

多維度網頁文件分類模式

楊士霆

南華大學資訊管理學系
stingyang@mail.nhu.edu.tw

黃家偉

南華大學資訊管理學系
stingyang@mail.nhu.edu.tw

摘要

現今管理者多以資訊分類方式進行管理，以便讓需求者尋找所需之資訊，然而於網頁分類議題當中，多數管理者依靠網頁分類機制為依據將網頁分類，但此方式缺乏以撰寫者角度為考量，此外，由於僅考慮網頁當中文字資訊，但網頁外部特徵亦有可能影響需求者瀏覽結果，造成需求者與管理者對網頁內容理解不同。因此本研究發展一套「多維度網頁文件分類模式」，以網頁使用者與網頁設計者角度解析網頁文件文字資訊與顯示特徵，接著於「網頁文件意象風格判定」模組判定網頁文件情緒類別；於「網頁文件專業類別判定」模組則判定網頁文件之專業類別。

關鍵詞：網頁文件分類、知識管理、色彩心理學、眼動追蹤、標籤區域

第一章、研究背景

由於網路普及與便利性，人們已習慣透過書本轉而從網路上獲取知識，然而網路資訊量龐大，造成知識網頁需求者難以從大量網頁中，尋找到所需之網頁資料，因此如何管理或篩選網頁文件，讓知識網頁需求者有效地尋找所需網頁資訊，乃成為目前網頁分類、管理中重要議題之一。為了吸引需求者瀏覽網頁，網頁設計者常藉由網頁色彩配置、圖文編排等，以吸引需求者閱覽，然而此方式容易造成需求者於瀏覽網頁文件時，因網頁文件視覺特徵對於知識網頁需求者情緒與閱讀效率等產生影響，因此，知識提供者除了推薦需求者適當領域之網頁外，並需兼顧網頁文件視覺特徵對需求者情感影響。然而知識提供者歸類網頁文件資訊方式，以及篩選適合之色彩配置方式，雖可藉由網頁分類技術以得知網頁類別，但此分類技術缺乏設計者撰寫網頁文件之微感觀點為考量，因而未能取得具設計者表達網頁文件之原意，導致網頁無法被準確地分類，此外，知識提供者亦無針對網頁文件色彩配置作為需求者情緒影響之判定機制，因而造成知識提供者僅能以自身觀點判定，但此方式可能因知識網頁提供者與需求者觀點不同，而造成需求者觀看此份網頁文件感受不同。綜合上述，其既有之運作模式如圖 1 之 AS-IS Model 所示。

