

# 台灣觀光資訊入口網站瀏覽行為分析

王佳鳳<sup>1</sup>

洪惟泉<sup>2</sup>

休閒事業管理系講師<sup>1</sup> 休閒事業管理系副教授<sup>2</sup>

## 摘要

台灣觀光資訊入口網站 (<http://taiwan.net.tw>) 每天約有一萬人次上網，在這無形中已經儲存著數以兆計由 0 與 1 所組成的磁性記憶與資料，我們運用資料採礦技術嘗試將龐大資料轉化成寶貴的資訊以供相關決策人員參考。

本研究時間為 2008/2/1 到 2008/4/30 共三個月時間，針對台灣觀光資訊入口網站之網站日誌檔 (log files) 透過 WEB TREND、SPSS、SPSS CLEMENTINE 等軟體工具，對其作相關瀏覽行為之分析。

主要分析結果為：一、網路使用方面：每人上網瀏覽頁數平均為四頁，且停留時間以一分鐘內占最多人數，上網時段以平日人數較多於週末，一天內平均有三階段上網人潮，分別為早上九點到十一點、下午三點到五點、晚上八點到十一點。二、交叉分析方面：點選時間與旅遊型態有顯著差異。三、關聯分析方面：不同旅遊景點類型其相關點選的旅遊景點不同。本研究最後並提出相關建議及策略供產、官、學界參考。

**關鍵字：**資料採礦、關聯分析、網站日誌檔、台灣觀光資訊入口網站

## The Analysis of the Browsing Behavior in the Tourism Information Portal of Taiwan by Applying Data Mining Technology

Peggy Wang<sup>1</sup> Wei-chuan Howng<sup>2</sup>

Instructor, Department of Leisure Business Management<sup>1</sup>

Associate professor, Department of Leisure Business Management<sup>2</sup>

## Abstract

There are at least 10,000 people browsing the Tourism information Portal of Taiwan (<http://taiwan.net.tw/>) and billions of the magnetic memory and data have already being stored, which are composed of 0 and 1. We provide references, which we transfer the great amounts of data into valuable information via Web mining technique to the decisive members.

The study period is three months—from 2008/2/1 to 2008/4/30, and it focuses on the log files of Tourism information Portal of Taiwan, which was analyzed via WEB TREND、SPSS、SPSS CLEMENTINE.

The results are as follows:

1. Internet usage: Averagely, each person browses four pages each day, and most people's browsing time is within one minute. There are more people surfing the Internet during weekdays, and there are least people surfing the Internet on weekends. There are 3 periods of the Internet—the first period is from 9: 00 am to 11:00 am; the second period is from 3: 00 pm to 5:00 pm; the third period is from 8: 00 pm to 11:00 pm.
2. Cross-analysis: the click time and types of traveling show great differences.
3. Connected analysis: The types of different tourists attractions are also different from click tourists attractions.

We offer the suggestions and strategies according the above results to the industry, the administrator, and the academic.

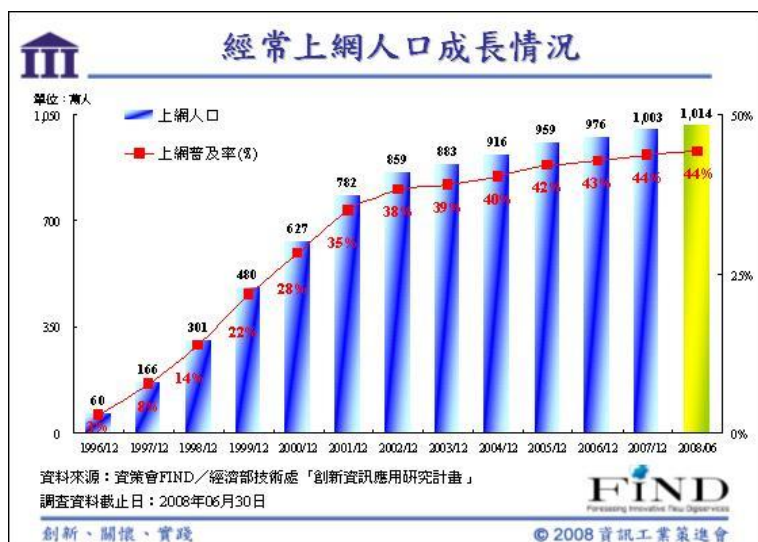
**Key Words :** web mining , [relational analysis](#) , [web side log files](#), the Tourism information Portal of Taiwan

