String
Dict ← json.loads ← String

Dcard 有非官方的 Python 模組

更多使用細節請參考[此文](#)

```
from dcard import Dcard
def filter(metas):
    return [m for m in metas if m['likeCount'] >= 100]
#return [m for m in metas if "女友" in m['topics']]

d=Dcard()
f=d.forums('photography') # 攝影版
m=f.get_metas(num=50,callback=filter) #list
p=d.posts(m).get(comments=False)
r=p.parse_resources() #list: try r[0][1]
done,fails=p.download(r)
print('Got %d pics' % done if len(fails)==0 else 'Error!')
```

心理學案例研究：臉書上的偏好行為

由按讚之頁面可推測個人資訊

1

Users' Facebook Likes

55,814 Likes

	art	cnn.com	(...)	BMW
User 1	1	1	...	0
User 2	0	1	...	1
User 3	1	0	...	0
(...)
User n	1	1	...	0

User – Like Matrix
(10M User-Like pairs)

58,466 Users

2

Singular Value Decomposition

3

Prediction Model

Using Logistic or Linear Regression
(with 10-fold cross validation)

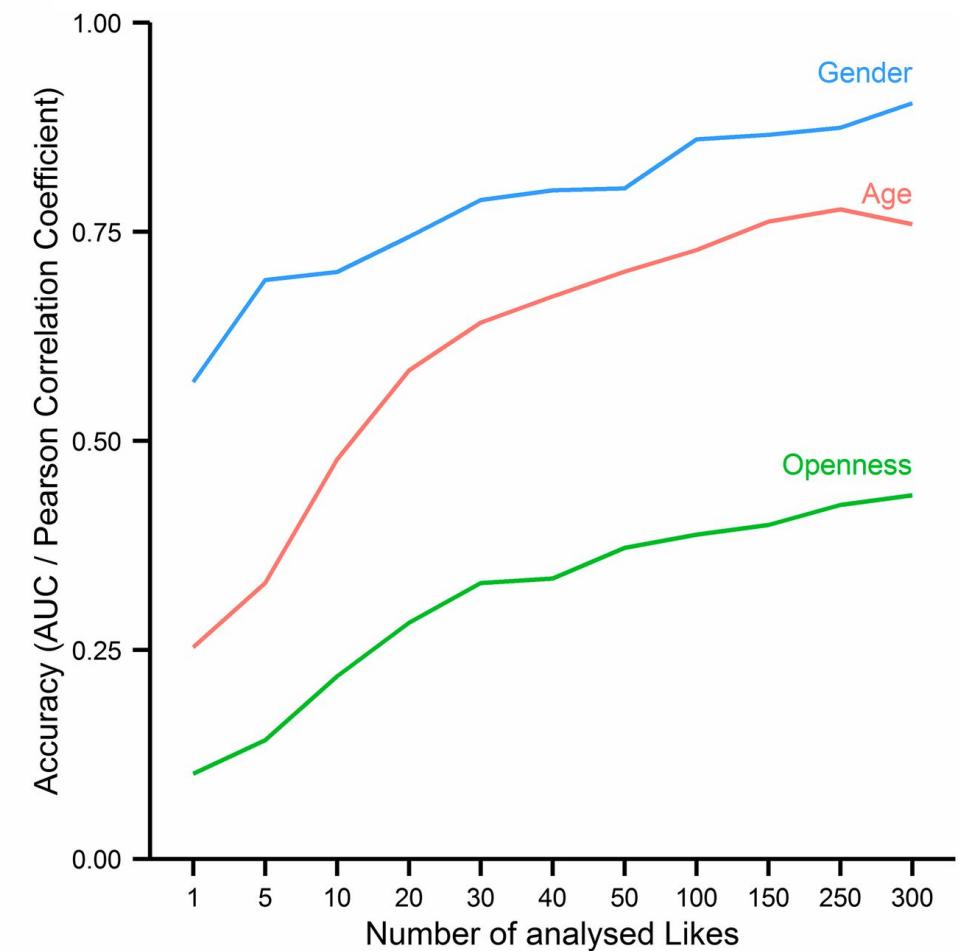
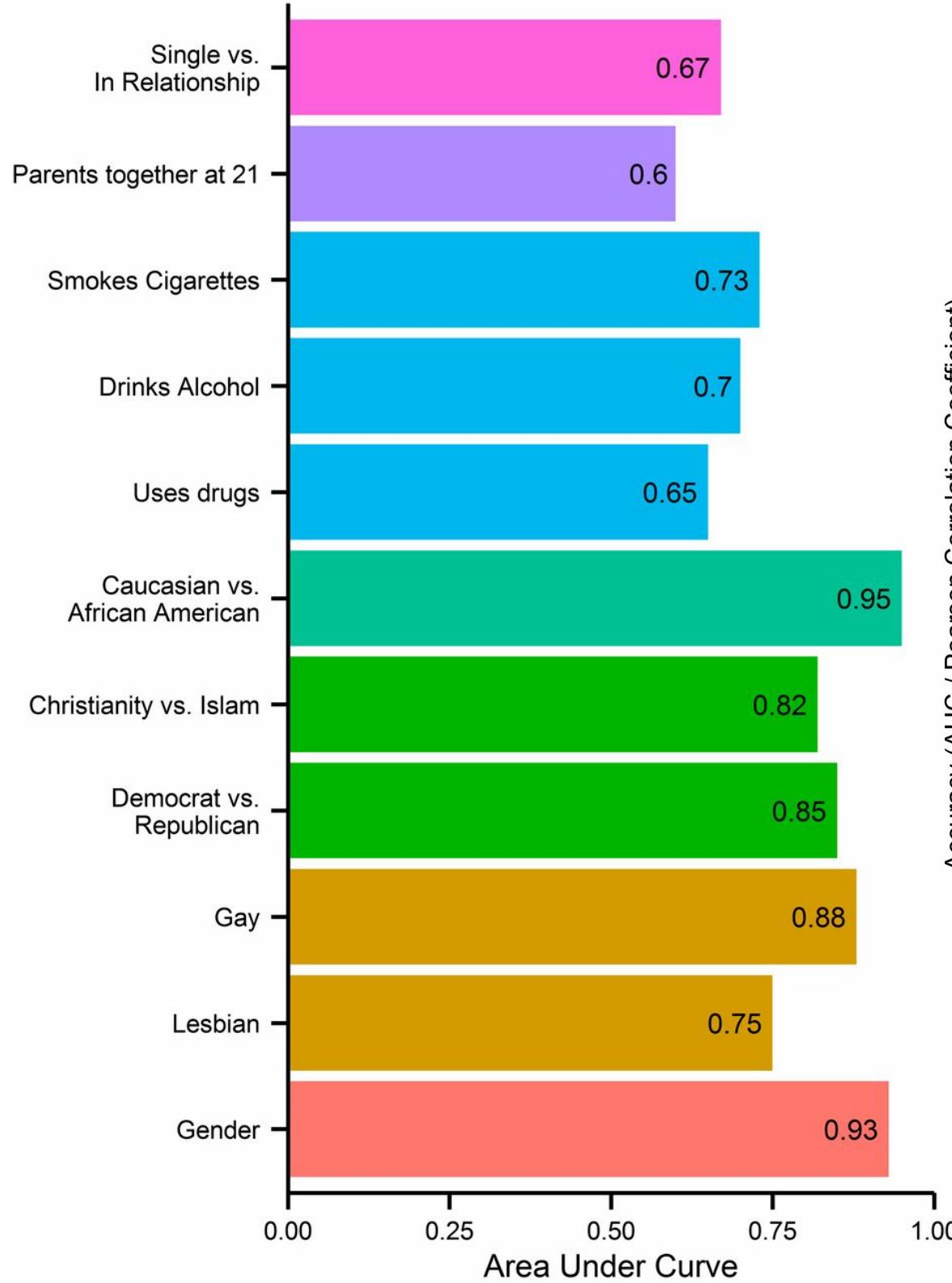
$$\text{e.g. } \text{age} = \alpha + \beta_1 C_1 + \dots + \beta_n C_{100}$$

Predicted variables

Facebook profile: age, gender, political and religious views, relationship status, proxy for sexual orientation, social network size and density

Profile picture: ethnicity

Survey / test results: BIG5 Personality, intelligence, satisfaction with life, substance use, parents together?



Kosinski et al., 2013, PNAS



Facebook APIs

Facebook APIs 的詳細介紹可以看這裡

原頁面：<http://www.facebook.com/taiwanpsy>

純資料：<http://graph.facebook.com/taiwanpsy>

Explorer可幫助我們找出還有何資料可看：

圖連接：[https://graph.facebook.com/taiwanpsy?fields=albums.fields\(link\)&access_token=...](https://graph.facebook.com/taiwanpsy?fields=albums.fields(link)&access_token=...)

