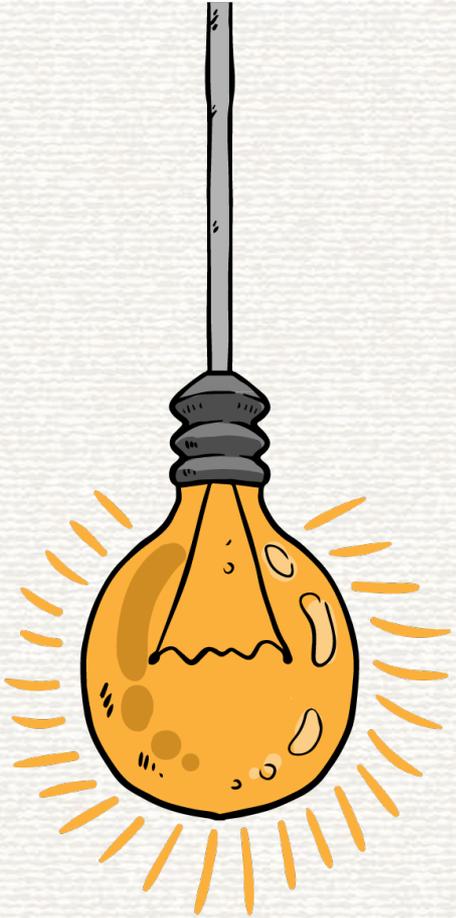




Linebot+爬蟲-製作心得

劉倩好



目录

1. 軟體使用

2. Linebot結合爬蟲

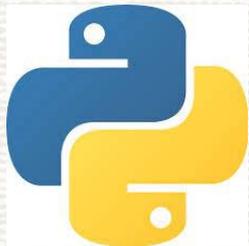
3. 爬蟲程式問題

4. 匯入Linebot問題

A large, horizontal, orange brushstroke graphic that serves as a background for the text. The stroke is composed of many overlapping, slightly curved lines, giving it a textured, hand-drawn appearance. It is centered on the page and spans most of its width.

1.軟體使用

軟體使用



Python+Visual Studio Code

程式撰寫、聊天機器人與
Line developer連結

Git+Heroku

爬蟲與聊天機器人串接

Line Developers

效果呈現

A large, horizontal, orange brushstroke graphic that serves as a background for the text. The stroke is composed of many overlapping, slightly curved lines, giving it a textured, hand-drawn appearance. It is centered on the page and spans most of its width.

2. Linebot結合爬蟲

Linebot資源下載

```
Microsoft Windows [版本 10.0.17134.590]
(c) 2018 Microsoft Corporation. 著作權所有，並保留一切權利。

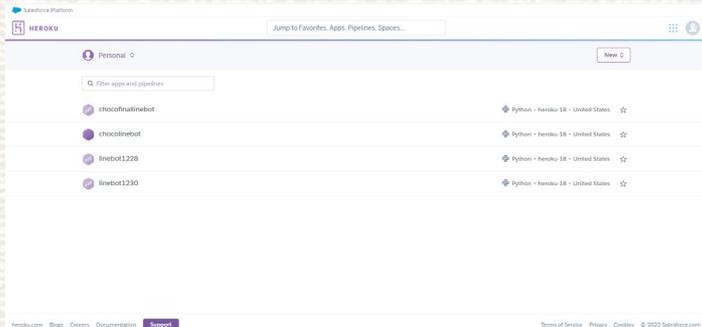
C:\Users\balim>cd C:/

C:\>git clone https://github.com/maso0310/linebot.git
Cloning into 'linebot' ...
remote: Enumerating objects: 5, done.
remote: Total 5 (delta 0), reused 0 (delta 0), pack-reused 5
Unpacking objects: 100% (5/5), done.

C:\>
```

打開 Visual Studio Code，開啟終端機後輸入 `cd C:/`，讓等一下要存取的資料放置在 C 槽中，接著輸入 `git clone https://github.com/maso0310/linebot.git`，將 github 裡 `maso0310/linebit` 的資料下載到自己的 C 槽中

Heroku創建與Linebot上傳



1. heroku login

2. heroku git:remote -a APP Linebot1230

3. git init

4. git add.

