『巴哈姆特資訊看版探勘』

F109118121_許唐維

i.Data crawl

We crawl three topics in bahamut website, and we use manual headers to access because bahamut has anti-crawl feature.

```
bsn_links =['60030', '60001', '60559']
bsn_categories=['電腦應用綜合討論', '電視遊樂器綜合討論', '智慧型手機']
base_url = 'https://forum.gamer.com.tw/'

HEADERS = {
    'User-Agent': 'Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/81.0.4044.92 Safari/537.36',
}
```

First, we get all article pages from every topic at the first page.

Then we get the article total reply page to get all the replies.

```
def get_article_url_list(forum_url):
    r = requests.get(forum_url, headers=HEADERS)
    if r.status_code != requests.codes.ok:
        print("載入失敗")
        return []

    article_url_list = []
    soup = BeautifulSoup(r.text,'lxml')
    item_blocks = soup.select('table.b-list > tr[class="b-list__row b-list-item b-imglist-item"]')
    for item_block in item_blocks:
```

```
        title_block = item_block.select_one('.b-
list__main__title')
        article_url = f"https://forum.gamer.com.tw/{title_block.ge
t('href')}"
        article_url_list.append(article_url)

    return article_url_list



def get_article_total_page(soup):
    article_total_page = soup.select_one('.BH-pagebtnA > a:last-
of-type').text
    return int(article_total_page)




#主題頁面資訊
def get_article_info(article_url):
    soup = BeautifulSoup(requests.get(article_url, headers=HEADERS
).text,'lxml')


    article_title = soup.select_one('h1.c-
post__header__title').text

    article_total_page = get_article_total_page(soup)  #獲得總樓層
的數量

    reply_info_list = []
    for page in range(article_total_page):
        crawler_url = f"{article_url}&page={page + 1}"
        reply_list = get_reply_info_list(crawler_url)
        reply_info_list.extend(reply_list)
        random.uniform(1, 3)

    article_info = {
        'title': article_title,
```

```
        'url': article_url,
        'reply': reply_info_list[0],
        'category': category
    }
    return article_info

def get_reply_info_list(url):

    reply_info_list = []
    soup = BeautifulSoup(requests.get(url, headers=HEADERS).text,'
lxml')
    reply_blocks = soup.select('section[id^="post_"]')

    for reply_block in reply_blocks:

        reply_info = reply_block.select_one('.c-
article__content').text
        reply_info = re.sub(r'\n+',"", reply_info)
        reply_info_list.append(reply_info)
    return reply_info_list
```

We use dataframe to save our data, which display below.

	item_id	category	title	content	link
0	電腦應用綜合討論_1	電腦應用綜合討論	【心得】原價屋讓我不再想去了	本人在二月中的時候在原價屋買了一個24吋螢幕，直到上禮拜才發現電腦有歪斜的情況，致電給原價屋...	https://forum.gamer.com.tw/C.php?bsn=60030&snA...
1	電腦應用綜合討論_2	電腦應用綜合討論	【問題】請問ppsspp模擬器玩部分遊戲DS4手把問題	如提玩古墓奇兵周年紀念跟第三次生日用DS4手把玩用左類比操控發現不大靈敏人物都用走路的後來調...	https://forum.gamer.com.tw/C.php?bsn=60030&snA...
2	電腦應用綜合討論_3	電腦應用綜合討論	【閒聊】維修費 是否合理	前幾天電腦壞了 拿去維修店家告知是 電源供應器 和 主機板壞了 需要更換後來換了 電源供應...	https://forum.gamer.com.tw/C.php?bsn=60030&snA...
3	電腦應用綜合討論_4	電腦應用綜合討論	【問題】想問一下主機板選擇問題？	Rt 最近要組電腦打算買b550系列的主機板有爬文說主機板建議買大張不買小張也就是買atx ...	https://forum.gamer.com.tw/C.php?bsn=60030&snA...
4	電腦應用綜合討論_5	電腦應用綜合討論	【問題】CPU選擇請益	最近想組一臺桌機現在那臺筆電用了快6年 硬體慢慢跟不上了但在找CPU選購的時候不太清楚要買A...	https://forum.gamer.com.tw/C.php?bsn=60030&snA...

