

以網站日誌分析系所網站對招生之影響

陳建伯	陳雯玉	郭文曲	黃立行
銘傳大學電腦與通訊學 系	臺灣科技大學資訊管理 系	銘傳大學電腦與通訊學 系	大同大學資訊工程所
jbchen@mail.mcu.edu.tw	wychen@mail.mcu.edu.tw	wckuo@mail.mcu.edu.tw	lshuang@mail.mcu.edu.tw

摘要

目前台灣的高教競爭激烈，各校為了招生無所不用其極。然而對於高中生或是家長來說，要了解一個系所的發展方向及特色，最直接的方法應該就是透過系所的網站。本文主要的目的，是利用系所網站的存取日誌檔，來觀察高中生在決定報考系所前，是否會有大量的網頁存取。經過分析後可以發現，在三月份繁星以及個人申請時，網站的存取數量的確有明顯的增加。同時我們也觀察到系所的招生名額，和網站的存取數量有顯著的關係，也就是招生名額越多的系所，網站的存取數量越大。另外我們也發現，各系所網站被下載檔案，『課程架構』明顯高於其他的檔案，由此可知，高中生及家長最關心的是未來四年所要修習的課程資訊。

關鍵字：網站日誌；系所招生

The Impact of Website Log Analysis on the Recruitment

Abstract

At present, Taiwan's higher education competition is fierce, and colleges are doing everything they can to recruit students. However, for high school students or parents, the most direct way to understand the direction and characteristics of a department is through the website of the department. The main purpose of this paper is to use the access log file of the website of the department to observe whether there will be a large number of web accesses before high school students decide to apply for the department. After analysis, it can be found that the number of accesses to the website did increase significantly in March when the Star Plan and individual application. At the same time, we also observed that the number of recruitment quotas in the department has a significant relationship with the number of accesses to the website, that is, the more recruitment quotas, the greater the number of websites accessed. In addition, we also found that the "course structure" was downloaded significantly higher than other files. It can be seen that high school students and parents are most concerned about the course information to be studied in the next four years..

Keywords: *website log, recruitment*

1. 前言

現今數位化時代快速發展，數位資料也呈現爆炸性的成長，各類底層大數據平台提供多種類的存儲方式，然而對於企業學校來說，儲存大量數據與高速計算效能僅僅是滿足企業學校進行數據分析的硬體基礎，如何發掘這些海量的數據產生應用價值，走好數據分析這最後一哩路，引導企業學校進行戰略決策卻是至關重要的一步。

然而，目前高等教育面臨挑戰，競爭激烈，因此想藉由分析點閱本校各系網站的流量、來源 IP 位址、參觀網站時間、參觀網頁、訪客作業系統、使用的瀏覽器、連結網站的來源、連結網站的關鍵字…等，從不同面向，觀察各種統計報表，讓各系所得以了解網路使用者瀏覽本校各系所網站的流量、以及在該網站的存取行為，觀察主要點選何種資訊檔案，證明好的網站設計內容對於吸引使用者目光是重要的，同時期望藉由改善網站內容、更新資訊及設計方向，進而提升本校的形象、提升曝光率，藉以吸引各國學生來本校就讀。

本文利用數據分析的方法，從網站上收集資訊與日誌(log)進行多面向分析，從不同系所相互比對觀察，比較系所招生人數與瀏覽數的關係，並從每月、每日甚至是到每小時，觀察流量點擊數，以及每一來源位址為何、統計每次停留時間，以做為改進網站之依據。

2. 文獻探討

由於本研究主要之目的是分析評估網站流量、網站日誌檔並找出相互之間關聯性，進而優化網站，提高知名度。因此，於本節中將過去的相關文獻做整理與歸納，以發展出本研究的主要架構與研究方法。

2.1 網站計量分析

網站分析 (Web Analytics) 是一種網站訪客行為的研究。於商務應用背景來說，網站分析特別指的是來自某網站蒐集來的資料，利用分析、報告，以決定網站佈局是否符合該網頁要達成之目標。除此之外，衡量網路流量的過程，還可以用作商業和市場研究的工具，以及評估和提高網站的有效性。[7]

