運用網站探勘於健康資訊網站績效及行為分析—健康九九衛生教育網站為例

The Study of Applying Web Usage Mining Technology to the Web Performance and Browsing Behavior Analysis on the of Health99 Portal

曾暉仁[a,] ,葉雨婷[a],羅友聲[a], 劉建財[a*]
Hui-Jen Tseng[a,], Yu-Ting Yeh[a], Yu-Seng Lo[a], Chien-Tsai Liu[a*]
[a] *臺北醫學大學醫學資訊研究所*
110 *台北市信義區吳興街* 250 *號 醫學綜合大樓* 4 *樓*
[*] *通訊作者*: *劉建財*, ctliu@tmu.edu.tw

摘要

本研究以網站伺服器之網站日誌檔(Web log)瀏覽紀錄作為資料來源，並以網站績效分析及網站探勘技術方法來進行瀏覽使用者對網站瀏覽行為分析，將所得的分析結果提出建議，讓網站經營者了解民眾瀏覽使用者行為及喜好需求，並強化未來整體網站架構設計、經營決策，以及行銷規劃之參考依據。

關鍵字：網站使用探勘、網站日誌檔、關聯法則、健康九九衛生教育網

Abstract

The data of the reserch comes from the browsing records of the Web log stored in the web server, together with the analysis of the effects of the web and the web usage mining. The rendered results will offer the web *administrator the users' browsing habit*s and preferences, and, according to the results, help with the future design, management and marketing of the website.

Keywords: Web Usage Mining, Web log, Association rule, Health 99 Education Resource.

1、 *前言*

民眾取得健康資訊管道除了藉由醫師及護理人員衛教、醫療機構宣傳品、人際傳播之外，媒體似乎已經成為許多民眾對疾病或健康認知的重要管道[1]。另外，行政院衛生署藉由健康九九網站提供健康衛教資訊，其目的希望利用網路的便利性、即時性及不受時間與空間限制的特性，讓民眾能很方便地獲得需要的健康資訊[2]。目前對健康資訊的傳播，能否達成效益尚未可知，如何讓民眾能獲得所需的健康資訊，進而提升健康素養及生活品質與國家競爭力，是值得探討的議題。

近年來由於網際網路與通訊科技迅速發展，上網人口數不斷增加，根據台灣網路資訊中心(2009)調查台灣地區上網人口已突破 1,580 萬人，整體人口上網比率為 70.95%，此結果顯示出目前台灣網路使用普及化[3]。由這些數據可解讀出在網路上無時無刻都有著大量的資料在進行傳輸。在這資訊爆炸的時代，網際網路大海中，已經儲存著數以兆計由 0 與 1 所組成數據資料。若能運用資料探勘(Data Mining)的技術來探勘出隱藏其中珍貴的行為特徵資訊，便可提供經營者做為決策判斷的參考[4]。現行的網路探勘(Web Mining)是由 Etzioni[1996]首先提出，並定義為「使用資料探勘技術於網際網路的文件及服務中，發現及萃取出隱含的資訊」。Cooley, Mobasher, and Srivastava [1997]將網路探勘分成兩大類：網站內容探勘(Web Content Mining)與網站使用探勘(Web Usage Mining)；Kosala and Blockeel[2000]根據網頁之間的連結架構，提出第三類的網站結構探勘(Web Structure Mining)[5,6]。在網站內容探勘重點是利用檢索資訊、資訊擷取等技術，根據網站本身的內容進行資料探勘，其內容包括網頁文字、超連結及目錄結構…等文字探索[6,7]；網站結構探勘主要是利用圖形演算法來描述網站結構和表示連結的模式，可以利用此種模式來將網站進行分類；而網站使用探勘則重視找出瀏覽使用者在網站上的瀏覽型態式樣(Pattern)分析，以網站日誌檔為資料來源進行網站探勘[7]。

本研究將藉由對健康九九衛生教育網站之網站日

誌檔進行網站績效分析(Web Performance Analysis)和瀏覽使用者行為型態式樣分析，協助網站經營者更了解民眾對健康衛教資源取得的行為及喜好需求，並提供網站架構設計、經營決策和未來網站行銷規劃之參考，創造彼此雙贏的局面。