Then we use CkipLab to tokenize our content and count the word frequency, which shows below.

```
category_freq = []
word = []

for i in range(0, len(df2)):

    tokens = ws(df2.all_contents[i])
    tokens_pos = pos(tokens)

    word_pos_pair = [list(zip(w,p)) for w, p in zip(tokens, t
okens_pos)]

    with open('stops_chinese_traditional.txt', 'r', encoding=
'utf8') as f:
        stops = f.read().split('\n')

    allowPOS=['Na','Nb','Nc','VA','VAC','VB','VC']

    tokens_v2 =[]
    for wp in word_pos_pair:
        tokens_v2.append([w for w,p in wp if w not in stops a
nd (len(w) >= 2) and p in allowPOS])

    tokens_pos = pos(tokens_v2)
    word_pos_pair = [list(zip(w,p)) for w, p in zip(tokens_v2
, tokens_pos)]

    word.append(tokens_v2[0])

    keyfreqs =[]
    filtered_words =[]

    for wp in word_pos_pair:
        word_frequency(wp)
    counter = Counter(filtered_words)
    keyfreqs.append(counter.most_common(200))

    category_freq.append(keyfreqs[0])
```

	category	all_contents	tokens	freq
0	電腦應用綜合討論	[本人在二月中的時候在原價屋買了一個24吋螢幕，直到上禮拜才發現電腦有歪斜的情況，致電給原價...	[原價屋, 螢幕, 禮拜, 電腦, 情況, 致電, 原價屋, 電話, 購買, 華碩, 皇家,...	[(螢幕, 28), (電腦, 13), (問題, 12), (硬碟, 12), (中國, ...
1	電視遊樂器綜合討論	[D.I.C.E. Awards是由遊戲界具有影響力的人物組成的互動藝術與科學學院都(Ac...	[D.I.C.E. Awards, 遊戲界, 影響力, 人物, 互動, 藝術, 科學, 學院...	[(遊戲, 83), (小屋, 44), (板綜板, 41), (玩家, 27), (瑪利歐...
2	智慧型手機	[因為合約下個月到期了目前xz3用兩年半了 開始一直過熱平常玩遊戲都是一些載不吃性能的主要還...	[合約, 遊戲, 性能, 拿來, 影片, 拍拍, 生活照, 手機, 發表, 價錢, 品牌, ...	[(手機, 29), (使用, 11), (螢幕, 11), (系統, 9), (更新, 9...

We use freq to present our midterm project with Django.

ii. Django

We use Django to create analysis website.

1. Create a new Django project

Step 1: Create a folder named “midtest”

```
mkdir midtest
```

Step 2: Go into the folder “midtest”

```
cd midtest
```

Step 3: Create a project configures folder named website_configs

```
django-admin startproject website_configs .
```

2. Create an APP

Step 1: Create an APP named “app_top_keyword”

```
django-admin startapp app_top_keyword
```

Step 2: setting website_configs’ s settings.py and urls.py

Setting setting.py

```
import os

ALLOWED_HOSTS = ['127.0.0.1']

INSTALLED_APPS = [
     ……
    'app_top_keyword',
]

TEMPLATES = [
    {
      ……
     'DIRS': [os.path.join(BASE_DIR, 'templates')],
      ……
```

Setting urls.py

```
from django.contrib import admin
from django.urls import path
from django.urls import include
urlpatterns = [
    #path('admin/', admin.site.urls),
    path('topword/', include('app_top_keyword.urls')),
]
```

3. Setting app_top_keyword’ s views.py and create urls.py

Setting views.py

```
from django.views.decorators.csrf import csrf_exempt
from django.shortcuts import render
from django.http import JsonResponse
import pandas as pd
from wordcloud import WordCloud, STOPWORDS, ImageColorGenerator
import matplotlib.pyplot as plt
# render 渲染網頁
def home(request):
```

```
    return render(request, 'app_top_keyword/home.html')
# read df
df_topkey = pd.read_csv(
    'app_top_keyword/dataset/baha_dataset_freq.csv', sep=',')
# prepare data
data = {}
for idx, row in df_topkey.iterrows():
    data[row['category']] = eval(row['top_keys'])
# We don't use it anymore, so delete it to save memory.
del df_topkey
# POST: csrf_exempt should be used
# 指定這一支程式忽略 csrf 驗證
@csrf_exempt
def api_get_cate_topword(request):
    cate = request.POST.get('news_category')
    # cate = request.POST['news_category']  # this command also wo
rks.
    topk = request.POST.get('topk')
    topk = int(topk)
    print(cate, topk)
    chart_data, wf_pairs = get_category_topword(cate, topk)
    response = {
        'chart_data': chart_data,
        'wf_pairs': wf_pairs,
    }
    print(response)
    return JsonResponse(response)
def get_category_topword(cate, topk=10):
    wf_pairs = data[cate][0:topk]
    print(data)
    words = [w for w, f in wf_pairs]
    freqs = [f for w, f in wf_pairs]
    chart_data = {
        "category": cate,
        "labels": words,
        "values": freqs}
    return chart_data, wf_pairs
print("app_top_keywords--類別熱門關鍵字載入成功!")
```

