

## JobGuide

專題編號：106-CSIE-S021  
執行期限：105年第1學期至106年第1學期  
指導教授：王正豪 教授  
專題參與人員：103590021 林冠璋  
103590451 張倚瑄

### 一、摘要

本專題 JobGuide 是從批踢踢網站平台擷取關於各項工作的評論以及相關資訊，並存進後端資料庫中再加以分析，最後再以網頁方式呈現分析結果。

然而 JobGuide 跟現有的徵才網站相比，我們多分析了各區段的就業潮，以此可以看到台灣現狀以及科技發展的方向，可以用來預估下一個熱門產業。

其中分析出來的結果包括：此間公司的評價、平均薪資、就業潮...等等，另外也以圖表的方式呈現各個分析的結果。

**關鍵詞：**Mysql、Python、Spark、Tomcat。

### 二、緣由與目的

在這科技進步的時代，找工作不再只是透過報章雜誌的廣告，數以萬計的工作機會都可透過網路搜尋而得。而這一筆一筆的工作機會代表的不只是求職者的需求，更是大數據時代的重要資料來源。

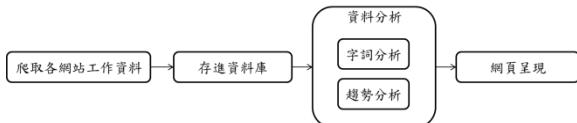
JobGuide 的緣由是因為在大三的時候發現找實習的困難性，所以想到或許可以試著做一個整理公司及工作機會的平台，但是在接著考慮細部內容時才驚覺，這些資料不該只是用來參考，更可利用大數據的特性來分析一些台灣的未來趨勢和現象。

### 三、研究報告內容

#### (一) 解決方法

先利用 Python 寫好的爬蟲系統，爬取批踢踢實業坊平台的工作資訊存進資料庫，再使用 Python 簡易演算法做各類分析，如公司評價會用字典分析等等；最後再以網頁方式呈現最後結

果，如圖一。



(圖一) 解決步驟

#### (二) 使用技術與方法

##### 1. 資料庫

在存取資料方面，我們使用 Mysql [1]來存取從各工作網站所擷取的資料，並透過 phpMyAdmin 以網頁介面來管理(新增、刪除)資料庫，如圖二。

##### 2. 爬文系統

在取得資料來源方面，我們使用 Python 語言撰寫爬蟲程式[2]，透過網頁的標籤頁將網頁內的資訊做切割，並把所需的資料屬性存進後端的 Mysql 資料庫。

##### 3. 分析資料

使用 Python MapReduce[3]做分析運算，如就業潮部分，用時間與各項工作去做分析，來觀看此刻哪一種職業為現在就業趨勢，另外如評價部分，會以字典法[4]來評斷；

##### 4. 網頁呈現

在網頁呈現方面，我們使用 HTML5、CSS、JavaScript。在 HTML5 以及 CSS 方面設計網頁版型，並加入 JS 來增加視覺效果，使得介面互動更為活潑。用

Webservice 方式連結資料庫，把已分析好的結果呈現在前端使用者介面。

## 四、實驗成果

分析結果可以透過網頁（圖三）的方式呈現，其中在導覽列部分會有：搜尋公司及職缺、近期薪資與就業熱潮、熱門職缺，三大Banner。

## ● 搜尋公司及職缺：

此部分會以使用者搜尋的公司或職缺；如搜尋公司，則會呈現該公司的評價、該公司的職缺以及對於該公司所有的評論；如搜尋職缺，則會呈現有哪間公司有此工作機會。

#### ● 近期薪資與就業熱潮：

在就業熱潮分析結果以及台灣各產業平均薪資都會個別以長條圖呈現，其中時間區段以一年為單位。

#### ● 就業領域：

其中平均薪資的部分，我們將產業類別分類成11大類來進行分析（參考1111網站工作分類[5]）。

## 五、結論

在搜尋公司頁面部分，網頁有搜尋欄可以輸入感興趣的公司。

在平均薪資和就業潮分析結果，我們將以長條圖的方式呈現，其中就業朝部分運算結果如圖四。

参考文献

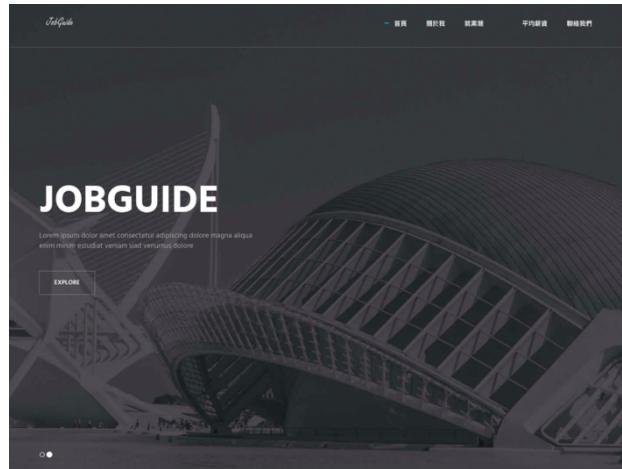
- [1] Mysql 資料庫使用  
<https://goo.gl/mtAUda>
  - [2] Python 網路爬蟲  
<https://goo.gl/rmw1Wc>
  - [3] MapReduce  
<https://goo.gl/83Gwkm>
  - [4] 李啟菁，「中文部落格文章之意見分

析」，2010/7，P.7 <https://goo.gl/hEtBQw>

## [5] 1111 網站工作分類

<https://goo.gl/SmmjRr>

(圖二)



(圖三)

```
17/06/20 08:37:30 INFO scheduler.Taskscheduler[IM
('12month', 410)
('11month', 570)
('10month', 506)
('09month', 490)
('08month', 497)
('07month', 522)
('06month', 516)
('05month', 519)
('04month', 502)
('03month', 549)
('02month', 508)
('01month', 411)
*****2016*****
```

(圖四)