2、材料與方法

本研究的研究對象為一般民眾及衛教人員之網路瀏覽使用者，以行政院衛生署國民健康局之健康九九網站日誌檔進行網站績效分析及使用者對網站瀏覽使用行為探勘。網站績效分析包括瀏覽使用者基本資訊特徵、區域來源佔比、每月瀏覽人數、會員註冊數、停留時間及分佈佔比、瀏覽主題區塊排行、網頁點閱率及排行、平均每人瀏覽頁數及佔比…等指標。使用者瀏覽行為探勘則著重於研究瀏覽使用者的瀏覽目的及瀏覽型態式樣。本研究研究架構如圖 1 所示。

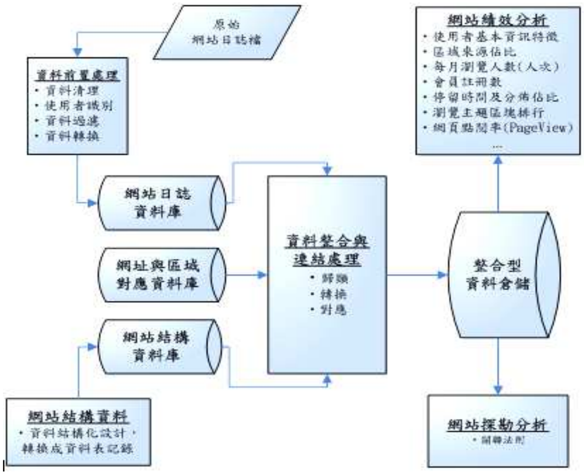

圖 1 研究架構

2.1 網站績效分析

根據 Eric E. Peterson 定義:網站績效分析是針對網站流量(Web Traffic)各式資料的評量(Assessment) ，包括網站交易資料(Web-based Transaction) 、網站主機效能(Web Server Performance)、可用性研究(Usabilities)、使用者送出資料(User Submitted Information)…等，藉由評量，對使用者在網站上使用經驗取得概括性的了解[8]。以下為部份常用網站績效指標定義：

瀏覽人數(Visitor)—同一個使用者不管瀏覽多少網頁，或多久的時間，都只能算作一個訪客數。

會員註冊數—已加入會員的使用者人數。

停留時間—指使用者造訪網站瀏覽網頁所花的時間。

點閱率(PageView)—又可稱為網頁瀏覽數，指整個網頁的文件被使用者成功請求(Request)的次數。

2.2 網站瀏覽行為分析

網站使用探勘是針對網站日誌檔進行使用者的瀏覽行為分析，從資料中發現多數使用者對網站的存取瀏覽行為特徵的流程[7]。分析的資料來源多為瀏覽使用者與網站互動的所留下的資訊，包含網站日誌檔、交易資料、會員註冊資料…等；透過對所留下的紀錄分析出瀏覽使用者的瀏覽行為模式。常用的技術包括關聯法則(Association Rule)、序列型樣(Sequential Pattern)、分群(Clustering)…等技術應用。

2.3、關聯法則(Association Rule)

關聯法則主要是找出在某一事件或是資料中會同時出現的東西。運用在於找出資料庫中資料項目(Item)間彼此的關聯性，探討同一事件中，某一項目出現時，另一個項目也在該事件出現的機率。關聯法則於網站使用探勘的應用，主要是找出使用者瀏覽網頁時，網頁之間的相互關聯關係，找出哪些網頁常會被一同存取。例如：60%的使用者參觀過 A 網頁時，也瀏覽了 B 網頁等；利用這些相關性，可做為網站架構設計參考，讓使用者在尋找有關聯的網頁之間的瀏覽路徑能夠比較直接，能夠更容易的瀏覽到想觀看的網頁[5]。關聯法則的定義：假設 I 是交易資料庫中所有項目的集合，T 是全部交易資料的的集合，一筆交易資料 Tj，$Tj \in T$，是由一些項目所形成的集合，稱為項目集（Itemset），項目集的項目個數 $|Tj| \geq 1$。若在項目集 X 與 Y 之間有一關聯法則，則表示式為：$X \rightarrow Y$；X、

$$Y \subseteq I，且 X \cap Y \neq \phi 。$$

關聯法則的產生可藉由支持度(Support)及信賴度(Confidence)來決定關聯法則是否為有效的規則，說明如下：關聯法則 $X \rightarrow Y$ 的支持度，表示在所有的交易集合中，同時包含有 X 與 Y 的比率值，即資料庫中 $X \cup Y$ 的交易所占的百分比，記為 Support$(X \cup Y)$：

$$\mathrm{Support}(X \rightarrow Y) = \frac{(X \cup Y)\text{交易數量}}{\text{總交易數量}}$$