5. git commit -am'ok'

6. git push heroku master

更新資料用

```
C:\linebot>heroku login
heroku: Press any key to open up the browser to login or q to exit:
Opening browser to https://cli-auth.heroku.com/auth/cli/browser/cae4d6ff-b0ab-41e3-a97b-ab63bcd756d?requestor=SPMYMTY.gzpbqAAAA4xITEUxjuyjIDxy4
zIM4GAPs13bh2AMIAWGA.iim551Fx1o9MEns1NTG7AF7eBLN21-x9tZrFm_-OyI
Logging in... done
Logged in as adt107127@gm.ntcu.edu.tw

C:\linebot>heroku git:remote -a linebot1230
set git remote heroku to https://git.heroku.com/linebot1230.git

C:\linebot>git init
Reinitialized existing Git repository in C:/linebot/.git/

C:\linebot>git add .

C:\linebot>git commit -am'ok'
On branch master
Your branch is ahead of 'origin/master' by 72 commits.
(use "git push" to publish your local commits)

nothing to commit, working tree clean

C:\linebot>git push heroku master
Everything up-to-date
```

Linebot連結

```
app = Flask(__name__)
static_tmp_path = os.path.join(os.path.dirname(__file__), 'static', 'tmp')
# Channel Access Token
line_bot_api = [REDACTED]

# Channel Secret
handler = WebhookHandler([REDACTED])
```

告訴程式碼需要連結的Linebot

Line_bot_api: Messaging API中

Channel_Secret: Basic settings中

News Products Documentation FAQ Glossary Community Blog

TOP > 人工智慧應用 > 人工智慧

Terms of use URL
optional

App types Bot

Permissions PROFILE

Channel secret

Assertion Signing

Key

News Products Documentation FAQ Glossary Community Blog

TOP > 人工智慧應用 > 人工智慧 > Messaging API

Auto-reply messages Disabled

Greeting messages Enabled

Channel access token

Channel access token (long-lived)

Policies About trademarks Found any problems? Please use our [inquiry form](#)

Family sites English

A large, horizontal, orange brushstroke graphic that serves as a background for the text. The stroke is composed of many overlapping, slightly curved lines, giving it a textured, hand-drawn appearance.

3.爬蟲程式問題

爬蟲程式問題

```
# Google 搜尋
import requests
from bs4 import BeautifulSoup

# Google 搜尋 URL
google_url = 'https://www.google.com.tw/search'
dd = "url"

# 查詢參數
my_params = {'q': '寒流'}

# 下載 Google 搜尋結果
r = requests.get(google_url, params = my_params)
# 確認是否下載成功
if r.status_code == requests.codes.ok:
    # 以 BeautifulSoup 解析 HTML 原始碼
    soup = BeautifulSoup(r.text, 'html.parser')

    # 解析 HTML 原始碼
    #print(soup.prettify())

    # 以 CSS 的選擇器來抓取 Google 的搜尋結果
    items = soup.select('div.kcrvY > a[href!=""]')

    for i in items:
        print("標題: " + i.text)
        # 網址
        print("網址: " + i.get('href')[7:9999])
```

```
標題: 「非常寒冷」寒流襲台魯澳3天連兩天暴冷 | 生活新聞 | 生活 | 聯合新聞網udn.com > 生活
網址: https://udn.com/news/story/7266/4348864&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQFn0ECA5Q8A&usg=AOvVaw0cDkuBBUhmuna0z5VvM0R
標題: 天氣即時預報- 本次寒流最冷期間, 即將到來--... | Facebookwww.facebook.com > weather.taiwan > posts
網址: https://www.facebook.com/weather.taiwan/posts/161499152856599&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQFjADegQIBhAB&usg=AOvVaw3d132uTpfj6ubB60aL97m
標題: 寒流web2.nmns.edu.tw > PubLib > Newsletter
網址: http://web2.nmns.edu.tw/PubLib/Newsletter/90/160/16.htm&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQFjAIegQICBAB&usg=AOvVaw1p9ZzSrsq-FZbDuhNpA56e
標題: 寒潮- 維基百科, 自由的百科全书zh.wikipedia.org > zh-tw > 寒潮
網址: https://zh.wikipedia.org/zh-tw/%25E5%25AF%2592%25E6%25BD%25AE&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQFjANegQIBxAB&usg=AOvVaw09HTyi0L8pFw7ErasY8mSk
標題: 冷氣團寒流知識- 台灣氣象知識網 - Google Sitesites.google.com > site > home > leng-qi-tuan-han-liu-zhi-shi
網址: https://sites.google.com/site/taiwanqixiangzhishiwang/home/leng-qi-tuan-han-liu-zhi-shi&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQFjADegQIChAB&usg=AOvVaw10h-xKomCKRnFkSPM_Fu16
標題: 最強寒流來襲 | 三立新聞網www.setn.com > project
網址: https://www.setn.com/project.aspx%3FProjectID%3D1060&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQtwIwD3oECAQQAg&usg=AOvVaw11e26po2w7PloXc-0Rd2v
標題:
網址: https://www.setn.com/project.aspx%3FProjectID%3D1060&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQtwIwD3oECAQQAg&usg=AOvVaw0c8q4vcXkKH_TArFSEmQe
```

```
標題: 「非常寒冷」寒流襲台魯澳3天連兩天暴冷 | 生活新聞 | 生活 | 聯合新聞網udn.com > 生活
網址: https://udn.com/news/story/7266/4348864&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQFn0ECA5Q8A&usg=AOvVaw0cDkuBBUhmuna0z5VvM0R
標題: 天氣即時預報- 本次寒流最冷期間, 即將到來--... | Facebookwww.facebook.com > weather.taiwan > posts
網址: https://www.facebook.com/weather.taiwan/posts/161499152856599&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQFjADegQIBhAB&usg=AOvVaw3d132uTpfj6ubB60aL97m
標題: 寒流web2.nmns.edu.tw > PubLib > Newsletter
網址: http://web2.nmns.edu.tw/PubLib/Newsletter/90/160/16.htm&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQFjAIegQICBAB&usg=AOvVaw1p9ZzSrsq-FZbDuhNpA56e
標題: 寒潮- 維基百科, 自由的百科全书zh.wikipedia.org > zh-tw > 寒潮
網址: https://zh.wikipedia.org/zh-tw/%25E5%25AF%2592%25E6%25BD%25AE&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQFjANegQIBxAB&usg=AOvVaw09HTyi0L8pFw7ErasY8mSk
標題: 冷氣團寒流知識- 台灣氣象知識網 - Google Sitesites.google.com > site > home > leng-qi-tuan-han-liu-zhi-shi
網址: https://sites.google.com/site/taiwanqixiangzhishiwang/home/leng-qi-tuan-han-liu-zhi-shi&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQFjADegQIChAB&usg=AOvVaw10h-xKomCKRnFkSPM_Fu16
標題: 最強寒流來襲 | 三立新聞網www.setn.com > project
網址: https://www.setn.com/project.aspx%3FProjectID%3D1060&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQtwIwD3oECAQQAg&usg=AOvVaw11e26po2w7PloXc-0Rd2v
標題:
網址: https://www.setn.com/project.aspx%3FProjectID%3D1060&sa=U&ved=2ahUKewjX1tKk68XsAhXjYsBHQ3oAJEQtwIwD3oECAQQAg&usg=AOvVaw0c8q4vcXkKH_TArFSEmQe
```