網站分析可以被定義為各種數據的評估工具，包括網站流量、以網站為基礎的交易、網站伺服器績效、可用性研究、用戶提交的資訊以及幫助網站經營者瞭解線上瀏覽者的經驗。有兩種主要方法可以用來採集分析的資訊：網頁頁籤(Page tagging)以及利用網站伺服器日誌檔(Web server log files) [1][4][5][6]。

頁面標籤(Page tagging)，或稱網蟲(Web bug)，是透過識別標籤在一個網站或多個網站中進行，主要是瀏覽者數量到監控整個瀏覽者對網頁的行為，這些都有免費的標籤服

務可用，當這個網頁有被訪客(Visitor)所使用，頁面識別標籤和資訊就會被傳送到計數軟體(Counter software)，以收集這些資料供日後分析，例如：網頁計數器以及cookie的運用[1][3][4]。90年代中期，基本上網頁計數器就包含在網頁中的圖像，顯示有多少次該圖像被讀取過，而圖像內容就是顯示大約該頁被拜訪的次數。90年代晚期，這種做法延伸至微小不可見圖像，而且改由透過 JavaScript 等方式請求關於網頁和訪客的訊息被傳遞到其他的第三方伺服器端，而這些資料隨後被網站分析公司處理並產生詳細的統計資料。[7]

另外一種是使用網頁伺服器的日誌檔(log files)，在1990年代之後web技術興起，網站日誌檔 (web log files) 成為另一種新型態的電腦日誌檔，記錄著訪客透過瀏覽器對網站伺服器所提出的請求 (requests)，以及與伺服器互動的行為歷程，可以提供許多珍貴的資訊，有助於網站行銷、發展與管理，同時也可藉以瞭解使用者行為、特質等。它收集大量訪客的事件，如點擊數 (hits)、網頁檢視 (page views)、網站停留時間 (time on site) 等。[7]藉由各種分析工具或是統計方法，更精確地瞭解網站營運成效及網站使用狀況，作為網站內容改善及掌握使用者動態的依據。

Hamid等人[2]曾從日誌分析應用研究的觀點中，提出日誌分析具有以下優點：(1) 日誌檔的資料是未經過濾且為系統自動收集，因此記錄資料較為客觀；(2) 日誌檔的資料都是非強制性的，逐步記錄使用者真實的動作，不同於填答問卷的調查法，使用者可能會因為當下的心情選擇、改變或隱藏他們心中真正的想法及行為；(3) 利用日誌檔分析可較問卷調查法或由訪談的可信度較為高；(4) 日誌分析是一項既有效又真實收集歷程資料的方法；(5) 日誌分析是一項有效的假設驗證方法；(6) 日誌分析是一項有效評估系統績效與效能的方法。以上都說明日誌的確可以從中得到一些脈絡與證據。

2.2 網站分析工具

目前市面上存在著各式各樣分析網站日誌的工具，主要以最有效方式取得有用的資訊。選擇適合的分析工具也是一大難題，本文就現在市面上的功能進行比較，以針對不同的議題選擇適合的工具軟體，降低分析過程中繁雜的處理工作。一般來說，Web日誌分析軟體基本上都可以有訪問次數和訪問者人數、停留時間和造訪時間、不同時間點的瀏覽數、訪客所在國家、頁面的瀏覽量、最多瀏覽的頁面、下載檔案型別、使用者的裝置、使用者的瀏覽器、搜索的關鍵字、HTTP錯誤等。本文以市面上較為著名的軟體中評比，詳細資料如表1。

表1 Web日誌分析軟體比較

軟體名稱	供應商	許可證	使用語言	支援平台	報表格式
Google Analytics	Google Inc.	Free	Java Script	Windows/Mac/ Linux/Solaris	HTM/PDF/ CSV
Web Log	Alentum	Proprietary	In-built	Windows	HTML/PDF/

Expert	Software				CSV
AWStats	AWStats Inc.	GNU-GPL	Perl	Windows	HTML/PDF
Piwik	Matomo.org	GNU-GPL	Python/Ruby	Windows/Mac/ Linux/Solaris	HTML/PDF
GoAccess	MIT Licensed	Open Source	C	Unix/Linux/ Mac	HTML/CSV/ JSON
Open Web Analytics	Open web analytics	GNU-GPL	PHP	Windows	HTML
Webalizer	Webalizer	GNU-GPL	C	Unix/Linux/ Mac	HTML
W3Perl	W3Perl.com	GNU-GPL	Perl	Windows/ Unix/Linux/ Mac	HTML
Analog	Community Development	GNU-GPL	C	Windows/ Unix/Linux/ Mac	HTML

