

政治即時通

企業資料通訊期末報告

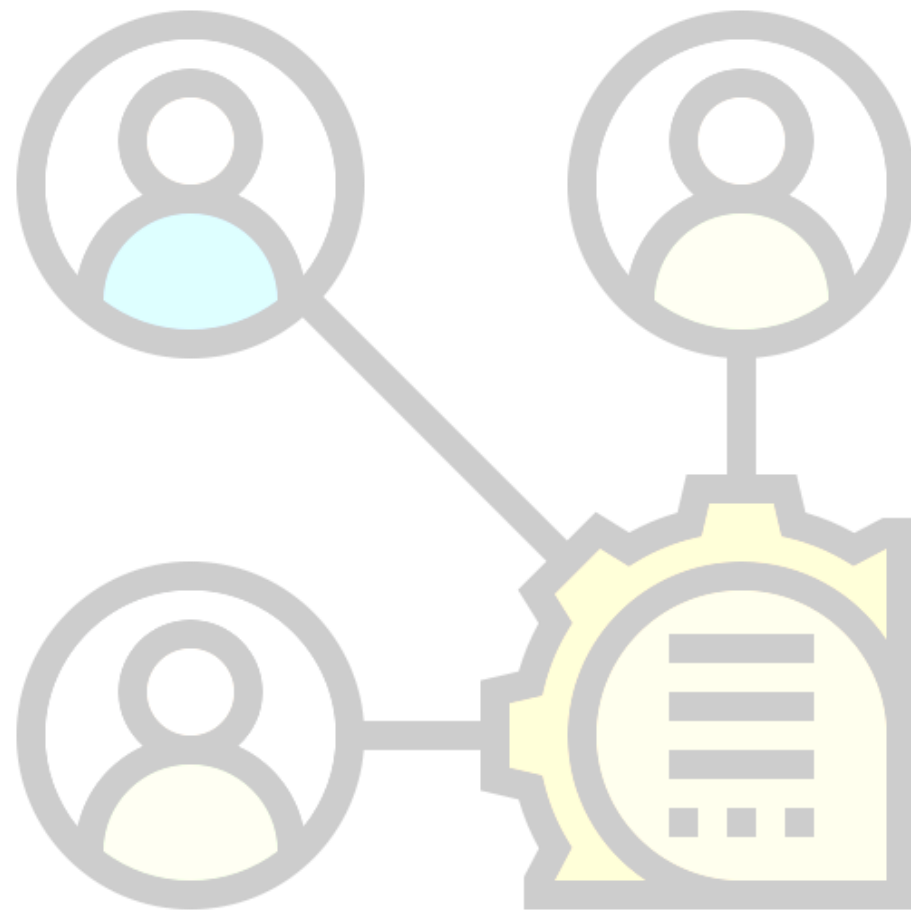
[第十組] B064020008黃靖涵
B064020031陳妍君
B051010016龔雪燕