Create app_top_keyword urls.py

In folder app_top_keyword, create a python file named “urls.py”

create app_top_keywrod/urls.py

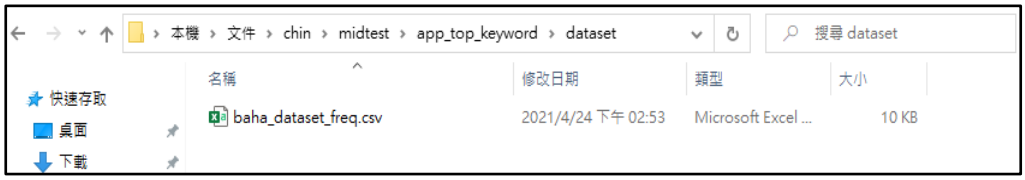
Setting app_top_keywrod/urls.py

```
from django.urls import path
from app_top_keyword import views
# Declare a namespace for this APP
app_name = 'app_top_keyword'
urlpatterns = [
    # For home
    path('', views.home, name='home'),  # app_top_keyword:home
    # For Ajax
    path('api_get_cate_topword/', views.api_get_cate_topword),
]
```

4. loading dataset

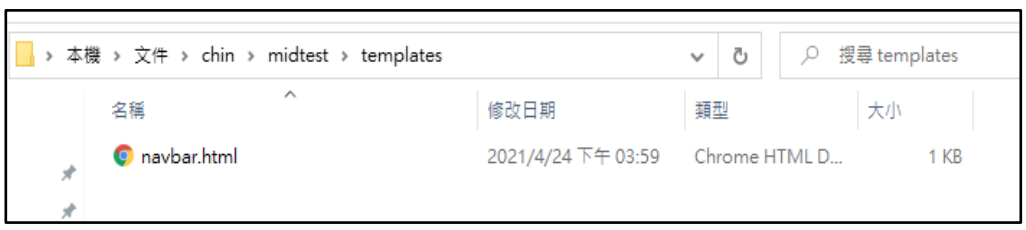
In folder app_top_keyword, create a file named “dataset” and

loading freq.csv.



5. create tempates in file midtest, then create navbar.html

Create midtest/tempates/navbar.html

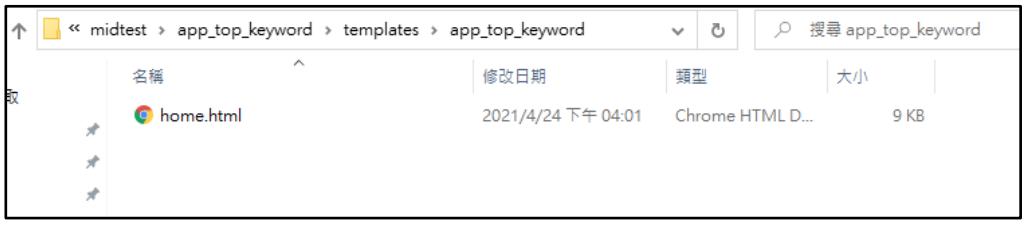


Setting nabar.html

```
<div class="col-lg-12 mb-2">
    <nav class="navbar navbar-expand-lg navbar-
light" style="background-color: #e3f2fd;">
        <a class="navbar-brand" href="#">巴哈網站大數據</a>
        <button class="navbar-toggler" type="button" data-
toggle="collapse" data-target="#navbarSupportedContent"
                aria-controls="navbarSupportedContent" aria-
expanded="false" aria-label="Toggle navigation">
            <span class="navbar-toggler-icon"></span>
        </button>
        <div class="collapse navbar-
collapse" id="navbarSupportedContent">
            <ul class="navbar-nav mr-auto">
                <li class="nav-item">
                    <a class="nav-
link" href="{% url 'app_top_keyword:home' %}">熱門關鍵詞分析</a>
                </li>
            </ul>
        </div>
    </nav>
</div>
```