:

## 壹、緒論

### 一、研究緣起

由於網際網路與通訊科技迅速發展，經常上網人口數不斷增加，根據資策會 ACI-FIND 最新調查數據顯示，截至 2008 年 6 月底為止，我國經常上網人口[達 1,014 萬人，網際網路連網應用普及率為 44%(如圖一)，且有逐步爬升的現象。這些數據顯示出網路上無時無刻都有著大量的資料在進行傳輸。在這網際網路資訊的爆發中，已經儲存著數以兆計由 0 與 1 所組成的磁性記憶與資料。這些網路上隱藏的資料透過資料採礦 (Web Mining) 幫助決策判斷，將龐大資料採礦出有意義、可利用的知識，加以應用。



圖一、經常上網人口成長情況 資料來源 2008 資訊工業策進會

本研究擬利用網路採礦技術與統計方法，以觀光局網頁為例，進行實證研究。有鑑於政府積極推展觀光，有意在資訊網路方面發展與投資，而台灣觀光資訊網站每天約有一萬人上網站瀏覽，意圖找尋些相關的旅遊資訊，若能了解這些潛在的客群，解析瀏覽者對各旅遊景點偏好，並將資訊科技，應用在觀光產業發展，利用電腦技術尋找出可用資訊，以提供政府及各單位團體規劃決策之參考。

## 二、研究目的

由於此網站提供多種旅遊產品及旅遊資訊內容，供網路使用者瀏覽與搜尋的旅遊入口網站，因此將探討不同的瀏覽者在不同階段之需求；以及針對不同類型旅遊之網頁設計是否應符合不同瀏覽者需求提出相關經營方針。因此，本研究主要研究目的如下：

- (一) 針對主要上觀光局網站的瀏覽者作瀏覽行為分析。
- (二) 針對時間作瀏覽者行為分析。
- (三) 針對主要產品URL之位置作瀏覽分析。

## 貳、研究方法

### 一、研究範圍與對象

本研究以國內交通部觀光局網站之中文的觀光入口網頁為研究範圍，研究對象為上此網頁的不同瀏覽者。（觀光局網站：<http://taiwan.net.tw/>）

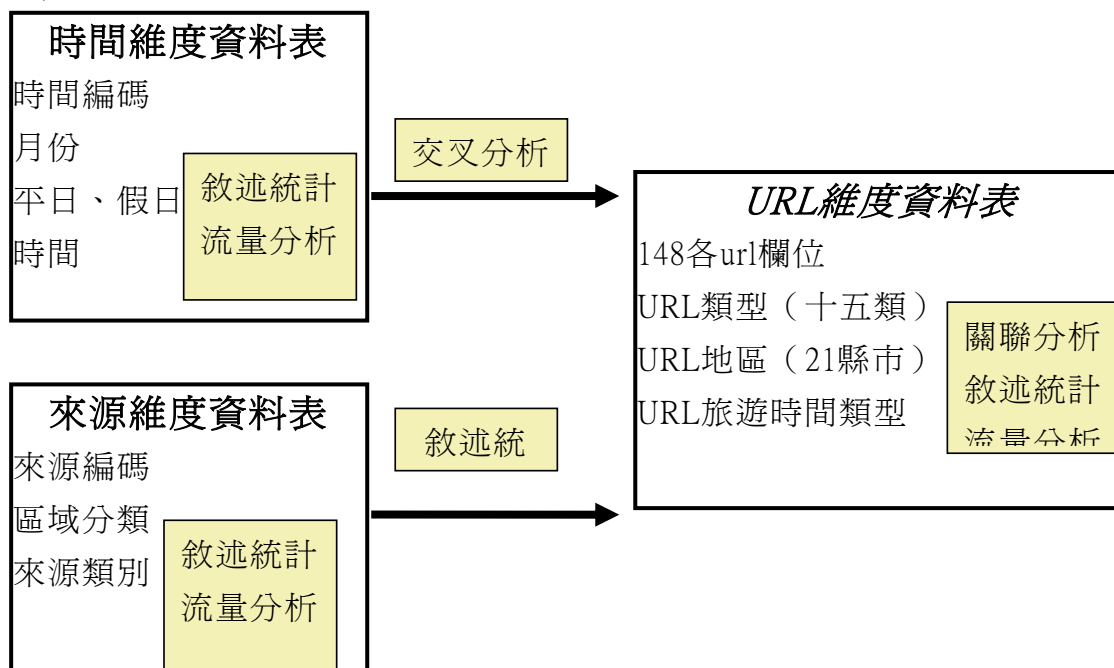
本研究透過觀光局協助取得 2008 年 2 月 1 日到 2008 年 4 月 30 日 IIS 格式的 LOG 檔案；其資料容量約有 20G，每一天的的檔案在不壓縮的情況下介於 130MB 至 200MB 之間，且一天約有十萬筆資料。