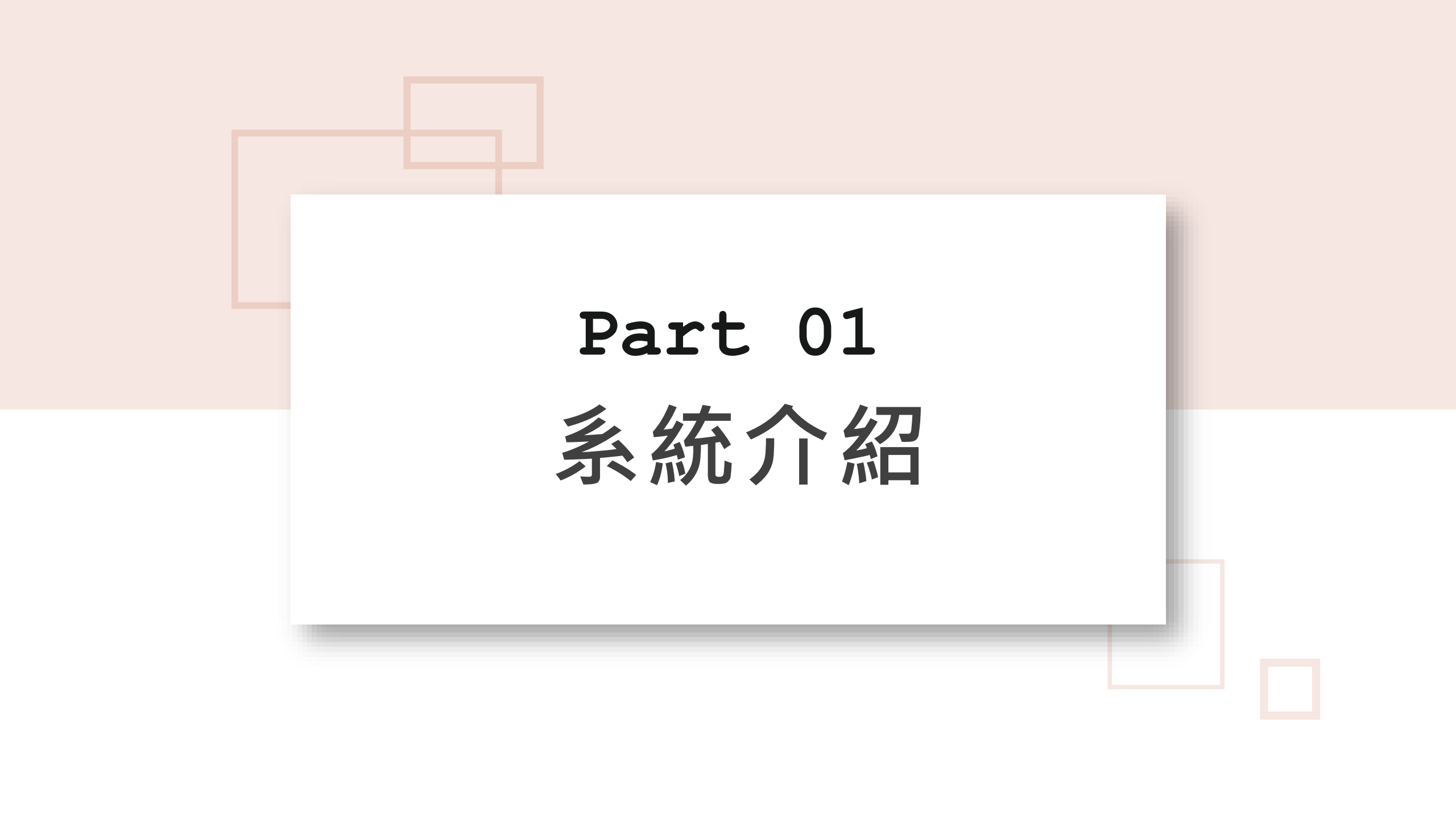
OUTLINE

Part 01
系統介紹

Part 02
操作說明

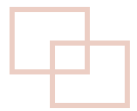
Part 03
程式開發





Part 01

系統介紹



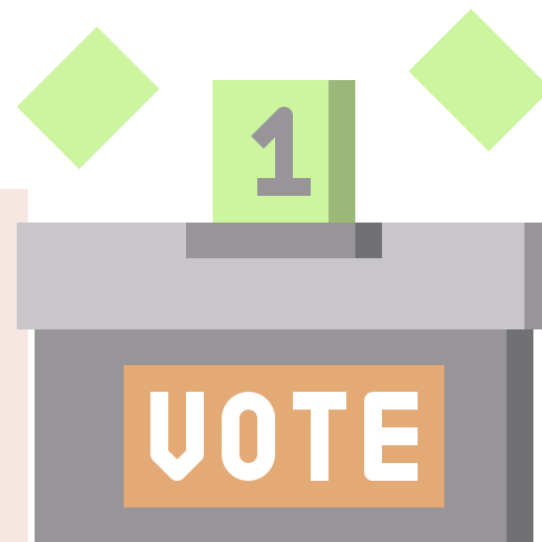
開發動機

別讓小島
沉沒



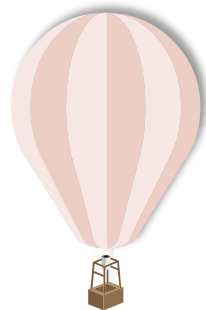
因應時事

配合2020年總統大選，提供使用者查詢PTT最新話題討論以及最即時的新聞資訊。





系統特色



NETWORKING

透過php的curl function擷取
PTT政黑版及各新聞網站資料

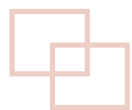
1. PTT政黑版

系統使用者輸入關鍵字，並將關鍵字
用**GET**的方式傳至ptt.php，再使用
curl的方式到PTT政黑版搜尋相關
推文。擷取**最新的十篇推文**，並顯示
於網頁下方的iframe中。

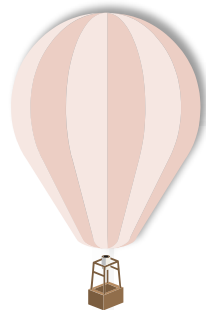
PTT SEARCH:

Keyword:

SEARCH!



系統特色



NETWORKING

透過php的curl function擷取PTT政黑版及各新聞網站資料

2. 各新聞網站

一樣藉由系統使用者輸入關鍵字、選取新聞網站，並用**GET**的方式傳至news.php，再使用**curl**的方式到該新聞網站擷取**最新的十則新聞**，並顯示於網頁下方的iframe中。

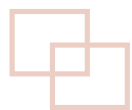
NEWS SEARCH:

Keyword:

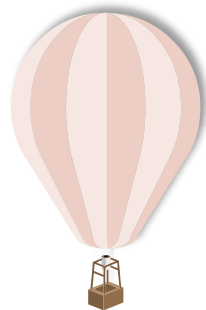
選舉

- ☐ TVBS
- ☒ 聯合新聞網
- ☐ 風傳媒
- ☐ 蘋果日報
- ☐ ETtoday
- ☐ 中央社
- ☐ 自由電子報
- ☐ 三立新聞網

SEARCH!