(註：GNU General Public License，簡稱 GNU-GPL，指使用免費軟體的許可證，保證終端用戶可以自由運行、學習、共享和修改軟體。)

本文選擇使用 AWStats 作為分析日誌之輔助軟體，該軟體主要是一個可生成網路流量、FTP 或郵件伺服器統計圖的免費開源軟體，使用 Perl 語言撰寫，擁有跨平台使用的優點。

3. 研究方法

日誌是由系統自動產生的數據，用於存儲有關操作系統內發生事件的訊息，存儲每個時間的數據，都可以以數據透視表或文件的形式呈現。在日誌文件裡，記錄是按時間排列，因此，日誌分析是從日誌文件中提取有價值訊息的任務。本節將敘述處理日誌分析的步驟與架構。

3.1 資料收集與清理

數據分析第一步最重要就是收集數據，在來源數據中收集日誌 log，這些來源包含如計算機系統、代理伺服器、網頁日誌數據等，不同的伺服器裡也存在著不同的日誌，

例如 Web 伺服器日誌、應用程序日誌、系統日誌等等。本研究即收集本校各系所網站的相關日誌，統計的數據如點擊數 (hits)、網頁檢視 (page views)、網站停留時間 (time on site)、下載檔案類型 (download file type) 等。

一旦正確安排了日誌數據，就必須執行數據清理過程，這是因為可能存在損壞的日誌數據，這會造成分析結果的錯誤。而導致數據損壞的原因可能包含：存儲日誌數據的磁碟損毀、應用程序異常終止、輸入/輸出時的干擾、系統中存在病毒等等。

3.2 資料結構化

日誌數據既龐大又複雜，因此，日誌數據的呈現方式將直接影響使用者與其他數據相關聯的能力與觀察。日誌的觀察因從不同的日誌中藉由深入調查中獲得，因此相同意義的數據應遵行相同的命名規則，以及相同的資料結構，以利後續分析。

3.3 資料分析與獲得結果

最後，就是分析日誌數據的結構化形式。這一步驟可以透過各種方法來執行，例如建立模型、標準化、統計、使用機器學習的分類、相關分析等等。圖 1 為日誌分析的流程架構圖。



圖 1 日誌分析流程圖

4. 實驗結果

以目前台灣高教的多元入學管道趨勢來看，每年三月的個人申請以及繁星推薦入學占錄取人數比例越來越高。以個人申請招生名額來看，目前規則是一名學生可以申請六個校系，而且根據大學招生委員會統計，自 103 年以來招生人數年年增加，從 103 年招生 49733 人，104 年 54322 人，105 年大增 1410 人達 55732 人，106 年更持續到達 55954 人，107 年則略減為 56060 人，但率取率高達有 81.36%，可見此時是高中生根據一月考試的學測成績尋找各大學各科系的錄取條

件以及該科系大學四年需要學習的課程內容、課程架構、各系所老師等資訊。因此，為吸引更多高中生來校就讀，提供清楚、正確的資訊、視覺化網頁、容易找尋相關資料的各系所網頁在此時就呈現出非常重要的地位和角色。觀察其中某些學系網站的流量數，從 2018 年一月到至今，發現三月皆為最高峰。如圖 2 系所網頁流量總數所示。