文敘述：[https://graph.facebook.com/taiwanpsy?fields=posts.limit\(5\).fields\(description\)&access_token=...](https://graph.facebook.com/taiwanpsy?fields=posts.limit(5).fields(description)&access_token=...)

...



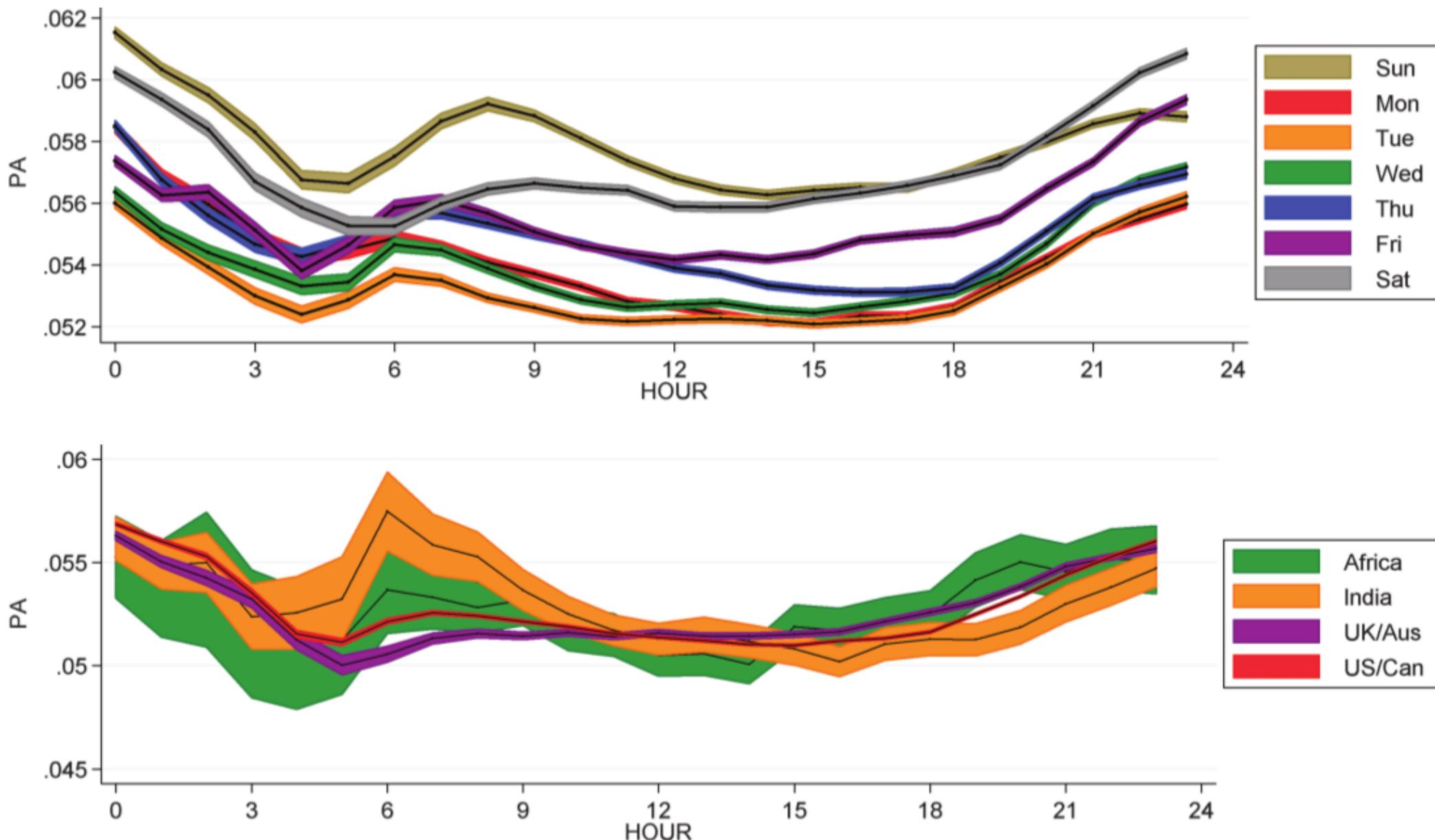
Python for Facebook

Facebook 的 Python SDK 詳細介紹可以看這裡

```
import json
def show(j_in):
    print(json.dumps(j_in,ensure_ascii=False,indent=2))

import facebook
token='...' # replaced by your access token
graph=facebook.GraphAPI(token)
show(graph.get_object('me',fields='id,name,email'))
show(graph.get_connections('me','friends'))
#graph.put_object('me','feed',message='Hello from class!')
```