關聯法則 $X \rightarrow Y$ 的信賴度，是用來確定超過 Support 事件的可信程度，也就是在包含有 X 的交易集合中，也同時包含有 Y 的比率值，記為 Confidence$(X \rightarrow Y)$，Confidence 的表示方式為：

$$\text{Confidence}(X \rightarrow Y) = \frac{\text{Support}(X \cup Y)}{\text{Support}(X)}$$

其中用來判定是否為高頻項目集的門檻值稱為最小支持度（Minimum Support）；而判定能否由高頻項目集形成規則的門檻值稱為最小信賴度（Minimum Confidence）。使用者可以藉由調整這兩個過濾規則的門檻值，來找出最合適的規則[5,9,15]。

2.4 原始網站日誌檔

原始網站日誌檔是透過國民健康局之研究計劃而取得 2008 年 1 月 1 日到 2008 年 12 月 31 日的網站日誌檔；其資料容量約有 51.8GB。此日誌檔為微軟 IIS 日誌檔格式(Microsoft IIS Log Format)，檔案格式包含:登入日期時間(Datetime)、客戶端位址(Client IP Address)、來源(Referrer)、客戶端(User Agent)、請求方法(Method)、URI 請求(URI Query)、HTTP 狀態(HTTP Status)、點心(Cookie)、…等欄位資訊[10]。

2.5 資料前置處理

此步驟是針對進行網站使用探勘時，主要資料來源為網站日誌檔，根據 Cooley 的定義，網站日誌檔前處理的動作，提出一套完整的處理流程[11]，說明如下：

資料清理(Data Cleaning)：此程序為重要流程之一，去除不當的資料，提高資料處理速度，本研究網站日誌記錄檔內，並非所有資料都具備探勘意義，故需刪除圖檔、程式檔、格式檔及動畫檔…等紀錄，例:.gif、.jpg、.bmp、ico、.css、.js、.swf、…等檔案。

使用者辨識(User Identification)：先以 Cookie Session 欄位歸戶成為單一瀏覽使用者，當 Cookie Session 欄位若為空值(Null)時，再加入 IP 來源及 User-Agent 欄位當做歸戶判斷原則依據[12]。

資料過濾：針對 User-Agent 及 Status 欄位進行過濾。User-Agent 欄位過濾目的是要找出網路蜘蛛(Spider)將它去除，例:wwwrobot、Googlebot、Yahoo、msnbot、WebTrends、…等字串條件過濾。Status 欄位過濾目的是要將正常的瀏覽紀錄保留下來，去除雜訊，例:200、206、301、302、…等狀態條件被過濾後保留[10]。

資料轉換：主要應用於使用者辨識時，依判斷歸戶法則，設計出函數資料轉換程式，轉成為所需的資訊。

2.6 網站日誌資料庫

此資料庫為原始網站日誌檔，經過資料前置處理後，將已整理過的網站日誌資料儲存起來的地方。

2.7 網址與區域來源對應資料庫

使用 MaxMind 公司之產品 GeoIP Country 資料庫做為國家區域之 IP 來源識別[13]，以及加入財團法人台灣網路資訊中心之 IPv4 核發分配表做為 ISP 來源識別[14]，增加瀏覽使用者基本資訊特徵價值。

2.8 網站結構資料庫

因健康九九網站之結構資訊也俱參考意義，故先以網站導覽為結構，經資料結構分析設計後，將資訊轉存成資料庫，並在網站績效分析及網站探勘時可加以運用。健康九九網站建置於 1998 年 10 月，係衛生署為透過網路提供民眾健康相關資訊而設立，至今收錄了近 3,300 件衛生教育教材，線上教材總數約 1,300 件。可依教材類別、年齡層、媒體種類、關鍵字、…等進行搜尋找到所需教材。目前教材共分為 23 類[2]。

2.9 資料整合與連結處理

主要針對於多資料來源的資料進行整合、歸類、轉換和對應的處理，依不同的分析需求，設計函數程式轉換儲存成所需的資料格式及連結資料關係，例:國內外來源、政府與非政府來源、瀏覽主題區塊、教材類別、教材名稱…等資料連結處理。