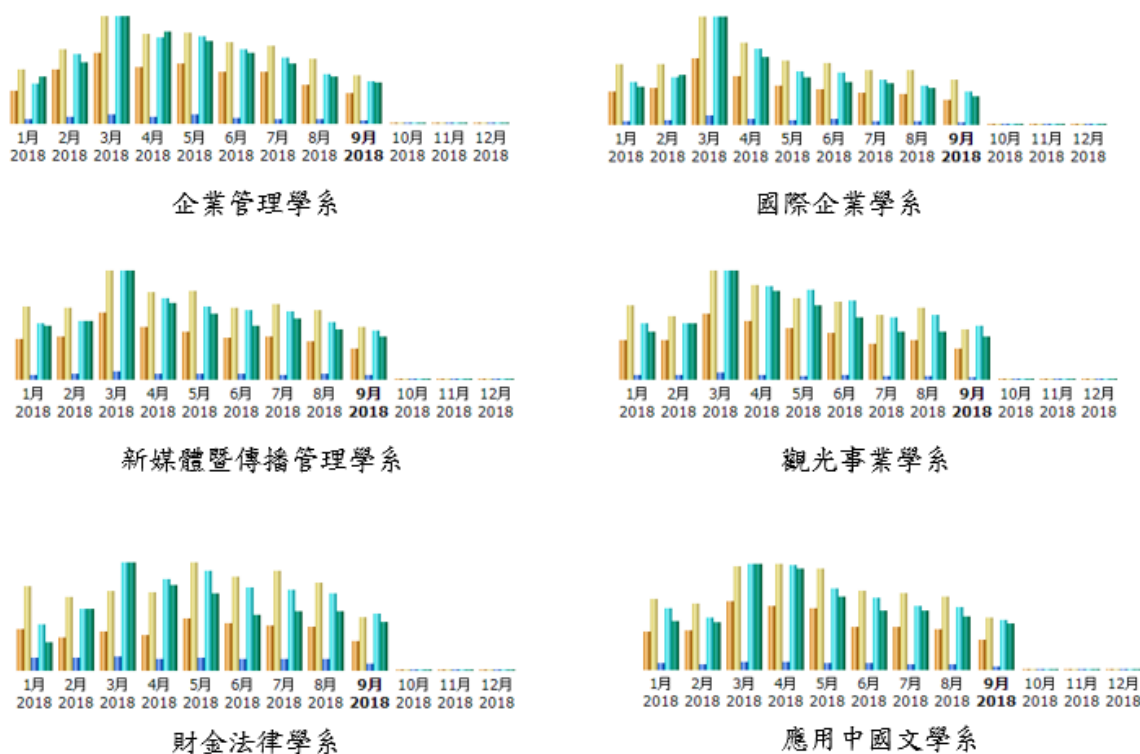


圖 2 各系所網頁流量總數

從圖 3 中發現以每年三月繁星與個人申請的入學管道，其中各系所的最低錄取分數、各系所招收的人數和 3 月最高峰的參觀人數對照結果，發現招生人數和參觀人數的數據曲線相互差距不大，表示是有相關的，但是很明顯發現最低錄取分數與招生人數和參觀人數是沒有相關的。

圖 4 則是大學入學的另一管道，每年 7 月考試完後得到的成績填寫志願，因此本研究再觀察 7 月參觀人數、與招生人數和錄取分數之間是否有相關。從圖中發現與圖 3 的結果類似，發現招生人數和參觀人數的曲線依舊相互差距不大，而與最低錄取分數則是一樣沒有任何太大的關係。