第一部份：將運用此 20G 資料，套用於 WEB TREND 軟體作瀏覽行為分析。

第二部分：擷取觀光資訊入口網站，中文版網頁之相關旅遊景點、型態之 URL 欄位，並運用 PERL 程式語言，將不屬於旅遊相關的欄位資料篩除，僅留下觀光局網站上第一層與第二層的旅遊景點位置，共擷取一百四十八個 URL 欄位的瀏覽資料，其檔案每天約一千筆資料，三個月合計有 121,864 筆資料。

## 二、研究架構

對於本研究目的：一、針對主要的瀏覽者作瀏覽行為分析、二、針對時間作瀏覽者行為分析、三、針對主要產品 URL 之位置作瀏覽分析，提出以下研究架構(參閱圖二)。



圖二、研究架構

## 三、研究問題探討

(一) 透過 WEB TREND 程式分析探討問題如下:

1. 網頁被讀取的次數；
2. 使用者來訪次數；
3. 使用者來訪停留時間；
4. 一日內各時段及一週內各日使用情形；
5. 來源網域作一整理分析；
6. 最常被瀏覽的網頁排行。

(二) 透過 data mining 觀念，所探討的問題如下：

- 『研究假設一、不同類型的旅遊型態及點選時間有顯著差異』
- 『研究假設二、主題遊樂園活力型感興趣之景點類型』
- 『研究假設三、求知性型者感興趣的旅遊類型』
- 『研究假設四、冒險有活力者感興趣的旅遊類型』
- 『研究假設五、商務旅遊者感興趣之景點類型』
- 『研究假設六、較高消費族群感興趣之景點類型』
- 『研究假設七、不同縣市對於高消費的喜好之來源地為何』

## 四、資料分析工具

本研究使用”WEB TREND 公司”所發行的”LOG ANALYZER”分析軟體，同時也使用了 PERL 程式語言，最後並利用 SPSS Clementine 應用軟體作關聯分析等。（參閱表一）

表一、資料分析工具

步 驟	方 法	用 途
清除、整理、轉化	PERL 程式語言	將純文字檔之 logs file 轉成以點分隔資料 (Csv)
資料庫、清除、整合	Excel 資料庫	僅留下有點選一百四十八個 URL 的瀏覽者
實作	WEB TREND	將所有 logs file 輸入之作流量分析。
	SPSS	卡方分析、檢定星期與 URL
	SPSS Clementine	關聯分析、探討各旅遊景點偏好

## 參、網路記錄檔案分析與發現

透過WEB TREND分析結果，本研究得到以下結論：

### 一、使用者行爲

觀光局資訊入口網站從使用者三個月之每天平均的瀏覽人數有11,803、13,214、13,548人次；瀏覽頁數為1,258,612、1,201,680、1,438,966頁，平均每人瀏覽頁數約4頁，平均停留時間也在1分鐘之內為多。