6. create app_top_key_word/templates/app_top_key_word /home.html

Create app_top_key_word/templates/app_top_key_word /home.html



Setting home.html

```
<!DOCTYPE html>
<html lang="en">
<head>
    <title>輿情分析平台</title>
```

```
    <meta charset="utf-8">
    <meta name="viewport" content="width=device-width, initial-
scale=1">
    <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/b
ootstrap/4.3.1/css/bootstrap.min.css">
    <script src="https://ajax.googleapis.com/ajax/libs/jquery/3.3.
1/jquery.min.js"></script>
    <script src="https://cdnjs.cloudflare.com/ajax/libs/popper.js/
1.14.7/umd/popper.min.js"></script>
    <script src="https://maxcdn.bootstrapcdn.com/bootstrap/4.3.1/j
s/bootstrap.min.js"></script>
</head>

<body>
    <div class="container">
        <div class="row">
            <!-- Here insert the navigation bar -->
            {%include 'navbar.html'%}
            <div class="col-lg-12">
                <h1>巴哈三個專版的關鍵詞</h1>
                <p>熱門度分析:可以了解專版關注那些重要的議題</p>
            </div>
            <!-- 新聞類別選單------------------------------>
            <div class="col-lg-6 mb-2">
                <div class="card">
                    <div class="card-header">
                        <h3 class="h6 text-uppercase mb-0">熱門關鍵
字瀏覽與繪圖(資料週期:資料截止時間為上週五)</h3>
                    </div>
                    <div class="card-body">
                        <!--新聞類別選單 form group-->
                        <div class="form-group row">
                            <label class="col-sm-3 form-control-
label">專版類別</label>
                            <div class="col-md-9">
                                <select id="cate-
selected" name="news_category" class="form-control">
                                    <!--<option>請選擇</option>-->
```

```
                                    <option>電腦應用綜合討論
</option>
                                    <option>電視遊樂器綜合討論
</option>
                                    <option>智慧型手機</option>
                                </select>
                                <small class="form-text text-
muted">請選擇專版類別
                                </small>
                            </div>
                        </div>
                        <!--form group-->
                        <!--熱門詞多少個?form group-->
                        <div class="form-group row">
                            <label class="col-md-3 form-control-
label">多少個熱門詞?</label>
                            <div class="col-md-9">
                                <input id="topk-
selected" name="topk" value="10"
                                    class="form-control form-
control-success">
                                <small class="form-text text-
muted">內定值為 10
                                </small>
                            </div>
                        </div>
                        <!--form group-->
                        <!--submit 按鈕 form group-->
                        <div class="form-group row">
                            <div class="col-md-9 ml-auto">
                                <button type="button" id="btn-
ok" class="btn btn-primary">查詢</button>
                            </div>
                        </div>
                        <!--form group-->
                    </div>
                    <!--card body-->
                </div>
```

```
                <!--column-->
            </div><!-- 區塊結束 -->
            <!-- 繪圖區塊----------------------------------------
---------------->
            <div class="col-lg-6 mb-5">
                <div class="card">
                    <div class="card-header">
                        <h3 class="h6 text-uppercase mb-0">熱門關鍵
字繪圖</h3>
                    </div>
                    <div class="card-body">
                        <canvas id="mychart"></canvas>
                    </div>
                </div>
            </div><!-- 區塊結束 -->
            <!-- 熱門關鍵字區塊----------------------------------
----------------->
            <div class="col-lg-6 mb-5">
                <div class="card">
                    <div class="card-header">
                        <h3 class="h6 text-uppercase mb-0">熱門關鍵
字</h3>
                    </div>