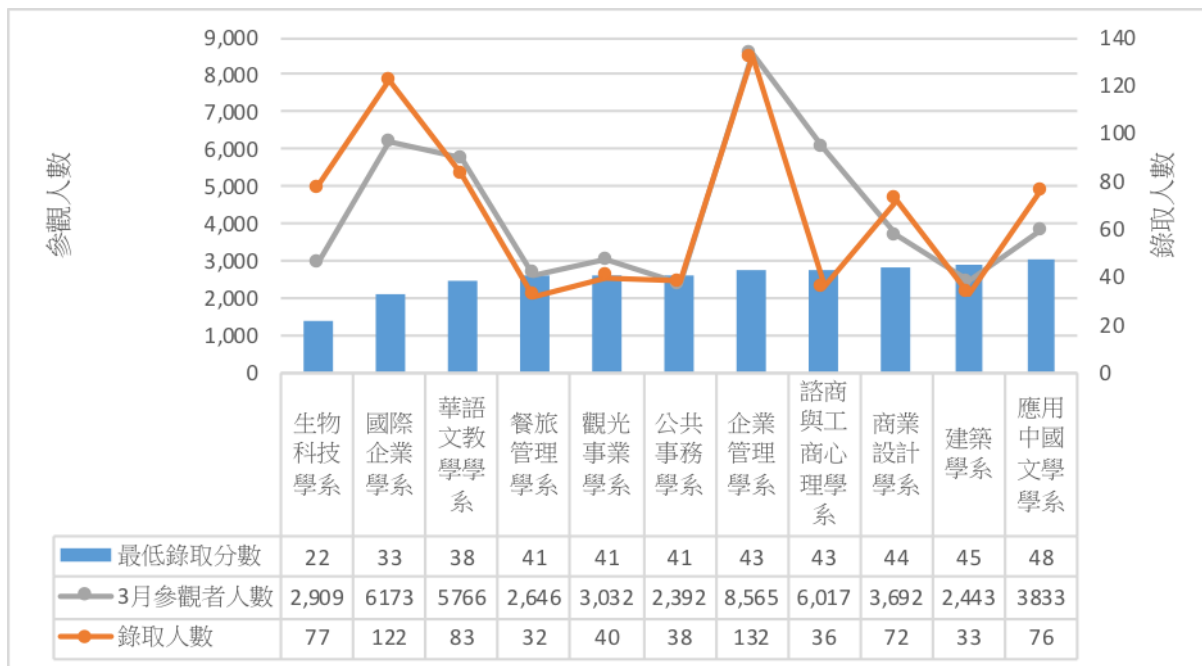


圖 3 各系所 3 月參觀人數、錄取人數與錄取分數之對應

圖 3 與圖 4 中發現諮商與工商心理學系的參觀人數與錄取人數同樣呈現較大的差距，有可能該系招收新生人數雖然不多，但該科系在目前大專院校中較無該類似的科系，大部分都是心理系或是應用心理系，才造成較多的參觀人數。

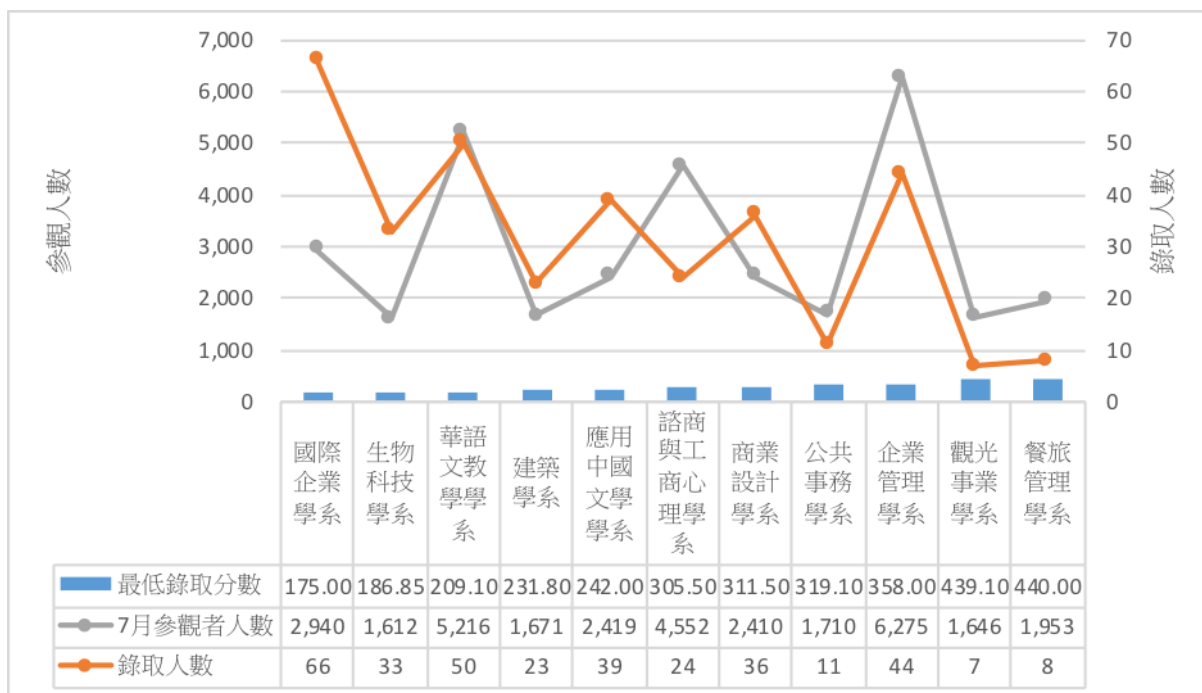


圖 4 各系所 7 月參觀人數、錄取人數與錄取分數之對應

從圖 3 的 11 個學系當中，其中的參觀人數最高者依序為企業管理學系、國際企業學系、諮商與工商心理學系、華語文教學學系、應用中國文學學系與商業設計學系，同樣的利用 3 月高峰期的流量觀察該月大部分的使用者都在各系網站中下載何種資訊，發現各系前 6 項下載檔案數及內容如表 2、表 3 與表 4。