系統特色



PROGRAMMING

1. HTML

- 網頁前端撰寫

2. CSS

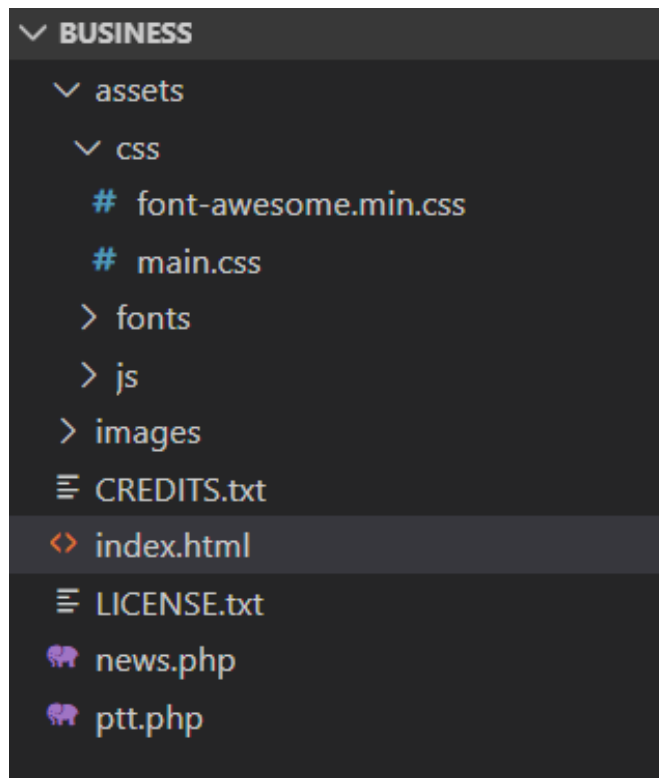
- 網頁前端美化

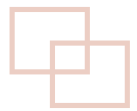
3. Javascript

- 動態網頁

4. PHP

- 網頁後端撰寫





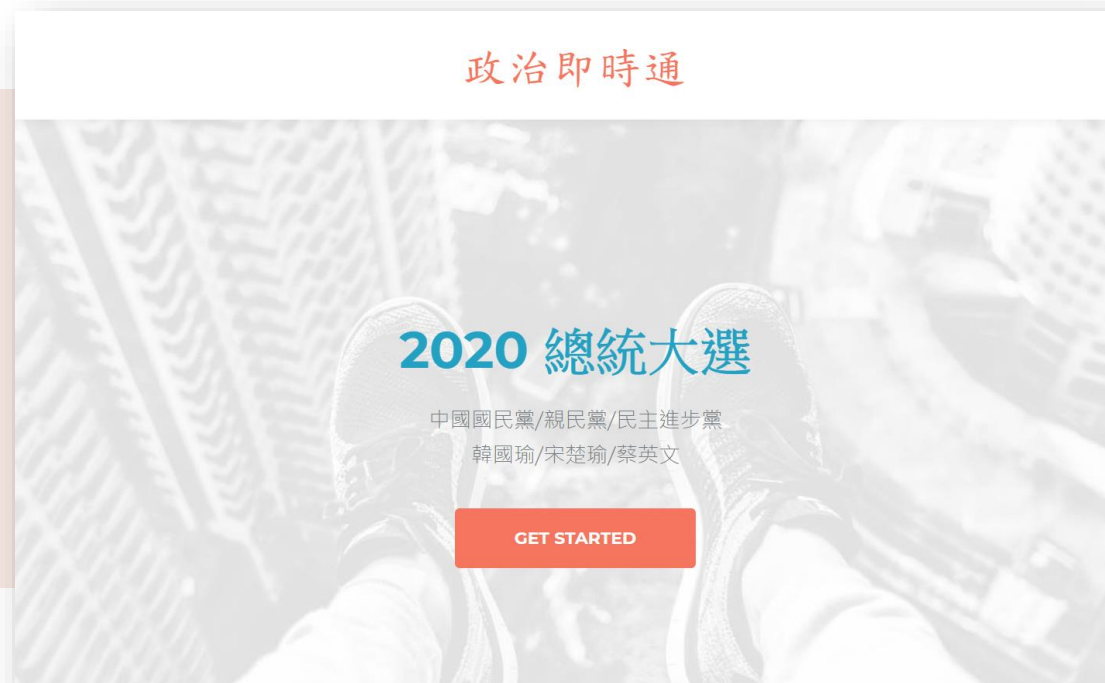
系統開發工具



XAMPP



PHP



CSS



HTML

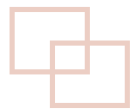


Javascript

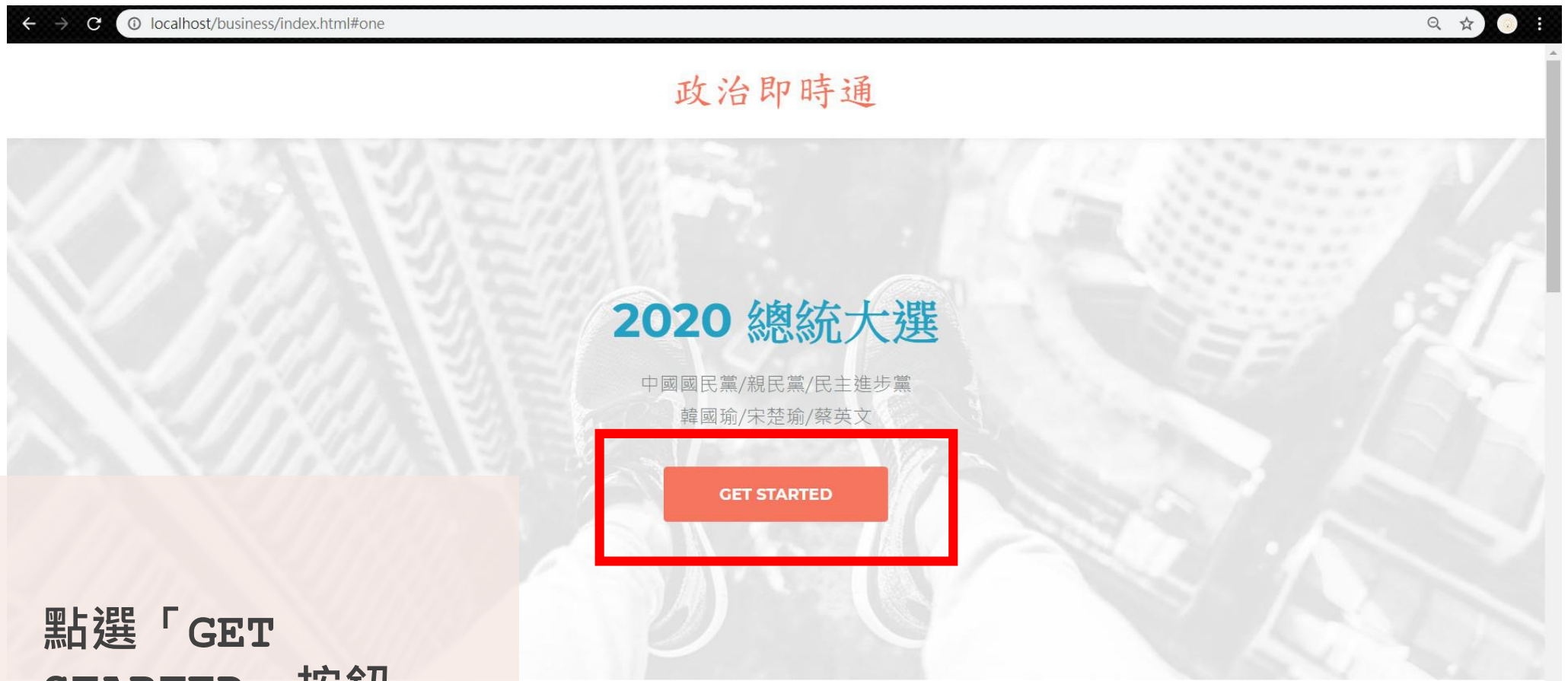


Part 02

操作説明

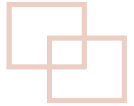


操作說明-PTT



01

點選「GET
STARTED」按鈕



操作說明-PTT

PTT SEARCH:

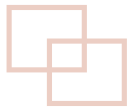
Keyword:

選舉

SEARCH!

02

輸入關鍵字並點選
「SEARCH!」按鈕



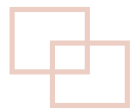
操作說明 - PTT

1. [新聞] 斷言2020大選蔡英文會贏 柯文哲：選民都
2. Re: [新聞] 柯媽爆料：柯文哲絕對會投蔡英文 蔡也會
3. [討論] 幫順邏輯！柯文哲是否支持小英清算自己？
4. Re: [新聞] 柯媽爆料：柯文哲絕對會投蔡英文 蔡也會
5. Re: [新聞] 柯媽爆料：柯文哲絕對會投蔡英文 蔡
6. Re: [新聞] 柯媽爆料：柯文哲絕對會投蔡英文 蔡也會
7. Re: [新聞] 柯媽爆料：柯文哲絕對會投蔡英文 蔡也會
8. Re: [新聞] 柯媽爆料：柯文哲絕對會投蔡英文 蔡也會
9. [黑特] 柯文哲無能，台北市竟然爆發如此嚴重的幣
10. Re: [新聞] 柯媽爆料：柯文哲絕對會投蔡英文 蔡