2.A 整合型資料倉儲

此階段為資料倉儲建立，經資料前置處理及資料整合與連結處理後，將多資料來源的資料整合建立在一個貯存庫中。

2.B 資料分析工具

本研究使用的分析軟體為微軟商業智慧解決方案所發行的 SQL Server 2005 資料庫，再利用 LogParser 軟體將網站日誌檔載入資料庫中。除此之外，也使用資料的函數設計程式，開發出為資料前置處理及網站績效統計等工作使用；網站探勘應用，也是採用資料庫之 Analysis Services 作為探勘實作工具，如表 1 所示。

表 1 網站探勘分析工具[15]

步驟	分析工具	說明
資料載入	LogParser2.2 版	將日誌檔載入資料庫

資料清理、過濾、整合、儲存、統計	SQL Server 2005 資料庫	資料清理、整合、過濾、使用者識別、轉換及網站績效分析
網站探勘實作	SQL Server 2005 Analysis Services	作為關聯分析等探勘分析

3、結果

3.1 網站績效統計分析

透過對網站績效分析探討問題，所得出結果如下:

(1)健康九九網站在 2008 年國內外 IP 來源數據分析可看出全年瀏覽人數分別為台灣平均每月佔 72,262 人及國外平均每月佔 47,061 人，合計平均每月為 119,323 人；依百分比來看台灣平均佔 60.6%及國外佔 39.4%；平均每位瀏覽使用者瀏覽頁數為 7 頁，如圖 2 所示。

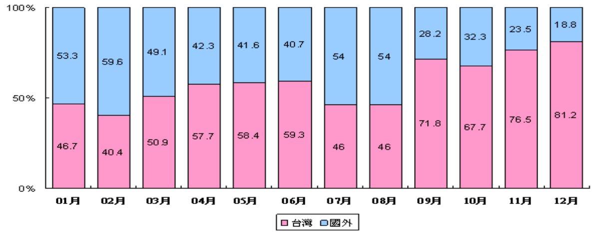

圖 2　2008 年國內外 IP 來源(本研究整理)

(2) 2008 年政府與非政府 IP 來源數據分析可看出全年瀏覽使用者分別平均政府佔有 19.2%、非政府佔 43.4 % 及其他佔 38.4%，如圖 3 所示。

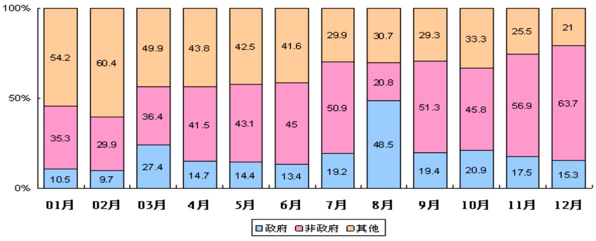

圖 3　2008 年政府與非政府 IP 來源(本研究整理)

(3)2008 年會員註冊人數據分析平均每月的註冊人數為 4,375 人，在九、十月的註冊人數明顯大幅增加約為 9,200 人，如圖 4 所示。

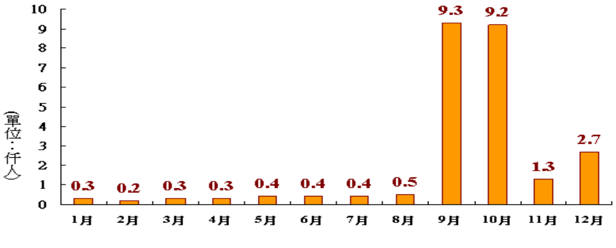

圖 4　2008 年會員註冊人數(本研究整理)

(4)2008 年瀏覽使用者來訪停留時間，每人平均停留時間在 3 分鐘內為 85.1%，佔大多數，如表 3 所示。

表 2　2008 年每人平均停留時間(本研究整理)

停留時間	累計百分比
1 分鐘	72.2%
3 分鐘	85.1%
10 分鐘	92.7%
30 分鐘	97.7%

(5)2008 年瀏覽主題區塊排行，以衛教資源區塊為最熱門排行第一名，故本研究先針對衛教資源區塊進行深入探究分析，如表 3 所示。

表 3　2008 年瀏覽主題區塊排行(本研究整理)

