*系級： 學號： 姓名：*

期末Python程式：網路擷取圖片程式

學習重點：網頁圖片擷取的應用

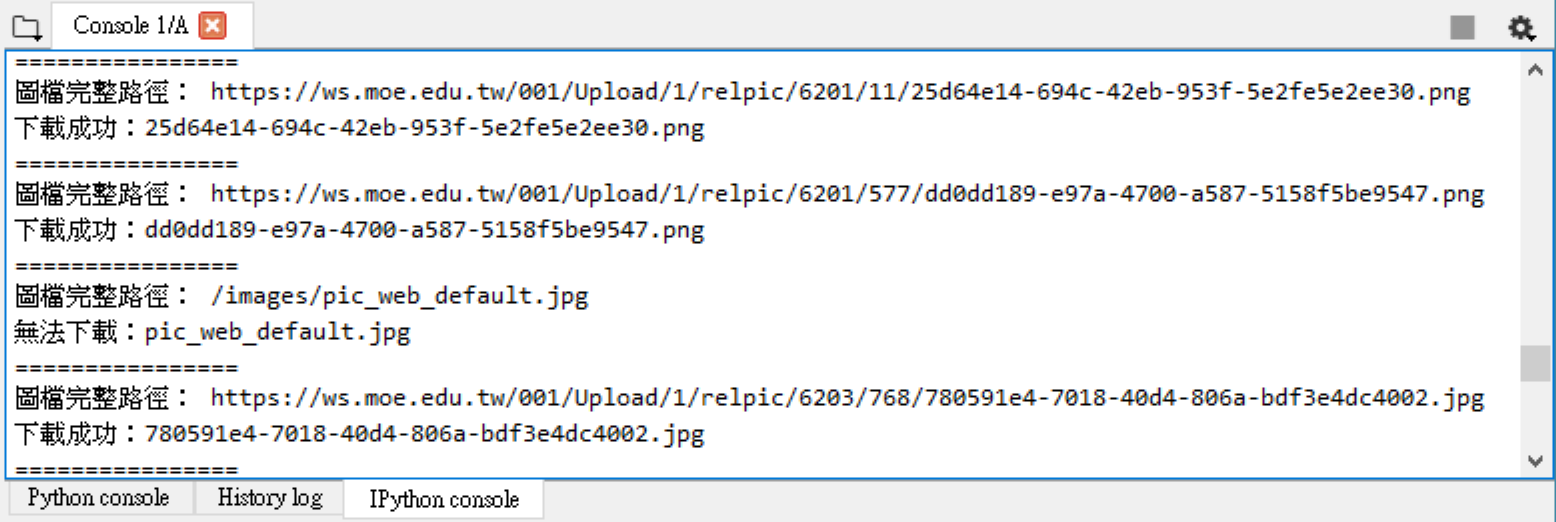
### 一、設計此程式的動機

### 二、程式設計目標

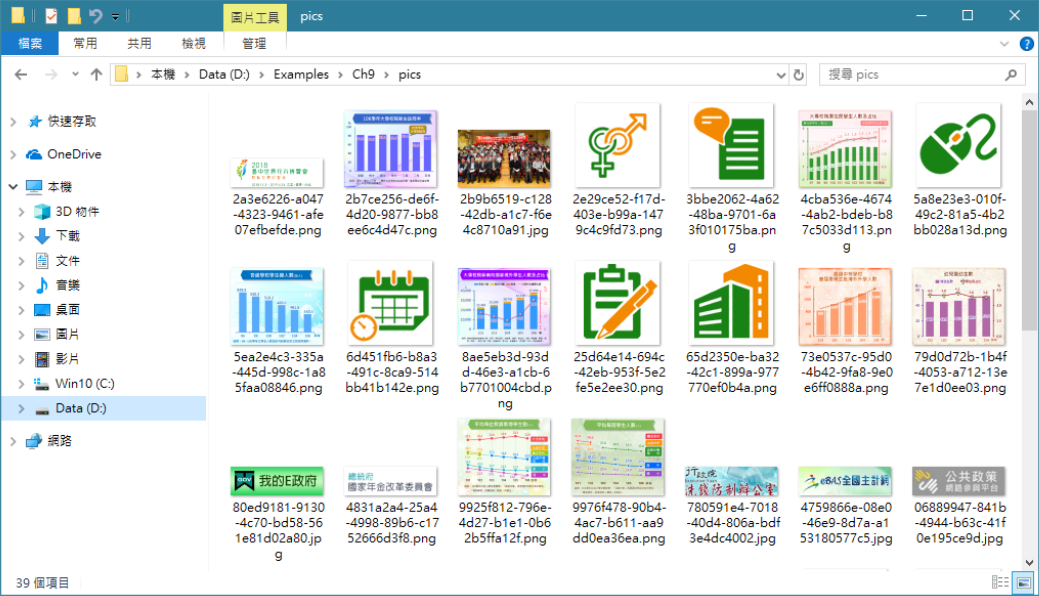
撰寫Python程式，自動擷取教育部網頁的圖片，教育部網站的網址為「https://www.edu.tw/」，其網站如圖所示。