### 二、瀏覽時段

從每個月瀏覽的趨勢看來，使用網站的人數，集中於工作日，假日的使用量明顯較低，而每天除了用餐與睡覺時段外，有三個高峰是人潮擁擠的時段，即早上九點～中午，下午兩點～五點，晚上八點後到十一點，三個人多的時段。

### 三、網頁排名

為得知觀光局資訊入口網站中，瀏覽者僅觀光相關景點點選率作探討，我們在觀光局資訊入口網站中篩選了一百四十八個旅遊景點之作排名，得知前十名多屬於網頁第一層位置的欄位，但某些網頁，如、主題樂園不是在建議行程當中，一般旅館、國際觀光旅館、一般觀光旅館也不是在首頁兩旁的建議欄位中，但這些頁面的點選率卻能勝過觀光局現有推薦行程之欄位，故建議當局可加以瞭解瀏覽者需求作網頁上的改進。

其次是觀察到其他受歡迎前十名的旅遊景點依序為，溪頭、日月潭生態二日遊、台北都會一日遊、烏來溫泉一日遊、台北士林夜市、東區商圈、烏來小吃三日遊、南

台灣二日遊、台北都會二日遊、台中美食二日遊、陽明山溫泉二日遊、北埔、內灣戀戀風情一日、北海岸一日遊、建議這些旅遊熱門的旅遊景點可以增加在網頁熱門景點中，讓瀏覽者更容易找到所需資訊。

#### 四、單一網頁元素之錯誤比例

在這三個月內，網頁上任一元素出現問題的機率，平均有3.5%，對於一個國家之入口網站，在品質的要求下，必須要再加強此網站的管理與維護。相信降低網頁元素錯誤的比例，民眾對觀光局資訊入口網站才會深具信心，外國人士對我國觀光資訊入口網站也會更滿意。

### 肆、結論與建議

#### 一、結論

針對資料採礦技術分析結果，本研究得到以下結論：

##### (一)交叉分析

觀光遊憩參與者，需要非常重視觀光遊憩產品的包裝與設計，最主要是要在離峰時段創造需求。如星期一、二時，在網路上可以與旅行社合作，提出相較便宜的交通票價，或住宿折扣訊息等，做特殊的目標行銷廣告。星期四、星期五時則可以加強「都會型」景點的配套方案，如台北都會行，配上當季的美食展覽等，虛實合一的網站才會達到乘數效應。

##### (二)關聯分析

透過關聯式法則的分析後，目的是以主題性的探討方式從大量瀏覽記錄中找出了瀏覽之間的相關聯性，看似兩個獨立的行程或景點，經由資料採礦的處理後，得以結合在一起，讓我們得到了一些寶貴的資訊並提供給網站管理者作參考。

遊樂園因其趣味性高容易吸引各地遊客前往，資料採礦過程發現瀏覽者在瀏覽主題樂園資訊的過程較多點選北邊的遊憩點選，旅遊時間設定為二日，瀏覽過程亦搭配溫泉與美食資訊一併了解。故宮博物院位於外雙溪，瀏覽者在安排計畫過程也傾向連結附近士林的夜市與北投溫泉等都會一日遊的行程，並未因為此求知性較強的故宮之旅而選擇同質性較高的文化活動。而秀姑巒溪泛舟有季節與刺激性的考量，其關聯分析結果偏向同質的戶外型活動，如浮潛，其他的瀏覽途徑仍傾向與地區性的旅遊資訊結合。而商務人士點選的會議展覽，多為公事所需，多來自四面八方的人士，故也較重視住宿的品質，所以在會議展覽的網頁上應提供相關住宿資訊，以解決商務人士擔憂的住宿問題。除了住宿之外，也因為會議展覽常在交通方便的大都市舉行，故可推薦商務人士的相關行程以傾向於區域性的活動為主。國際觀光旅館、航空為高消費的旅遊型態，可以配合高消費的活動類型如主題樂園，而同質性的住宿方面也是此高消費人士所關切的，可以加以配套。