課堂作業：使用Google搜尋爬出寒流資料

困難：出來的爬蟲結果呈現亂碼無法順利連結

原因：程式碼結束後出現亂碼

爬蟲程式解決

```
import requests
from bs4 import BeautifulSoup

from linebot import (
    LineBotApi, WebhookHandler
)
from linebot.exceptions import (
    InvalidSignatureError
)
from linebot.models import *

def search_crawler(my_params):
    print(my_params)
    google_url = 'https://www.google.com.tw/search'
    information = requests.get(google_url, params = {'q':my_params})

    if information.status_code == requests.codes.ok:

        # 以 BeautifulSoup 解析 HTML 原始碼
        soup = BeautifulSoup(information.text, 'html.parser')

        # 以 CSS 的選擇器來抓取 Google 的搜尋結果
        items = soup.select('div.kCrYT > a[href^="/url"]')
        text = []

        for i in items:
            wantdelect=i.get('href').find('&sa=U&ved=2')
            text.append("標題" + i.text+'\n')
            text.append("網址: " + i.get('href' [7:wantdelect]+''\n'))
            print(i.text)

        string = ''.join(text)
        print(string)
        message = TextSendMessage(text=string)
        return message
```

<https://udn.com/news/story/7266/4348864&sa=U&ved=2ahUKewjX1tKK68XsAhXJyYsBHQ30AJ EQFnoECAsQBA&usg=AOVVaw0cDkuBBUhmnun aozSvv>