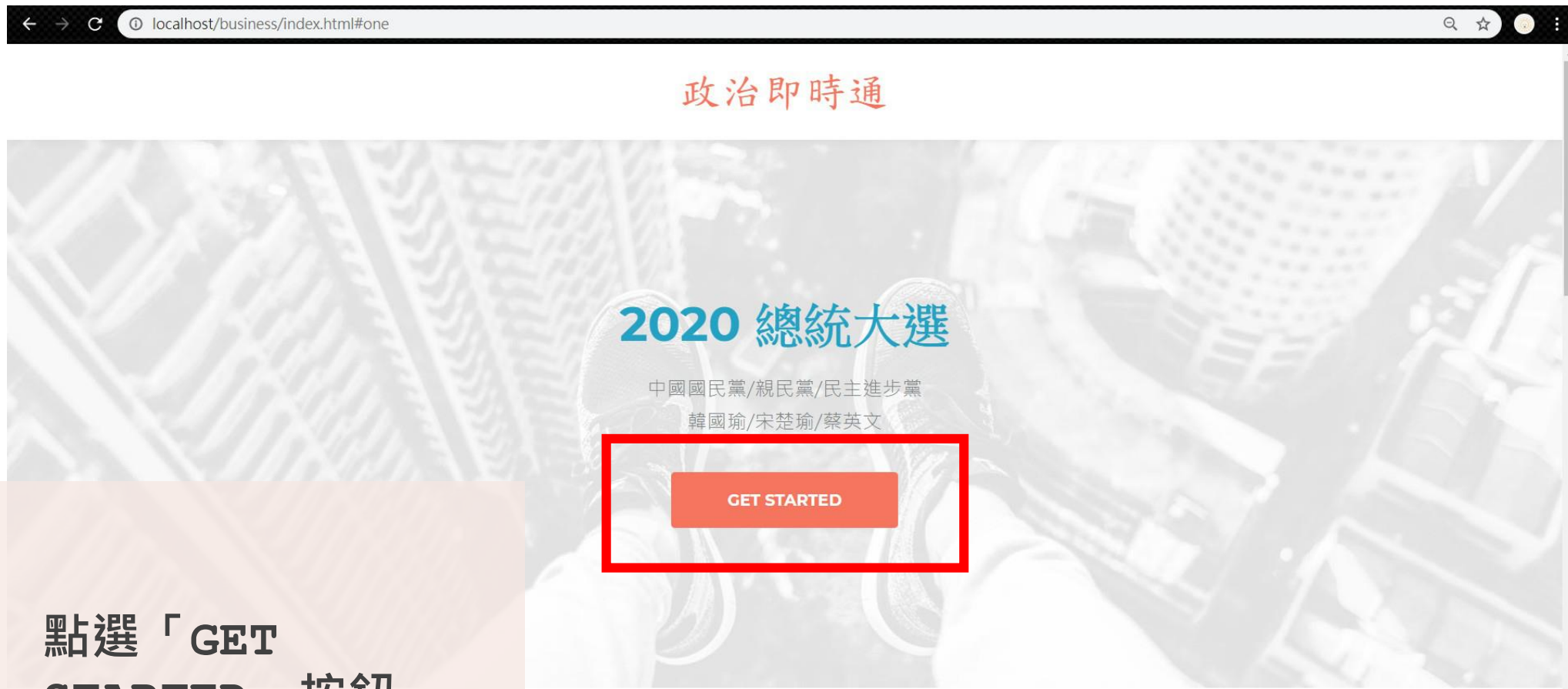
03

依照關鍵字搜尋結果，將顯示最新十篇推文於下方iframe中。點選欲瀏覽的推文後，便會開啟新分頁。

The screenshot shows a web browser displaying a PTT forum thread. The address bar shows the URL: ptt.cc/bbs/HatePolitics/M.1577357743.A.438.html. The forum header includes the board name '批踢踢實業坊' and the sub-board '看板 HatePolitics'. The thread title is '[新聞] 斷言2020大選蔡英文會贏 柯文哲：選民都'. The author is 'girl10319 (凱莉)' and the post time is 'Thu Dec 26 18:55:40 2019'. The main content of the post discusses the 2020 presidential election, mentioning that the author's mother (柯媽) has leaked information that the author (柯文哲) will vote for蔡英文 (Cai Ingwen). The author expresses uncertainty about the outcome, stating that while they believe in the 'Iron Han' (鋼鐵韓粉) who support 韓國瑜 (Han Kuo-yu), they also acknowledge the 'Iron Han' who support 蔡英文 (Cai Ingwen). The author concludes by saying they will vote for 蔡英文 (Cai Ingwen) because they believe she will win.

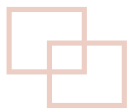


操作說明－新聞網頁



01

點選「GET
STARTED」按鈕



操作說明－新聞網頁

NEWS SEARCH:

Keyword:

選舉

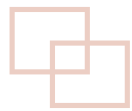
- ☐ TVBS
- ☒ 聯合新聞網
- ☐ 風傳媒
- ☐ 蘋果日報
- ☐ ETtoday
- ☐ 中央社
- ☐ 自由電子報
- ☐ 三立新聞網

SEARCH!

偷偷按照新聞黨派顯示出來了

02

輸入關鍵字，並欲察看的點選新聞媒體，並點選「SEARCH!」按鈕



操作說明－新聞網頁

1. 賴清德直搗藍營會大安區：守護台灣成老量 第一選擇謝佩芬
2. 罷韓尬挺韓！陳其邁不出席遊行：對高雄有信心
3. 澎湖立委選舉在地人「內戰」 歐中慨扮深耕軍公教「加持」
4. 男網路分享「同婚讓台灣成愛滋樂園」 法官認定「謠言」判罰3千
5. 韓國瑜12/25上任市長滿周年！Wecare高雄下周四遞連署書 啟動第一階段罷
6. 嘆蔡英文「不是壞人」 陳佩琪：為何放任國家讓派系吃乾抹淨？
7. 不反你們這對活寶天兵兄弟要反誰？常美冰店發500冰棒支持罷韓
8. 外銀看2020年全球金融市場 追求穩健投資得留意3關鍵
9. 為35年前的殺警案自清 國民黨彰化立委蕭景田的真心告白
10. 綠營連江縣立委候選人「改編傳統軍歌」 跨四鄉五島向馬祖軍民致意



03

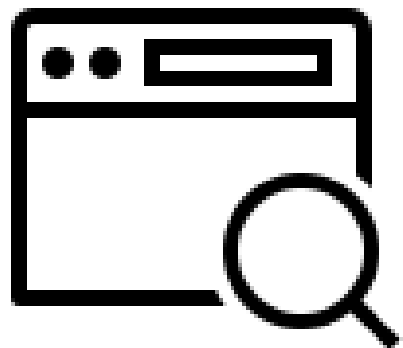
依照輸入的關鍵字及選擇的新聞媒體，將顯是最新十則新聞資訊於下方iframe中。點選欲瀏覽的新聞後，便會開啟新分頁。



Part 03

程式開發

分為三個部分：



首頁UI

前端UI相關設計
(陳妍君)