## 二、建議

### (一)針對觀光局的相關建議：

從分析結果看來，若網頁下方提供推薦行程者，引導率較高，若能在每個頁面下都有相關建議行程，將可達到更人性化的推薦機制，並提升與瀏覽者間的互動。

- 1.網站可多加行銷不同類型的景點，因為由景點的排名看來，很多景點都欠缺人次瀏覽，使其瀏覽數量差距盛大，相對的在實體上的遊客人數也會差距大，若能分散大眾旅遊的人潮，將可避免尖峰時段人擠人的擁擠現象發生。
- 2.配套方案可以針對同質性與異質性的景點作配套。
- 3.藉由專業旅遊雜誌或網站進行廣告，建立地方形象。
- 4.增加網頁互動介面與留言版的相關訊息討論區。
- 5.製作生態紀錄片，透過知名媒體進行宣傳推廣，吸引不同區域的 遊客。

### (二)對產業界的建議

由於觀光旅遊業其本身即為服務業之一種，故旅遊服務亦無法儲存，即具有不可儲藏性(perishability)。不可儲藏性，容易造成旅遊業者離、尖峰人數懸殊的問題，若透過調查分析可以更瞭解現況，減低懸殊問題。

- 1.在平日方面，業者可藉由資料採礦關聯分析，辨認於離峰時段出遊之遊客特性，以其為目標市場進行溝通促銷。
- 2.在遊憩活動種類方面，可經由設計獨具特色之觀光遊憩活動，吸引遊客願意於離峰時段出遊，此屬於行銷組合之範疇，如透過關聯分析：找出相關產品組合，結合各地方業者以策略聯盟方式對旅客做主題式的安排行程。
- 3.因為由網站分析可以得知旅遊景點排名，因此建議業者利用套票的方式以受歡迎的旅遊活動（如：泡湯、美食、主題樂園等活動）搭配較不熱門的活動來作為促銷，除了能吸引遊客在離峰時段至觀光景點遊玩，也希望藉此帶動一些其他週邊產業的發展。
- 4.行銷推廣方面：平均而言，遊客對於業者所提供的促銷活動之接受度皆相當高，因此在行銷推廣上，本研究建議業者在舉辦各項促銷活動時，假日與非假日進行差異化訂價，不定期舉辦促銷活動，並可以針對不同遊客偏好、時段作宣傳，舉辦不同性質的促銷活動；如星期一二為住宿促銷日，星期三為區域型活動促銷日，星期五、星期六為都市型配套促銷日等。
- 5.透過社團或團體舉辦之活動來吸引團客，如：學校單位、公家機構、私人企業等。
- 6.企業必須界定採取網路資料採礦之目標，以作為整體策略之準則。多數企業採取網路資料採礦多為支援公司整體行銷策略，亦有部份企業僅將網路作為其行銷管道之一，並將其與傳統行銷管道區隔。

進年來興盛的網路傳播有其一定的效果，這亦是國內旅遊市場在資訊傳播上可多加利用的一項資源。

### (三)未來研究方向

#### 1.調查時間延長：

由於僅有三個月的資料可作分析，若觀光局能保存一年以上的資料，將延長調查時間，進而針對不同季節不同時間作探討、比較。

#### 2.路徑分析：

將使用者存取檔案之URL之順序紀錄，並規劃瀏覽動線。從路徑分析讓提供網站管理者瞭解網站的結構，並修正不必要之網頁連結，將網頁結構合理化。

#### 3.連續、順序模型（Sequential Patterns）分析：

此分析可運用cookies做追蹤，應用時間框架歸納顧客行為模式。

#### 4.增加對多媒體物件之調查分析：

此次研究對於網頁內容的萃取與表示方式，僅針對觀光局資訊入口網站部分景點進行分析，而去除了多媒體的檔案格式。未來可朝多媒體內容分析邁進

#### 5.增加多國語言版本之分析：

觀光局資訊入口網站有多種語言，但此次研究僅針對中文版網頁分析，而觀光局網頁目前有八種語言版本，若能針對不同語言網頁分析，將更能瞭解國外人士對於我國感興趣的觀光景點，也能因此針對外國人作不同促銷與推廣。