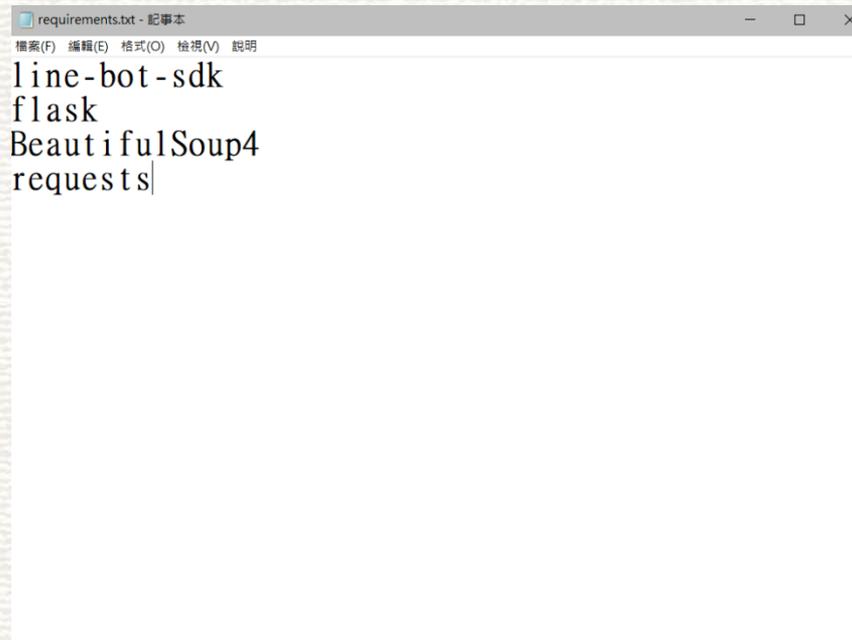
解決:

- 1.刪除&sa=U&ved= 2與後面的亂碼
- 2.刪除「https://」(可有可無)
- 3.印出第8個數字後到wantdelect前的內容
(0~7)

A large, horizontal, orange brushstroke graphic that serves as a background for the text. The stroke is composed of many overlapping, slightly curved lines, giving it a textured, hand-drawn appearance.

4. 匯入Linebot問題

爬蟲程式匯入Linebot問題1



```
requirements.txt - 記事本
檔案(F) 編輯(E) 格式(O) 檢視(V) 說明
line-bot-sdk
flask
BeautifulSoup4
requests|
```

問題：匯入Linebot後沒有任何反應

原因：套件缺少

解決：將所需套件名稱打在requirements.txt文件中

於本機端中輸入pip install requirements.txt

安裝缺少的套件

爬蟲程式匯入Linebot問題2

```
elif '爬蟲' in msg:
    msg = msg.split(' ')
    key_word = msg[1]
    message = search_crawler(key_word)
    line_bot_api.reply_message(event.reply_token, message)

import requests
from bs4 import BeautifulSoup

google_url = 'https://www.google.com.tw/search'
information = requests.get(google_url, params = {'q': '活動'})

if information.status_code == requests.codes.ok:

    # 以 BeautifulSoup 解析 HTML 原始碼
    soup = BeautifulSoup(information.text, 'html.parser')

    # 以 CSS 的選擇器來抓取 Google 的搜尋結果
    items = soup.select('div.kCrYT > a[href^="/url"]')
    text = []

    for i in items:
        text.append(i.text+'\n')
        text.append(i.get('href')[7:9999]+''\n')
        print(i.text)

string = ''.join(text)
print(string)
```

<https://www.google.com/search?q=%E5%AF%92%E6%B5%81>

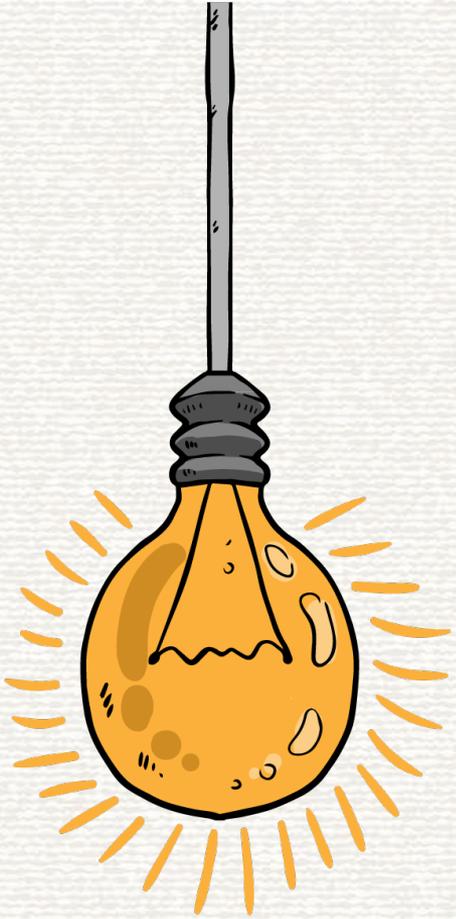
問題：匯入Linebot後爬蟲無法反應

原因：Linebot不知道我正在爬蟲

解決：用「,」告訴Linebot需要爬蟲與爬蟲資料,

EX：寒流, 爬蟲

event.reply_token：搜尋文字、message：爬蟲



謝謝大家!