心理學案例研究：Tweets 情緒分析



Golder & Macy, 2011, Science

心理學案例研究：政黨與快樂

共和黨支持者自陳較為快樂卻較多負面情緒

Positive emotion word lists

	PANAS-X Positive affect	PANAS-X Joviality	LIWC Positive emotion	Happy emoticons
Conservatives				
Liberals	0.88 < 1	0.80 < 1	0.95 < 1	0.83 < 1

Negative emotion word lists

	PANAS-X Negative affect	PANAS-X Sadness	LIWC Negative emotion	Sad emoticons
Conservatives				
Liberals	1.44 > 1	1.32 > 1	1.07 > 1	0.77 < 1



Twitter API 1.1

Twitter APIs 的詳細介紹可以看這裡

踹看看：

<https://twitter.com/search?q=%E5%8A%A0%E5%A4%A7>

<https://api.twitter.com/1.1/search/tweets.json?q=%E5%8A%A0%E5%A4%A7>

Twitter APIs 幾乎都要通過 OAuth 身份認證才能用

Application Settings

Keep the "Consumer Secret" a secret. This key should never be human-readable in your application.

Consumer Key (API Key) HDTBqZDAOuZJyDhsr5Z5tw

Consumer Secret (API Secret) nKULI80bwwedY4jtk4UTt2dXotoIVOIPQ8fUVK2LDo

Your access token

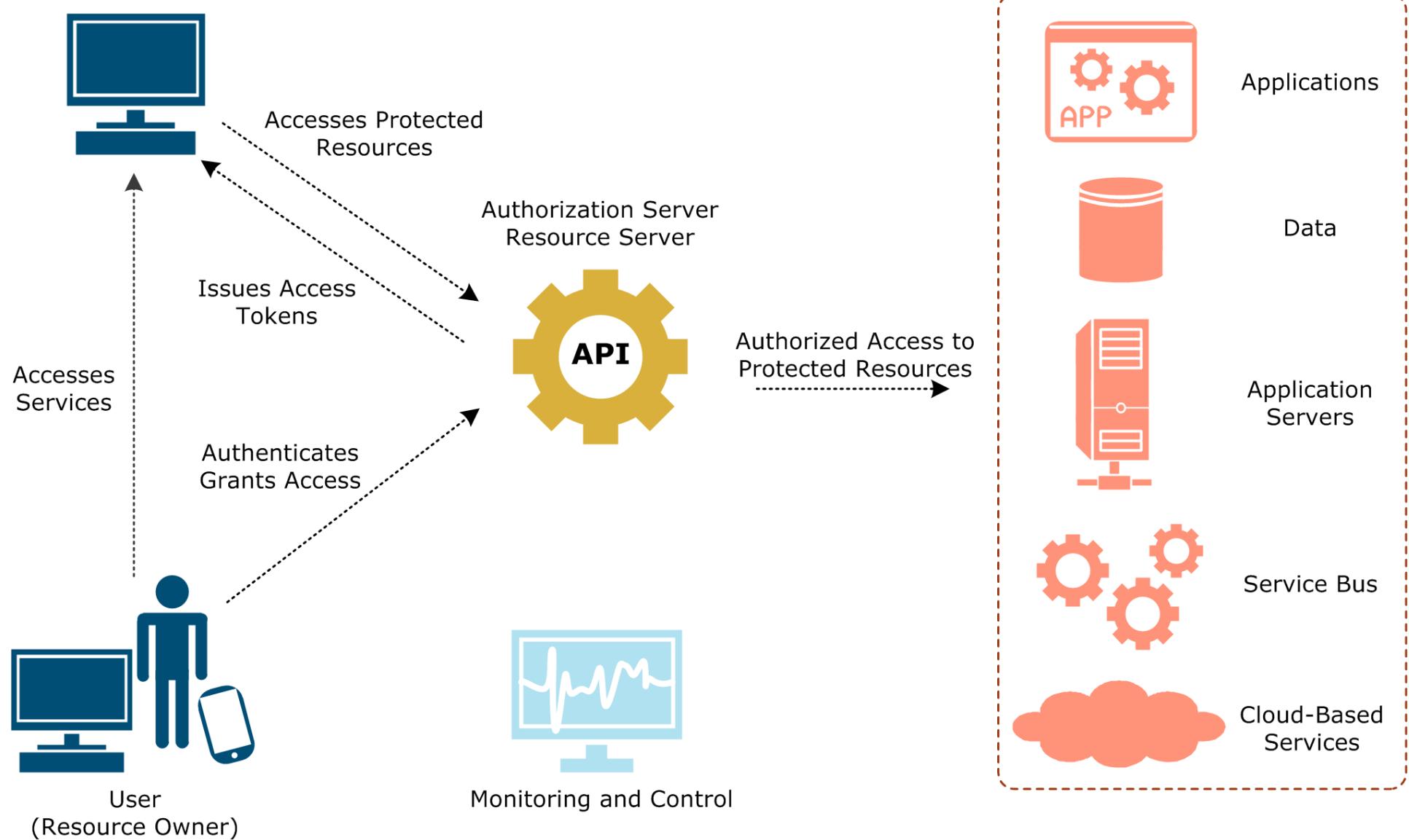
This access token can be used to make API requests on your own account's behalf. Do not share your access token secret with anyone.