排行	瀏覽主題區塊	網頁瀏覽數
1	衛教資源	1,813,396
2	衛教主題館	572,393
3	影音劇場	265,942
4	珍愛生命	231,736
5	疑問醫答	132,034
6	線上健檢	130,385
7	下載專區	103,888
8	最新新聞	98,203
9	活動中心	97,210
10	會員中心	92,043

(6)衛教資源區塊平均每月的瀏覽人數有 11,812 人；平均每人瀏覽網頁數為 4 頁；平均網頁瀏覽數分佈 5 頁內佔 87.7%為最多，如圖 5 及 6 所示。

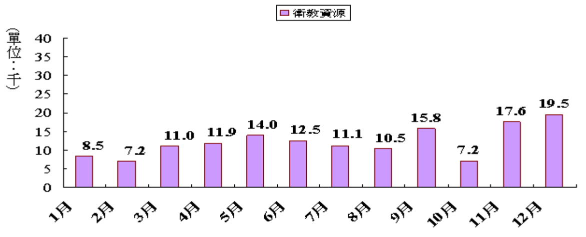

圖 5　2008 年衛教資源區塊瀏覽人數(本研究整理)

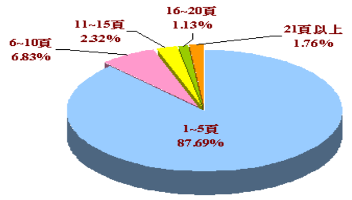

圖 6　2008 年衛教資源網頁瀏覽頁數分佈

(7)衛教資源區塊依教材類別統計，排行前十名佔 86.7%；排行前四名佔 64.7%分別為菸害防制、傳染病防治、慢性病防治、癌症防治等類別，如表 4 所示。

表 4　2008 年衛教資源區塊依教材類別排行

排行	教材類別	網頁瀏覽數
1	菸害防制	190,016
2	傳染病防治	75,703
3	慢性病防治	44,342
4	癌症防治	43,337
5	食品營養	29,207
6	病人安全	26,562
7	性教育	19,258
8	健康體能	15,747
9	兒童健康	15,376
10	婦女健康	14,191

(8)2008 年衛教資源區塊依教材名稱排行，前十名佔所有教材為 13.91%，教材主題內容分別為禁菸防治佔 5 筆及腸病毒佔 4 個筆之比率分佈，如表 5 所示。

表 5　2008 年衛教資源區塊依教材網頁排行

排行	教材名稱	網頁瀏覽數
1	本場所禁止吸菸，最高罰 1 萬元	9,549
2	趕走腸病毒　贏在常洗手	8,966
3	本場所禁止吸菸，違者罰 2 千~1 萬元	8,898
4	室內三人以上工作與公共場所禁止吸菸	8,768
5	本場所禁止吸菸，最高罰 1 萬元	8,666
6	『遺失的微笑』紀錄片	7,954
7	趕走腸病毒　贏在常洗手	6,079
8	禁止吸菸貼紙	6,070
9	腸病毒 71 型相關病徵及疑似重症轉診時機	5,569
10	腸病毒重症前兆病徵，掌握黃金治療時間	5,525

3.2 **關聯分析**

使用網站探勘分析工具所得出結果如下：

(1)依『相依性網路』工具主要分析目的在於藉由圖形描繪出衛教資源區塊依各教材類別的關聯性，由圖可看出[菸害防制]類別跟以下類別[中醫中藥、安全促進、正確用藥、視力聽力保健、心理健康促進、性教育、其他、…等]類別有著很強烈的關聯性，如圖 7 所示。

圖 7　2008 年衛教資源區塊之教材類別關聯圖

(2)支援度是指在購物籃分析中同時包含關聯規則左右兩邊物品的交易次數百分比，即支援這個規則的交易的次數百分比[5]。依圖可看出單一項目集如[菸害防制]、[傳染病防治]、[慢性病防治]、[癌症防治]等類別支援度最高；以二個項目集如[傳染病防治,菸害防制]、…等類別的支援度較高，如圖 8 所示。