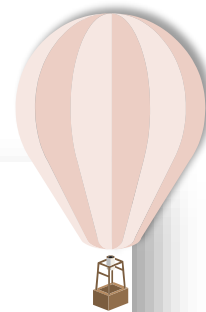
PTT搜尋

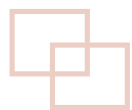
後端抓資料並顯示
(龔雪燕)



新聞搜尋

後端抓資料並顯示
(黃靖涵)





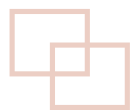
首頁UI-index.html (1/3)

```
<!-- Header -->
<header id="header">
  <a href="index.html" class="logo">政治即時通</a>
</header>

<!-- Banner -->
<section id="banner">
  <div class="content">
    <h1>2020 總統大選</h1>
    <p>中國國民黨/親民黨/民主進步黨<br/>韓國瑜/宋楚瑜/蔡英文</p>
    <ul class="actions">
      <li><a href="#one" class="button scrolly">Get Started</a></li>
    </ul>
  </div>
</section>
```

Header 、 Banner

將網頁之主旨顯示於header及banner，並創造一個按鈕「GET STARTED」，使用<a href>的方式連結到下方輸入關鍵字搜尋之頁面(#one)。



首頁UI-index.html (2/3)

```
<!-- One -->
<section id="one" class="search page">
  <div class="inner">
    <h1>PTT SEARCH:</h1>
    <h3>Keyword:</h3>
    <!--<form action="/action_page.php"-->
    <form method="GET" action="ptt.php" target="iframe">
      <input type="text" name="keyword"><br>
      <input type="submit" value="SEARCH!">
      <input name="url" type="hidden" value="https://www.ptt.cc/bbs/HatePolitics/search?q=">
    </form>
    <h1>NEWS SEARCH:</h1>
    <h3>Keyword:</h3>
    <form method="GET" action="news.php" target="iframe">
      <input type="text" name="keyword">
      <div>
        <br>
        <input type="radio" name="media" value="TVBS" id="bm1"><label for="bm1" style="color: #4F86C6;">TVBS</label>
        <input type="radio" name="media" value="聯合新聞網" id="bm2"><label for="bm2" style="color: #4F86C6;">聯合新聞網</label>
        <input type="radio" name="media" value="風傳媒" id="om1"><label for="om1" style="color: coral;">風傳媒</label>
        <input type="radio" name="media" value="蘋果日報" id="om2"><label for="om2" style="color: coral;">蘋果日報</label>
        <input type="radio" name="media" value="ETtoday" id="om3"><label for="om3" style="color: coral;">ETtoday</label>
        <input type="radio" name="media" value="中央社" id="gm1"><label for="gm1" style="color: #88dba3;">中央社</label>
        <input type="radio" name="media" value="自由電子報" id="gm2"><label for="gm2" style="color: #88dba3;">自由電子報</label>
        <input type="radio" name="media" value="三立新聞網" id="gm3"><label for="gm3" style="color: #88dba3;">三立新聞網</label>
      </div>
      <br>
      <input type="submit" value="SEARCH!">
    </form>
  </div>
</section>
```

Input type="text"

使用input type="text"創造一個關鍵字搜尋的可輸入方框，並將輸入的文字以**keyword**為name來傳送輸入的值。

Input type="submit"

使用input type="submit"創造一個button，以執行搜尋功能。

Input name="url" value=""

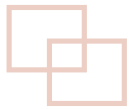
使用input name="url" type="hidden" value=""

來傳送PTT正黑版的網頁位址。

Input type="radio"

使用input type="radio"創造單選新聞網的選項，並利用<label>讓使用者點選文字即可以選擇該選項，而label for要填的值即對應到前方id之值。

另外，依新聞公司偏向之黨派不同，利用代表性顏色顯示於網頁上提醒使用者。



首頁UI-index.html (3/3)

```
<!-- One -->
<section id="one" class="search page">
  <div class="inner">
    <h1>PTT SEARCH:</h1>
    <h3>Keyword:</h3>
    <!--<form action="/action_page.php"-->
    <form method="GET" action="ptt.php" target="iframe">
      <input type="text" name="keyword"><br>
      <input type="submit" value="SEARCH!">
      <input name="url" type="hidden" value="https://www.ptt.cc/bbs/HatePolitics/search?q=">
    </form>
    <h1>NEWS SEARCH:</h1>
    <h3>Keyword:</h3>
    <form method="GET" action="news.php" target="iframe">
      <input type="text" name="keyword">
      <div>
        <br>
        <input type="radio" name="media" value="TVBS" id="bm1"><label for="bm1" style="color: #4F86C6;">TVBS</label>
        <input type="radio" name="media" value="聯合新聞網" id="bm2"><label for="bm2" style="color: #4F86C6;">聯合新聞網</label>
        <br>
        <input type="radio" name="media" value="風傳媒" id="om1"><label for="om1" style="color: coral;">風傳媒</label>
        <input type="radio" name="media" value="蘋果日報" id="om2"><label for="om2" style="color: coral;">蘋果日報</label>
        <input type="radio" name="media" value="ETtoday" id="om3"><label for="om3" style="color: coral;">ETtoday</label>
        <br>
        <input type="radio" name="media" value="中央社" id="gm1"><label for="gm1" style="color: #88dba3;">中央社</label>
        <input type="radio" name="media" value="自由電子報" id="gm2"><label for="gm2" style="color: #88dba3;">自由電子報</label>
        <input type="radio" name="media" value="三立新聞網" id="gm3"><label for="gm3" style="color: #88dba3;">三立新聞網</label>
      </div>
      <br>
      <input type="submit" value="SEARCH!">
    </form>
  </div>
</section>
```