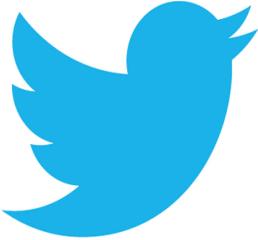
Access token 206228745-LujHvkviaSuBq4FH2Unx92xvYFCBYtoArQknciLF

Access token secret AbuFKDs8oWSFMcbwUOq9D4E0Blx4dh8tfzvQ93Utndvjp

OAuth 2.0

確保 A 能不知我們在 B 的密碼卻能存取 B 的部分資訊





Python for Twitter

有很多非官方的 Python 模組可以用
以下按照熱門程度高至低排列

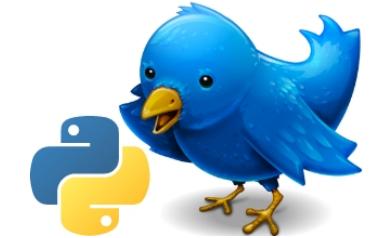
<https://github.com/tweepy/tweepy>

<https://github.com/bear/python-twitter>

<https://github.com/sixohsix/twitter>

更多 Tweepy 的文件可以看這裡

Tweepy , 你好 !



```
import tweepy
key1='...' # API key
secret1='...' # API secret
key2='...' # Access token
secret2='...' # Token secret
auth=tweepy.OAuthHandler(key1,secret1)
auth.set_access_token(key2,secret2)
api=tweepy.API(auth)
data=api.search(q=u'台灣',
                 geocode='22.9500,120.2000,500km')
for tweet in data:
    print(tweet.user.screen_name,tweet.text,'\n')
#for follower in api.followers_ids('iingwen'):
#    user=api.get_user(follower)
#    print(user.screen_name,'\t',user.name)
```

線上鬼城 Google+

[http://plus.google.com/+ 蘋果日報](http://plus.google.com/+蘋果日報)

← 蘋果日報

ABOUT



Home

Collections

Communities

Profile

People

Events

Notifications



蘋果日報

想看新聞？來台灣蘋果日報G+就對了！

159,924 followers

FOLLOW



Google+ APIs

API Calls

Most of the Google+ API follows a RESTful API design, meaning that you use standard HTTP methods to retrieve and manipulate resources. For example, to get the profile of a user, you might send an HTTP request like:

```
GET https://www.googleapis.com/plus/v1/people/userId
```

<https://www.googleapis.com/plus/v1/people/> + 蘋果日報

```
{
  "error": {
    "errors": [
      {
        "domain": "usageLimits",
        "reason": "dailyLimitExceededUnreg",
        .....
    }
}
```



沒授權不行？

Google 有各種服務的 APIs

使用前要先創立 Projects

The screenshot shows the Google Cloud Platform API Manager interface. On the left, there's a sidebar with icons for IAM & Admin, All projects (which is selected and highlighted in blue), IAM, Quotas, Service accounts, Labels, Settings, Encryption Keys, and Identity-Aware Proxy. The main area has a header with 'Google APIs Project' and a search bar. Below the header are buttons for 'CREATE PROJECT' (which is highlighted with a red box) and 'DELETE PROJECT'. A 'HIDE INFO' button is also present. The central part of the screen displays a 'Select a project' section with tabs for 'PERMISSIONS' (which is active) and 'LABELS'. A message at the bottom of this section says 'Please select at least one resource.'