                    <div class="card-body">
                        <ul id="topkeys"></ul>
                    </div>
                </div>
            </div><!-- 區塊結束 -->
        </div> <!-- row 結束-->
    </div> <!-- container 結束-->
</body>
</html>
<!-- chartjs 圖 js-->
<script src="https://cdnjs.cloudflare.com/ajax/libs/Chart.js/2.7.3
/Chart.min.js"></script>
<!-- 程式碼區 -->
<script>
// Write your JS code here!
```

```
call_ajax();
//let cate = $('#cate-selected').val();
//console.log(cate);

//let topk = $('#topk-selected').val();
//console.log(topk);
//alert(topk);
   $('#btn-ok').on('click', function () {
    console.log("按下按鈕");
    call_ajax();
  showTopKeys(wf_pairs);
   });
 //*新聞類別選單 select 被選中值有改變時，執行以下事件
   $('#cate-selected').on('change', function () {
     let cate = $('#cate-selected').val();
           console.log(cate);
   }); //event function
// Exercise#2: Define a function
// Please paste showTopKeys function here!
    //* 顯示關鍵詞資料函數
    function showTopKeys(items) {
        //先清除前一次的資料
        $('#topkeys').empty();

        //將內容加上 li 標籤附加起來，顯示在顯示區"topkeys"
        for (let i = 0; i < items.length; i++) {
            let item_li = "<li>" + items[i] + "</li>";
            $('#topkeys').append(item_li);
        }
    } //function
    // Call function when btn_ok is clicked
// Exercise#4: Define “call_ajax” function to perform Ajax
// Call ajax function when page is loaded and button is clicked.
   // See what the data received from backend API looks like.
   // Display word frequency pairs.
   function call_ajax() {
        let cate = $('#cate-selected').val();
        let topk = $('#topk-selected').val();
```

```
        $.ajax({
            type: "POST",
            //url: "/topword/api_get_cate_topword/",
            url: "http://127.0.0.1:8000/topword/api_get_cate_topwo
rd/",
            //url: "http://163.18.22.32:8000/topword/api_get_cate_
topword/",
            //url: "api_get_cate_topword/", //Not recommended!
            data: { "news_category": cate, "topk": topk },
            success: function (received) {
                console.log(received);
                let chart_data = received.chart_data;
                let wf_pairs = received.wf_pairs;
                console.log(wf_pairs);
                showTopKeys(wf_pairs);
                showChart(chart_data);
                //showChart(chart_data);
            } //success function
        }); //ajax
    } //call_ajax
 //**繪圖函數 showChart()
 function showChart(chart_data) {
// 畫圖需要的數據資料
let values = chart_data.values;
let labels = chart_data.labels;
let category = chart_data.category;
//第 1 個變數: 餵給 chart 的資料 data
let data = {
    labels: labels,
    datasets: [{
        label: category,
        data: values,
        backgroundColor: randomColors(values.length),
        borderColor: randomColors(values.length),
        borderWidth: 1,
    }],
};
//第 2 個變數: chart 的選項  指定 y 坐標軸從零開始顯示
```

```
let options = {
    scales: {
        yAxes: [{
            ticks: {
                beginAtZero: true
            }
        }]
    },
};
//取得在前面 html 區域欲顯示的圖代號
let canvas_mychrat = document.getElementById("mychart");


//**先清除前一個圖 再繪新圖
// 可以印出 barchart 物件是否存在
// console.log(window.barchart);
//先清除前一個圖 再繪新圖 if 有以下兩種寫法皆可
// if (window.barchart)  //若存在則為 true
// if (typeof (barchart) != "undefined"){
if (window.barchart) {
    barchart.destroy();
}
//**繪圖(產生一個圖物件變數名稱為 barchart)
// 必須全域變數--注意:前面不要有 let, var, const 等修飾詞
// 理由: 我們要讓它存在於網頁全域變數，
// 這樣我們才方便判斷是否有前一次的圖，如果存在有，要刪除之，否則，很多張圖會疊在一起
barchart = new Chart(canvas_mychrat, {
    type: 'bar',
    data: data,
    options: options,
});
//** 產生隨機顏色
function randomColors(num_colors) {
    let colors = [];
    for (i = 0; i < num_colors; i++) {
        let r = Math.floor(Math.random() * 255);
        let g = Math.floor(Math.random() * 255);
        let b = Math.floor(Math.random() * 255);
```

```
        let rgb = `rgba(${r},${g},${b},0.5)` // (red, green, blue,
 alfa) alfa 透明度
        colors.push(rgb);
    }
    return colors;
}
} //show chart function
</script>
```

7.Start Django Sever, show the midtest websire

Step 1: Start Django Sever

python manage.py runserver

step 2: 127.0.0.1:8000/topword/ show the webside