<form method="GET" ...>

在form中使用method="GET"來獲取後端所要呈現的資訊，
action="ptt.php"/action="news.php"表示所讀取的
資訊來源，

target="iframe"則是將獲取並要呈現的資訊顯示於下方
iframe中。



PTT 搜尋-ptt.php (1 / 4)

```
// store url and keyword
if(isset($_GET['keyword'])) {
    $keyword = $_GET['keyword'];
    $url = $_GET['url'];
}

// 設定url
if(isset($keyword)&&isset($url)) {
    $url.=$keyword;
    $pattern = '/<div\sclass\s=\s"title\s">\s*<a\shref\s=\s"(\S*)\s">(.*?)</a>\s*</div>/';
} else echo "NO !!!!!!!!!!!!!!!";
// echo $url."<br>";
```

存取GET的資料

先判斷GET的session值中有沒有存keyword和url的值，如果有的話再將值存入\$keyword和\$url這兩個變數中

存入輸入的關鍵字

先判斷有沒有存入\$keyword和\$url這兩個變數，如果有的話就將原網址加上想搜尋的\$keyword並以\$pattern（常規表示式）值擷取對應的資料，如果沒有存入\$keyword和\$pattern這兩個變數的話就顯示"NO!!!!!!!!!!!!!!"

**\$pattern: 擷取新聞標題、
連結網址**



PTT搜尋-ptt.php (2 / 4)

```
$ch = curl_init();

curl_setopt($ch,CURLOPT_SSL_VERIFYHOST,0);
curl_setopt($ch,CURLOPT_SSL_VERIFYPEER,0);
curl_setopt($ch, CURLOPT_CONNECTTIMEOUT, 10);
curl_setopt($ch, CURLOPT_URL, $url);
curl_setopt($ch, CURLOPT_FOLLOWLOCATION, true);
curl_setopt($ch, CURLOPT_COOKIE, 'over18=1');
curl_setopt($ch, CURLOPT_RETURNTRANSFER, true);

$data = curl_exec($ch);
curl_close($ch);
```

設定curl相關參數

`curl_init()` :

初始化curl

`curl_setopt()` :

設定網址 (用前面存的\$url變數)、
連結延遲等待時間、回傳等相關設
定

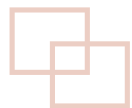
`curl_exec()` :

執行並將網頁原始碼存入\$data中

`curl_close()` :

關閉連結

CURLOPT_COOKIE=> "over18=1"
由於ptt有年齡上的瀏覽限制，所以要
先傳入yes(即數字1)至COOKIE才能進
入頁面爬蟲



PTT 搜尋-ptt.php (3 / 4)

```
// 找出 title & href  
preg_match_all($pattern,$data,$titles, PREG_PATTERN_ORDER);
```

找出需要的原始碼

利用preg_match_all這個function找出要擷取的原始碼，並存入\$titles陣列中
並用PREG_PATTERN_ORDER的排序方式存入陣列



PTT搜尋-ptt.php (4 / 4)

```
for($i=0; $i<=9; $i++) {  
    $href = $titles[1][$i];  
    $titles[1][$i] = "https://www.ptt.cc".$href;  
}
```

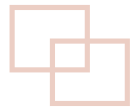
```
// echo top1~top10  
echo "<div style='color:#727a82; font-family:微軟正黑體;'>";  
for ($j=0; $j<=9; $j++) {  
    $rank = $j+1;  
    echo $rank.". ". "<a href='". $titles[1][$j]."' target='  
        _blank' style='color:#727a82; font-family:微軟正  
        黑體;'>". $titles[2][$j]. "</a><br>";  
}  
echo "</div>";
```

微調擷取的網址

由於ptt的文章連結是採用相對位置的連結，所以要將擷取到的網址前面再加上原網址

echo前十筆擷取資料

用for迴圈的方式將\$titles陣列中的擷取資料echo出來，並用html tag - <div>包住



新聞搜尋-news.php (1/5)

```
// store url and keyword
if(isset($_GET['keyword']) && isset($_GET['media'])) {
    $keyword = $_GET['keyword'];
    $media = $_GET['media'];
}
```

```
// 判斷media並設定對應的url
if(isset($media) && isset($keyword)) {
    switch ($media) {
        case '中央社':
            $url = "https://www.cna.com.tw/search/hysearchws
                .aspx?q=" . $keyword;
            $pattern = '/<li><a\shref=\\"(\S*)\\">.*?<div\sc
                lass=\\"listInfo\\"><h2>(.*?)<\h2>/';
            break;
        case '聯合新聞網':
            $url = "https://udn.com/search/result/2/" . $keywo
                rd;
            $pattern = '/<dt>\s*<a\shref=\\"(\S*)\\" \starget\
                =\_blank\\">\s*.*\s*<h2>(.*?)<\h2>/';
            break;
    }
}
```

