

修課推薦系統

專題編號: 103-CSIE-S026

執行期限: 102 年第 1 學期至 103 年第 1 學期

指導教授: 郭忠義

專題參與人員: 100820330 劉昶甫

100820334 林純正

100820340 陳威仰

一、摘要：

修課推薦系統是一個能夠根據在校時選修的不同課程來推薦適合修習的課程，是一個大資料分析的應用，此系統不只能用於分析課程推薦，還可以用於其他商務之物品推薦。

關鍵字：大資料分析、推薦系統、網頁、分散式運算

二、緣由與目的：

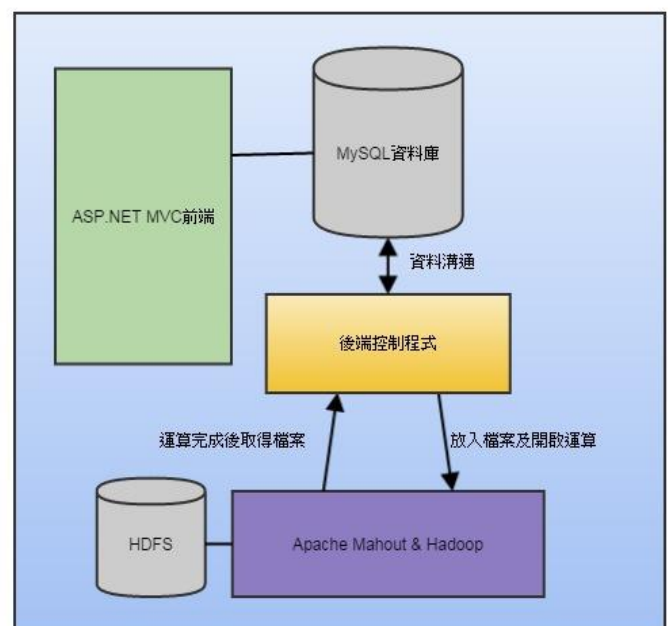
常常在面臨選課的時候，面對眾多課程時不會知道要如何選擇，所以我們就想到應用一般最常用於商務應用的推薦分析系統，轉而應用在選課上，於是就產生了這個修課推薦系統。

所以我們完成了一個可以處理大量資料並運行分析的推薦系統，不只是可以將修課資訊載入推薦，還可以應用於一般商務推薦運算，像是電影推薦、商品推薦等等。

三、專題架構：

此修課推薦系統架構大概可以分成四大部分: 分散式資料儲存系統(Hadoop 的 HDFS)、分散式運算系統(Hadoop 平台)、後端控制程式(含 Mahout 之視窗程式)、前

端系統(ASP.NET MVC 及 MySQL 資料庫)，主要是使用 java 語言撰寫，利用 Mahout 內建的分類分析演算法輔助資料的分析，使用 Ssh 與 Hadoop 連接，並放入 Hadoop 分散式平台上運算。而資料則要預先從資料庫上抓下來放入 HDFS 上以備的分析，而分析結果也需要經由後端控制程式放入 MySQL 中儲存，以方便前端網頁的存取，並透過一個簡易的網頁來查看資料庫中的資料(包含分析前與分析後之資料)。



資料分析種類:

(一)修課推薦分析:

使用已畢業及未畢業學生的相關修課資料，利用此資料來作分析，並推薦該使用者適合修習的課程。

(二)一般商務推薦:

使用的資料為電影評比資料，此資料是美國明尼蘇達大學利用開發的線上電影推薦系統所收集而來，利用 100 萬筆的推薦資料來驗證此系統的正確度及用於一般商務應用的可能性。

四、使用者操作及系統運作：

使用者透過在系統前端網頁登入帳號，可以查詢已修習的學分及課程，之後透過進入修課推薦頁面，查詢推薦列表，頁面會將使用者輸入的學號在資料庫中找尋事先分析好的結果，並顯示在頁面上供使用者瀏覽。

系統方面，系統在接收到使用者的學號後，會在資料庫中尋找該使用者之修課表單，由於推薦運算需要花費大量的時間，無法達到即時運算的功能，所以是透過事先進行的資料分析，輸出該使用者的修課資料及相關職場分析或修課推薦的結果。

五、相關背景與技術:

(一) ASP.NET MVC

ASP.NET MVC Framework 是微軟在 ASP.NET 中所添加的一組類別庫，能快速的利用此 Framework 完成 C# 網頁。

(二) MySQL

使用 MySQL 來當作資料儲存平台，用來暫時儲存分析前的模擬資料與儲存分析後的資料，主要用於前端網頁資料的直接連接。

(三) Apache Hadoop & HDFS

Hadoop 是一個應用分散式資料技術、提供巨量資料的分析技術及人工智慧巨量分析探勘技術的分散式作業環境，具備可擴充性，如果使用多節點處理

分散式的運算(MapReduce)，則可獲得更快的運算速度，而 HDFS 是 Hadoop 分散式的資料儲存空間，用來存放要分析的巨量資料。

(四) Apache Mahout

Mahout 主要是為了 machine learning 提供各種演算法的程式庫開放原始碼，裡面有包含許多非分散式及分散式的分析算法，而本專案則使用 Mahout 中提供的分析演算法完成資料推薦分析的部分。

六、結論：

修課推薦系統是一推薦系統之應用，但是由於我們無法取得學生對修課課程的評價，沒有評價的推薦運算準確度較低，而另一方面，資料量的多寡也會大大的影響推薦的正確性。所以未來可能可以利用教師評鑑中的課程推薦度來當作對評價，等具有足夠的資料量後，此系統將會有較大的可用性。

七、參考文獻：

[1] Mahout

<http://mahout.apache.org/>

[2] Hadoop

<http://hadoop.apache.org/>

[3] MySQL

<http://www.1keydata.com/tw/sql/sql.html>

[5] James Chen, Apache Mahout 於電子商務的應用

<http://www.slideshare.net/chaoyu0513/hit20130928-apache-mahout>

[6] 粉丝日志

<http://blog.fens.me/>