表 2 企業管理學系與國際企業學系 3 月下載檔案統計

企業管理學系		國際企業學系	
下載項目	下載總數	下載項目	下載總數
課程架構	1019	課程架構	1242
學分學程課程架構	159	學分學程課程架構	327
必選修科目表	149	大學入學申請分則	295
新聞：產學合作	111	面試流程與問題示例	267
新聞：推動數位行銷	110	職涯地圖	261
面試流程與問題示例	107	個人申請評定標準	258

表 3 諮商與工商心理學系與華語文教學學系 3 月下載檔案統計

諮商與工商心理學系		華語文教學學系	
下載項目	下載總數	下載項目	下載總數
課程架構	394	課程架構	493
學生證使用注意事項	181	學分學程課程架構	123
各班課表	180	華教學報	67
心理學參考知識	132	論文寫作格式	57
修讀輔系科目及辦法	118	實習實施細則	51
必選修科目表	110	實習報告	33

表 4 應用中國文學學系與商業設計學系 3 月下載檔案統計

應用中國文學學系		商業設計學系	
下載項目	下載總數	下載項目	下載總數
課程架構	380	課程架構	424
讀書心得徵文比賽	195	實習流程圖	267
職涯地圖	141	面試準備大解惑	263
實習實施細則	112	系所簡介	99
論文格式	48	面試注意事項	94
學生實習懶人包	33	輔系科目	79

由以上各表展現各學系下載檔案的類型來看，各系最重要就是課程架構，也就是大學四年應該修習的那些科目，以及系上所開的學分學程課程有哪些之外，最關係到高中生則是有關面試相關資料，下載數大約也是二百多人次。由於各系所放置的項目並無規定，所以像是諮商與工商心理學系、華語文教學學系以及應用中國文學學系就無面試相

關選項。再者，假設系上有實習課程，該系上之相關實習檔案下載次數也非常多。總觀來看，除了各系放置基本資料之外，各系的特色、專業知識以及新聞似乎也相當受到使用者的喜愛，下載次數也不少，像是企業管理學系的 2 則新聞、應用中國文學學系徵文比賽以及華語文教學學系的學報，說明各系放置特色訊息與資料，對使用者來說會對該系更有深一步得了解。

5. 結論

從台灣高教的入學管道時間，觀察 3 月和 7 月的網路流量與參觀人數，發現有些許的相關性，即各系錄取人數與參觀人數有相關性，且各網站放置的訊息越多越豐富，的確可以讓使用者快速得到他所需要的資訊。因此，好的網站提供清楚、最新資訊是必要的，可以吸引使用者，提升該系形象，且現在網頁資訊豐富，觀察使用者於各系網站每一頁其實停留不超過 30 秒為最多，因此提供清楚路徑，快速找尋到目標資訊是重要的，建議網頁架構不宜太多層。

由於現在每一系所的規定項目並不是一定，因此較難全面性的比較下載檔案次數，只能根據目前狀況歸類出大方向，未來還可以增加來源區域，分析使用者的來源地點，增加分析的有意義性。

6. 參考文獻

- [1] Croll, A., & Power, C. (2009), *Complete web monitoring*. Sebastopol: O'Reilly Media Inc.
- [2] Hamid R Jamali, David Nichlas and Paul Huntington. (2005), *The use and users of scholarly e-journals: a review of log analysis studies*. Aslib Proceedings: New Information Perspectives, 57(6), 554-571
- [3] Jansen, B. J. (2006), *Search log analytics: What it is, what's been done, how to do it*. Library and Information Science Research, 28, 407-432
- [4] Kaushik, A. (2010), *Web Analytics 2.0*. Indianapolis: Wiley Publishing Inc
- [5] Peterson, E. T. (2004), *Web analytics demystified. A maketer's guide to understanding how your web site affects your business*. Celilo Group Media and CafePress.
- [6] Peterson, E. T. (2006), *The big book of key performance indicators*. Portland/Boston/Madrid: Web Analytics Demystified.
- [7] Wikipedia , https://en.wikipedia.org/wiki/Web_analytics