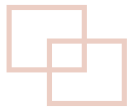
存取GET的資料

先判斷GET的session值中有沒有存keyword、media的值，如果有的話再將值存入\$keyword和\$media這兩個變數中

依照media值存對應的值

先判斷有沒有存入\$keyword和\$media這兩個變數，如果有的話就依照\$media存的值，存入相對應的\$url及\$pattern（常規表示式）值

\$pattern: 擷取新聞標題、連結網址



新聞搜尋-news.php (2/5)

```
case 'TVBS':
    $url = "https://news.tvbs.com.tw/news/searchresult/news?search_text=".$keyword;
    $pattern = '/<li><a\shref=\\"(\S*)\\" \sclass=\\"search_list_box\">.*?<div\sclass=\\"search_list_txt\">(.*?)<\div>/';

    break;
case '風傳媒':
    $url = "https://www.storm.mg/site-search/result?q=".$keyword;
    $pattern = '/<a\sclass=\\"card_link link_title\" \shref=\\"(\S*)\">\s*<p\sclass=\\"card_title\">(.*?)<\p>\s*<\a>/';

    break;
case '蘋果日報':
    $url = "https://tw.appledaily.com/search/result?querystrS=".$keyword;
    $pattern = '/<h2><a\shref=\\"(\S*)\" \s\target=\\"_blank\">\s*(.*?)\s*</';

    break;
case 'ETtoday':
```

依照media值存對應的值

先判斷有沒有存入\$keyword和\$media這兩個變數，如果有的話就依照\$media存的值，存入相對應的\$url及\$pattern(常規表示式)值

\$pattern: 擷取新聞標題、連結網址



新聞搜尋-news.php (3/5)

```
$url = "https://www.ettoday.net/news_search/do
      Search.php?keywords=".$keyword;
$pattern = '/<h2><a\shref\=\\'."(\S*)\".\s.*?>(.*?)
           <\a><\h2>/';

break;
case '自由電子報':
    $url = "https://news.ltn.com.tw/search?keyword=
          ".$keyword;
    $pattern = '/<a\ssclass\=\\'."tit\".\shref\=\\'."(\S*)\"
              \".*?>(.*?)<\a>/';

    break;
case '三立新聞網':
    $url = "https://www.setn.com/search.aspx?r=0&q=
          ".$keyword;
    $pattern = '/<div\ssclass\=\\'."newsimg-area-item-2
              \s\">\s*<a.*?href\=\\'."(.*?)\".\s.*\s*.*
              \s*.*\s*.*\s*.*\s*.*?>(.*?)<\div>/';

    break;
default:
    echo "media not found!!!!!!!!!!";
    break;
} // switch end
} else echo "NO MEDIAAAAAAAAAAAAAAAAA!!!!!!!!!!!!!!";
```

依照media值存對應的值

先判斷有沒有存入 \$keyword 和 \$media 這兩個變數，如果有的話就依照 \$media 存的值，存入相對應的 \$url 及 \$pattern (常規表示式) 值

如果都不是以上十個新聞網站的話就顯示 "media not found!!!!!!!!!!"

如果沒有存入 \$keyword 和 \$media 這兩個變數的話就顯示 "NO MEDIAA!!!"

**\$pattern: 擷取新聞標題、
連結網址**



新聞搜尋-news.php (4 / 5)

```
// CURL
$ch = curl_init();

// curl相關設定
curl_setopt_array($ch, [
    CURLOPT_URL => $url,
    CURLOPT_CONNECTTIMEOUT => 10,
    CURLOPT_RETURNTRANSFER => true,
]);
$data = curl_exec($ch);

// 關閉連接
curl_close($ch);
```

```
// 找出 title & href
preg_match_all($pattern,$data,$titles, PREG_PATTERN_ORDER);
```

設定curl相關參數

`curl_init()`:

初始化curl

`curl_setopt_array()`:

設定網址(用前面存的\$url變數)、連結延遲等待時間、回傳等相關設定存成一個陣列

`curl_exec()`:

執行並將網頁原始碼存入\$data中

`curl_close()`:

關閉連結

找出需要的原始碼

利用`preg_match_all`這個

function找出要擷取的原始碼，並存入\$titles陣列中

並用`PREG_PATTERN_ORDER`的排序方式存入陣列



新聞搜尋-news.php (5/5)

```
// 風傳媒&三立新聞網抓到的網址前面要加上原網址
if($media == '風傳媒'){
    for($i=0; $i<=9; $i++) {
        $href = $titles[1][$i];
        $titles[1][$i] = "https://www.storm.mg".$href;
    }
}else if($media == '三立新聞網') {
    for($i=0; $i<=9; $i++) {
        $href = $titles[1][$i];
        $titles[1][$i] = "https://www.setn.com/".$href;
    }
}
```

```
// echo top1~top10
echo "<div style='color:#727a82; font-family:微軟正黑體;'>";
for ($j=0; $j<=9; $j++) {
    $rank = $j+1;
    echo $rank.". ".<a href='".$titles[1][$j]."' target='_blank' style='color:#727a82; font-family:微軟正黑體;'>".$titles[2][$j].</a><br>";
}
echo "</div>";
```

微調擷取的網址

由於「風傳媒」及「三立新聞網」這兩個網站的文章連結是採用相對位置的連結，所以要將擷取到的網址前面再加上原網址

echo前十筆擷取資料

用for迴圈的方式將\$titles陣列中的擷取資料echo出來，並用html tag - <div>包住



Thanks