執行程式的時候，會抓取網頁原始碼上所有的「img」標籤資料，進行網頁上圖檔的下載，並且會印出圖檔下載的狀況，下載成功或無法下載，如圖所示。



程式會自動在「9-6-2.py」檔案的工作目錄，建立資料夾「pics」，並將網頁擷取下來的圖檔置於其中，其下載結果如圖所示。



### 三、參考程式碼

|  |  |
| --- | --- |
| 列數 | 程式碼 |
| 1  2  3  4  5  6  7  8  9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24  25  26  27  28 | #網路擷取圖片程式  import requests,os,urllib  from bs4 import BeautifulSoup  url='https://www.edu.tw/'  html=requests.get(url)  html.encoding="utf-8"  bs=BeautifulSoup(html.text,'html.parser')  pics\_dir="pics"  if not os.path.exists(pics\_dir):  os.mkdir(pics\_dir) #在工作目錄下建立目錄pics來儲存圖片  all\_links=bs.find\_all('img') #用串列取得所有<img>標籤的內容  for link in all\_links:  src=link.get('src') #讀取src屬性值的內容  attrs=[src] #將src字串轉成串列attrs  for attr in attrs:  if attr!=None and('.jpg'in attr or'.png'in attr):#讀取.jpg或.png檔  full\_path = attr #設定圖檔完整路徑  file\_n=full\_path.split('/')[-1] #從網址的最右側取得圖檔的名稱  print('================')  print('圖檔完整路徑：',full\_path)  try: #儲存圖片程式區塊  image = urllib.request.urlopen(full\_path)  f = open(os.path.join(pics\_dir,file\_n),'wb')  f.write(image.read())  print('下載成功：%s' %(file\_n))  f.close()  except: #無法儲存圖片程式區塊  print("無法下載：%s" %(file\_n)) |

### 四、程式碼解說

第2、3行：匯入各項所需的套件，包括：requests套件、os套件、urllib套件、BeautifulSoup套件等。

第4行：設定教育部網站網址「https://www.edu.tw/」給變數url。

第8～10行：設定存放圖片檔案的資料夾名稱，如果資料夾不存在，則在工作目錄下建立目錄「pics」來儲存圖片檔案。

第11行：使用find\_all( )函式取得所有<img>標籤的內容。

第12～14行：使用for迴圈對每一個<img>標籤做處理，讀取src屬性值的內容，然後將src字串轉成串列attrs。

第15～18行：對串列attrs內的每一項做處理，如果該項非空，且副檔名為「.jpg」或「.png」，則設定圖檔完整路徑給變數「full\_path」，並從網址的最右側取得圖檔的名稱指定給變數「file\_n」。

第21～26行：使用urlopen( )函式取得圖檔資料，並以二進位寫入模式「wb」，將下載的檔案以同檔名寫到工作目錄下之資料夾內，完成後印出「下載成功：」加上檔案名稱字串，最後進行檔案關閉。

第27、28行：當檔案下載失敗時，進入此「except」區塊，印出「無法下載：」加上檔案名稱字串。

### 五、對於大數據與程式設計導論課程的學習心得