支援	大小	項目集
26988	1	菸害防制
16490	1	傳染病防治
14718	1	癌症防治
12976	1	慢性病防治
7856	1	食品營養
6266	1	性教育
5884	1	病人安全
5847	2	傳染病防治, 菸害防制
5800	1	兒童健康
5288	1	健康體能
5064	1	其他
4450	2	癌症防治, 菸害防制
4418	1	婦女健康
4181	2	慢性病防治, 菸害防制
3914	1	安全促進
3611	2	慢性病防治, 傳染病防治
3420	2	癌症防治, 傳染病防治
3066	1	心理健康促進
2966	2	食品營養, 菸害防制
2919	2	慢性病防治, 癌症防治
2860	2	食品營養, 慢性病防治
2739	2	食品營養, 傳染病防治
2609	2	性教育, 菸害防制
2589	3	癌症防治, 傳染病防治, 菸害防制
2563	3	慢性病防治, 傳染病防治, 菸害防制
2352	2	食品營養, 癌症防治

圖 8　2008 年衛教資源區塊之教材類別關聯支援

(3)信賴度是指購物籃分析中有了左邊商品，同時又有右邊商品的交易次數百分比，也就是說在所有的購買了左邊商品的交易中，同時又購買了右邊商品的交易概率[5]；依圖可看出[中醫中藥, 物質濫用防制->癌症防治]、[中醫中藥,物質濫用防制->食品營養]…等類別的關聯重要度與機率強度極高，如圖 9 所示。

機率	重要性	規則
0.944	0.429	中醫中藥, 視力聽力保健 -> 菸害防制
0.940	0.691	中醫中藥, 物質濫用防制 -> 癌症防治
0.940	0.426	中醫中藥, 物質濫用防制 -> 菸害防制
0.939	0.643	物質濫用防制, 遺傳型疾病防治 -> 傳染病防治
0.924	0.420	物質濫用防制, 遺傳型疾病防治 -> 菸害防制
0.922	0.419	中醫中藥, 遺傳型疾病防治 -> 菸害防制
0.920	0.632	中醫中藥, 物質濫用防制 -> 傳染病防治
0.920	0.418	中醫中藥, 口腔保健 -> 菸害防制
0.915	0.417	中醫中藥, 心理健康促進 -> 菸害防制
0.915	0.416	中醫中藥, 性教育 -> 菸害防制
0.913	0.955	中醫中藥, 物質濫用防制 -> 食品營養
0.913	0.416	物質濫用防制, 健康保險 -> 菸害防制
0.912	0.631	口腔保健, 健康保險 -> 傳染病防治
0.911	0.632	物質濫用防制, 社區職場健康 -> 傳染病防治
0.911	0.631	中醫中藥, 心理健康促進 -> 傳染病防治
0.910	0.414	口腔保健, 社區職場健康 -> 菸害防制
0.908	0.630	視力聽力保健, 正確用藥 -> 傳染病防治
0.908	0.414	中醫中藥, 社區職場健康 -> 菸害防制
0.908	0.413	物質濫用防制, 社區職場健康 -> 菸害防制
0.904	0.676	物質濫用防制, 遺傳型疾病防治 -> 癌症防治
0.904	0.628	口腔保健, 心理健康促進 -> 傳染病防治
0.903	0.677	中醫中藥, 心理健康促進 -> 癌症防治
0.903	0.627	物質濫用防制, 健康保險 -> 傳染病防治

圖 9　2008 年衛教資源之類別關聯重要度與機率

4、討論與結論

本節先整體綜合本研究之發現如下說明：

以 2008 年國內外 IP 來源數據分析可得知除了台灣 IP 來源之外，國外的瀏覽使用者也佔一定的比率。再者以全年瀏覽人數合計平均每月為 119,323 人，與會員註冊人數平均每月為 4,375 人來比較，佔比差距極大，在九、十月的會員註冊人數呈明顯大幅增加。全年瀏覽使用者平均每位瀏覽頁數為 7 頁，瀏覽使用者來訪停留時間每人平均停留在 3 分鐘內為 85.1%，佔大多數，如何增加瀏覽使用者的上網停留時間也是網站經營者思考的問題之一，故可建議可不定期多次舉辦網路行銷活動來增加網站黏度。

在瀏覽主題及網頁區塊方面，以衛教資源區塊為最熱門排行第一名，故本研究先針對衛教資源區塊進行深入研究分析。衛教資源網頁平均每月的瀏覽人數分析有 11,812 人，以全年瀏覽人數合計平均每月為 119,323 人來比較，佔比差距極大，平均網頁瀏覽數分