#### 6.自動預測趨勢和行為

透過程式能讓資料採礦自動在大型資料庫中尋找預測性資訊，以往需要進行大量手工分析的問題如今也可以迅速直接由資料本身得出結論。

#### 7.廣告適配度

網路資料詳載使用者的網站瀏覽行為，亦可以計算點選標語廣告所顯示的行為偏好、哪些廣告是使用者較常訪問等，以作網頁配置之妥善安排。

#### 8.運用資料採礦技術進行商品交叉銷售

目前資料採礦技術在國內外的顧客關係管理上，最常使用在直效行銷、購物交叉行銷、客戶關係行銷、客戶服務或客戶流失分析等應用上。

#### 9.分類（classification）

按照分析對象的屬性分門別類加以定義，建立類組（class）。例如，將網站瀏覽者的屬性，區分為勇於冒險的瀏覽者、文藝趨向的瀏覽者、愛好蒐集資訊的瀏覽者、夜貓族群的瀏覽者等。並針對不同瀏覽者做有效的策略行銷。

#### 10.研究各國觀光局入口網站

研究各國觀光局入口網站發展現況及其推廣的策略，以提出我國觀光局入口網站行銷之建議。



11.針對不同觀光目的地進行比較研究

可在網路空間上看出旅客選擇旅遊路線模式的分佈情形，進而加以歸納，作為觀光遊憩系統規劃之參考。

## 參考文獻

- 吳凱雯(2001)：利用資料採礦技術提供網際網路使用者個人化服務，靜宜大學資訊管理學系碩士論文。
- 沈兆陽(2002)：資料倉儲與 ANALYSIS SERVICES SQL SERVER 2000 OLAP 解決方案，文魁。
- 楊勝博(1999)：隔週休二日國內旅遊參與型態影響之研究—以九族文化村、劍湖山世界遊樂區為例，逢甲大學建築及都市計畫研究所碩士論文，未出版
- 行政院經濟建設委員會都市及住宅發展處(1991)：台灣戶外遊憩政策之研究。
- 蔡麗伶(譯)(1990)：Mayo E. J. & L. P. Jarvis 著，旅遊心理學，揚智。
- 馬惠玲(2003)：台灣地區國內旅遊市場區隔變數之研究，逢甲大學建築及都市計畫研究所碩士論文。
- Berry, M. J., & Linoff, G.(1997). Data Mining Techniques for Marketing, Sales and Customer Support, Wiley.
- Berry, M. J., & Linoff, G.,Web, (2003).Sales and Customer Support, Wiley.
- Cooley, R., Mobasher, B., & Srivastava, J.(1997). Web Mining Information and Pattern Discovery on the World Wide Web, Proceedings of Ninth IEEE International Conference on Tools with Artificial Intelligence.
- Fayyad, U.(1998). Mining Database: Towards Algorithms for knowledge Discovery, IEEE Computer Society Technical Committee on Data Engineering..
- Gunn, C.(1988). Vacationscapes: Designing tourist regions. New York: Van Nostrand Reinhold.
- Marchionini, G.(1995). Information Seeking in Electronic Environments. New York: Cambridge University Press. ,p.100.
- Kelly, J., & Godbey, G.(1992).Time in time out. The Sociology of Leisure. pp.173-190.
- Kay T., & Jackson G.(1991). Leisure despite constraint: The impact of leisure constraint on leisure participation. Journal of Leisure Research..23 ( 4 ) , 301-313.
- Newman, M., & Landay, J.(2000). Sitemaps, Storyboards, and Specifications: A sketch of web site design practice. pp. 263-274.
- Pieter, A., & Dolf, Z. (1996). Data Mining, Addison Wesley Longman.
- 【網路資料】
- <http://ecommerce.vanderbilt.edu/cmepaper.revision.july11.1995/cmepaper.html>
- <http://www.computer.org/proceedings/hicss/0001/00015/00015042abs.htm>.
- <http://www.taiwan.net.tw/lan/cht/index/>
- <http://www.dgbasey.gov.tw/>

<http://www.twnic.com.tw>

<http://www.nua.com/surveys/>

[http://www.idc.com.tw/c\\_default.htm](http://www.idc.com.tw/c_default.htm)

<http://www.isc.org/>