API Keys

使用 APIs 前要先取得授權碼 (API key)

The screenshot shows the Google API Manager interface. The top navigation bar includes the Google logo, the project name "Informatics" (with a dropdown arrow), a search bar, and various icons for settings and notifications. The main menu on the left has sections for "API Manager" (selected), "Dashboard", "Library", and "Credentials" (which is highlighted with a red box). The central area is titled "Credentials" and contains three tabs: "Credentials" (selected), "OAuth consent screen", and "Domain verification". A large callout box highlights the "API key" section, which is described as "Identifies your project using a simple API key to check quota and access". Below it are descriptions for "OAuth client ID" and "Service account key". At the bottom of the callout box is a blue button labeled "Create credentials".

Google APIs Informatics

API Manager Credentials

Credentials OAuth consent screen Domain verification

API key
Identifies your project using a simple API key to check quota and access

OAuth client ID
Requests user consent so your app can access the user's data

Service account key
Enables server-to-server, app-level authentication using robot accounts

Create credentials

Google+ APIs (1/2)

還要記得 enable 對應的服務

The screenshot shows the Google APIs Library page. The left sidebar has three main items: Dashboard, Library, and Credentials. The 'Library' item is highlighted with a red box. The main content area is divided into sections: Social APIs, YouTube APIs, and Advertising APIs. The 'Social APIs' section contains the Google+ API, which is also highlighted with a red box. Other listed APIs in this section include Blogger API, Google+ Pages API, and Google+ Domains API. The 'YouTube APIs' section includes YouTube Data API, YouTube Analytics API, and YouTube Reporting API. The 'Advertising APIs' section includes AdSense Management API and DCM/DEA Reporting And Trafficking API.

≡ Google APIs Informatics ▾

API Manager Library

Dashboard

Library

Credentials

Social APIs

Google+ API

Blogger API

Google+ Pages API

Google+ Domains API

YouTube APIs

YouTube Data API

YouTube Analytics API

YouTube Reporting API

Advertising APIs

AdSense Management API

DCM/DEA Reporting And Trafficking API

Google+ APIs (2/2)

網址補上 API key 後即可通行

[https://www.googleapis.com/plus/v1/people/+ 蘋果日報
?key=AIzaSyD2D_DhejojYLHH1YY0qwcsayzVheC-QM4](https://www.googleapis.com/plus/v1/people/+ 蘋果日報?key=AIzaSyD2D_DhejojYLHH1YY0qwcsayzVheC-QM4)

```
{  
  "kind": "plus#person"  
  "objectType": "page",  
  "id": "110444452632515045074",  
  "displayName": "蘋果日報",  
  "tagline": "想看新聞？來台灣蘋果日報 G+ 就對了！",  
  ...  
}
```

Google 官方有提供 Python 模組

The screenshot shows the Google Developers website. The top navigation bar includes the Google Developers logo, a search bar with the placeholder "API Client Library for Pyth...", and a "Search" button. Below the navigation, a breadcrumb trail shows "Products > API Client Libraries > API Client Library for Python". The main title "API Client Library for Python" is displayed prominently in a large white font against a dark teal background.

使用時需要 OAuth 2.0 的授權碼

The screenshot shows the Google Cloud Platform API Manager Credentials page. On the left, a sidebar menu has "API Manager" selected, with options for "Dashboard", "Library", and "Credentials" (which is highlighted). The main content area is titled "Credentials" and contains tabs for "Credentials", "OAuth consent screen", and "Domain verification". A "Create credentials" button is visible. Below the tabs, there are four options: "API key", "OAuth client ID", "Service account key", and "Help me choose". The "OAuth client ID" option is highlighted with a red border. A tooltip for "OAuth client ID" states: "Identifies your project using a simple API key to check quota and access" and "Requests user consent so your app can access the user's data".