佈 5 頁內佔 87.7%為最多。衛教資源區塊依教材類別統計排行及關聯分析的支援度運用可了解瀏覽使用者對於菸害防制、傳染病防治、慢性病防治、癌症防治等類別的喜好，並可歸類為 2008 年最熱門的健康衛教類別。在 2008 年衛教資源區塊依教材名稱排行，熱門話題為菸害防治及腸病毒之比率分佈最高，由此可知國民健康局對這類的行銷活動很積極在推廣及宣導。

在關聯分析方面，由『相依性網路』工具分析，可了解當民眾瀏覽有[關菸害防制]的類別衛教資訊，同時也會去瀏覽跟以下類別[中醫中藥、安全促進、正確用藥、視力聽力保健、心理健康促進、性教育、…等]類別。以置信度來看可了解民眾同時去瀏覽[中醫中藥及物質濫用防制]的類別衛教資訊，也會去瀏覽[癌症防治或食品營養衛教]資訊機率很高。

綜合以上研究發現，可看出網站經營者需調整網站經營的方向和重點，以及可得知網路行銷活動能否達成效益，當作未來決策的參考資訊。另外，也可看出普遍的瀏覽使用者對健康資訊取的行為喜好需求之趨勢分析及關心的健康衛教資訊議題。

對目前市場上的網站分析工具，跟本研究運用方法做比較，市場上的工具的缺點是只提供固定套表式的統計方法功能，在局限功能下，無法進一步依設定主題來分析瀏覽使用者行為喜好，優點是無須撰寫任何程式，並可在短時間內產出結果；本研究運用方法，優點是可針對設定的主題來分析瀏覽使用者行為喜好，以及可靈活運用工具設計出程式和模型，缺點是須費時在撰寫程式，以及資料前置處理和資料整合與連結等工作處理上。

未來研究方向可往不同的角度切入設定新的研究議題，也可運用資料探勘之不同的演算法技術，例:序列型樣、針對網頁或使用者進行分群…等進行研究。

5、研究限制

整理目前遇到的問題如下，需大量硬碟空間及高階效能 CPU 來運算分析、網站日誌檔可分析資料有限、使用者 IP 來源沒有精確的 IP 資料庫可參照、內部 IP 來源無法分析來源問題、無法準確估計究竟有多少人上網、瀏覽器及代理伺服器的快取問題等問題。

6、致謝

感謝行政院衛生署國民健康局資助本研究計畫。

7、參考文獻

[1] 王淑華，台灣地區健康傳播模式與效益之研究ー從知溝理論觀點，世新大學傳播管理學研究所碩士論文，2005。

[2] 健康九九衛生教育網，http://www.health99.doh.gov.tw,2009.

[3] 台灣網路資訊中心（2009）。2009年01月「台灣寬頻網路使用調查」報告，http://www.twnic.net.tw/download/200307/0901a.doc

[4] 王佳鳳，運用資料採礦於台灣觀光資訊入口網站瀏覽行為分析，靜宜大學觀光事業學系研究所碩士論文，2004。

[5] 劉玉敏，線上漸進式網站使用探勘之研究－以產品推薦為例，雲林科技大學資訊管理系碩士論文，2002。

[6] R. Cooley, B.Mobasher and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web", In Proceeding of ICTAI'97, 1997.

[7] Bamshad Mobasher, Robert Cooley, Jaideep Srivastava, "Automatic Personalization Based on Web Usage Mining", Communications of The ACM, vol.43, pp.142-151, 2000.

[8] 連啓佑，網站績效分析與管理，文魁資訊股份有限公司，2006。

[9] 曾憲雄、蔡秀滿、蘇東興、曾秋蓉、王慶堯，資料探勘(Data Mining)，旗標出版，2008。

[10] Hypertext Transfer Protocol -- HTTP/1.1, http://www.w3.org/Protocols/rfc2616/rfc2616.html

[11] Robert Cooley, Bamshad Mobasher, Jaideep Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns", Knowledge and Information Systems, vol. 1(1), 1999.

[12] Bamshad Mobasher, Bettina Berendt, Myra Spiliopoulou, "KDD for Personalization", PKDD 2001 Tutorial, 2001.

[13] Geolocation and Online Fraud Prevention, MaxMind, http://www.maxmind.com,2009.

[14] TWNIC財團法人台灣網路資訊中心，IPv4核發分

配表，http://rms.twnic.net.tw/twnic/Search-non.htm

[15] 尹相志，SQL Server 2005 資料採礦聖經，學貫行銷股份有限公司，2006。