Create credentials → OAuth client ID → Type:Other

Python for Google+ (1/2)

```
import httplib2,json
from apiclient import discovery
from oauth2client import client,file,tools
CS='client_secret.json'
FLOW=client.flow_from_clientsecrets(CS,
scope=[
    'https://www.googleapis.com/auth/plus.login',
    'https://www.googleapis.com/auth/userinfo.email'
]) # See https://developers.google.com/+/api/oauth
store=file.Storage('sample.dat')
crdntl=store.get()
if crdntl is None or crdntl.invalid:
    crdntl=tools.run_flow(FLOW,store)
http=crdntl.authorize(httplib2.Http())
srvc=discovery.build('plus','v1',http)
me=srvc.people().get(userId='me').execute()
```

Python for Google+ (2/2)



測試其他功能：

```
people=srvc.people().search(query=u'台大').execute()  
print(json.dumps(people,ensure_ascii=False,indent=2))  
for person in people['items']:  
    print(person['displayName'],person['url'])
```

[YouTube APIs](#)更值得花時間探索

其他社群媒體的 APIs

請自己花時間探索

Line

Instagram

痞客邦

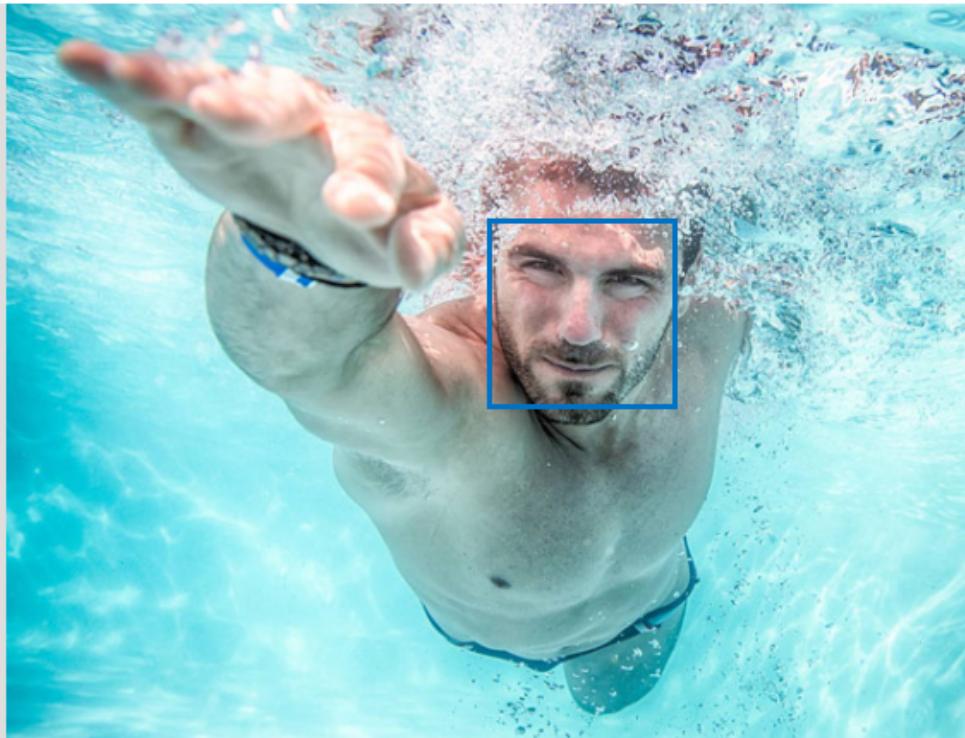
...



若沒提供則用下週方法

Beyond Social Media

有很多其他類型的資料庫或應用程式 APIs



FEATURE	VALUE
NAME:	
Description	{ "tags": ["water", "swimming", "sport", "pool", "person", "man", "frisbee", "ocean", "blue", "bird", "riding", "top", "standing", "wave", "young", "body", "large", "game", "glass", "pond", "playing", "board", "catch", "clear", "boat", "white"], "captions": [{ "text": "a man swimming in a pool of water", "confidence": 0.8909298 }] }
Tags	[{ "name": "water", "confidence": 0.9997857 }, { "name": "swimming", "confidence": 0.955619633 }, { "name": "sport", "confidence": 0.953807831 }, { "name": "pool", "confidence": 0.9515978 }, { "name": "person", "confidence": 0.889862537 }, { "name": "water sport", "confidence": 0.664259 }]
Image format	"Jpeg"

例如 Microsoft 的 Computer Vision APIs

本週作業

進一步研究 Social Media & Web APIs

使用 YouTube APIs研究熱門影片的特徵
(如片長、主題、影片內容等)

(可以是 data- or hypothesis-driven research)

精簡的分析報告中至少要兩張統計圖或表來支持論述
(一張 3 分；不算影片範例圖)

本週請繳交 .ipynb 檔

Game Over