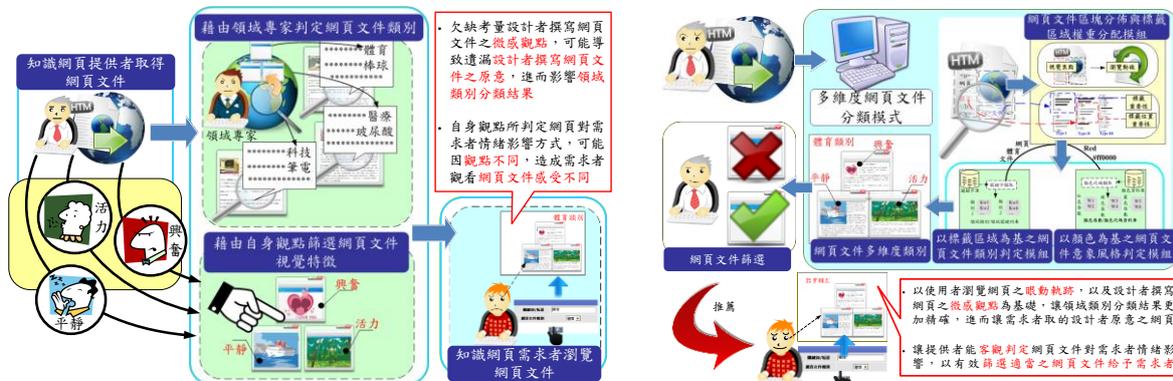


圖 1 AS-IS Model

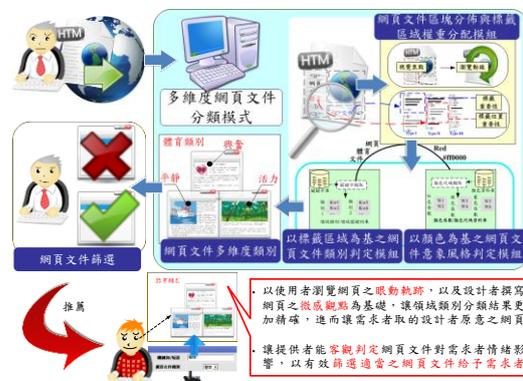


圖 2 TO-BE Model

有鑑於此，本研究如圖 2 之 TO-BE Model 所示，期望藉由網頁文件標籤所含色彩代碼與文字型資訊為解析資料，並考量使用者瀏覽網頁文件眼動軌跡，以及設計者撰寫網頁微感為基礎，以進行網頁文件類別判定，進而發展一套「多維度網頁文件分類模式」，以幫助知識網頁提供者有效推薦需求者適當領域之網頁文件，並篩選對需求者不適當情感影響之網頁文件。首先本研究「多維度網頁文件分類模式」乃由「網頁文件區塊分佈與標籤區域權重分配模組」以判定網頁文件中使用者所注視之重要區塊，以及解析網頁文件中含有設計者原意之重要資訊，並藉由使用者與設計者觀點進行「網頁文件意象風格判定模組」以判定該網頁色彩配置對於使用者情緒影響，以及「網頁文件專業類別判定模組」以針對該網頁資訊內容進行分類。

第二章、文獻回顧

於網頁文件探勘議題中，Lim et al (2005) 乃提出以網頁文件中 UML 與 HTML 語法或 Tag 等特徵作為網頁分類之分析資料，建構一套網頁自動分類系統，該研究乃提出可以擷取網頁中網址與 HTML 語法等特徵，進而提供後續研究之網頁文件分類的分析特徵與資料。Chen et al (2009) 先行針對複合表與關聯表所能引導搜尋之空間範圍及兩者間之搜尋關係，建構語義關連圖 SRG (Semantic Relationship Graph)，之後以天真貝式分類器 NB (Naive Bayes) 為主，開發一套以語意關連圖為基之多關聯天真貝式分類器，該分類器乃根據語意關連圖之分析結果，排除不必要之特徵與關聯性，進而避免產生無相關之連結。Furnkranz (2002) 乃先從網頁之本文中取得分類特徵，之後考量網頁文件中超連結特徵，連結至對應之網頁，並取得該網頁之所有分類特徵，並且整合

與目標網頁之特徵進行整合，以有效地針對目標網頁進行分類。**Youn 與 Jeong (2009)**則從文件中選擇決定性之特徵，並將此些特徵依比重進行分類，再針對分類後之文件特徵進行漸進式篩選，最後將篩選結果作排序，以取得排序為前之特徵值及其分類屬性，提高分類器之準確率。而 **Zhang et al (2009)** 先行利用主要成分分析法 (Principal Component Analysis) 分析網頁之主要構成要素以擷取特徵，並從特徵集中排除不相關之特徵，之後利用遺傳演算法逐步進化之特性，從特徵子集中選取適合之特徵做為分類預測屬性，最後再依預測結果完成分類任務。

使用者瀏覽網頁行為分析議題中，**Wang 與 Day (2007)** 乃利用眼動儀追蹤瀏覽者於瀏覽有意義路徑時，瀏覽者注意力的分佈，以得知路徑上廣告影響瀏覽者程度，並且該研究乃採用實驗室實驗法進行眼動儀測試。從該研究得知廣告訴求調整：廣告於路徑初期與末期應採用理性訴求，而中期應採用感性訴求。網站資訊安排：網站資訊初期應採用容易引起瀏覽者興趣之資訊，並且網站重要資訊放置順序乃應由左至右、有上至下。此外，**Schiessl et al (2003)** 研究則要求受試者上網找尋網頁文件中鏈結，同時利用人眼追蹤系統追蹤受試者。其追蹤結果發現男性於觀看網頁文件時，大部分先行瀏覽圖片較少瀏覽文字，而女性於觀看網頁文件時，則仔細瀏覽網頁文件之文字內容，故得知男性乃以圖片與寬鬆導向瀏覽為主、女性則以文字與精確導向瀏覽為主。另外於網頁研究中，**Oulasvirta et al (2005)** 利用眼球運動對網頁區域進行探討，該研究分為實驗組與對照組進行測試，測試使用者第一眼焦點與記憶最深刻網頁區域，實驗組為知道該網頁區域，而對照組則相反，並從實驗結果得知使用者第一眼焦點與記憶最深刻皆落於網頁左邊之區塊中。於顏色對於使用者視覺影響研究中，**Pearson 與 Schaik (2003)** 針對鏈結選單進行視覺式搜尋程序與交互式搜尋程序並重複測量，以得知鏈結選單配置與顏色對於使用者有何影響，其結果表示鏈結選單顏色為藍色明顯比紅色容易被發現，而鏈結選單配置於網頁文件左邊與右邊容易被使用者注意 (**Park & Noh 2002**)。

色彩心理學議題中，**Wexner (1954)** 針對八個顏色與十一個情緒進行研究，其研究結果指出藍色代表安全、舒適與溫和，紅色代表刺激、保護、反抗與興奮，橙色代表煩惱與沮喪，黑色代表消沉與有力，紫色代表高貴與神秘，黃色代表快活與快樂。另外，**Adams 與 Osgood (1973)** 則針對各種不同文化情感與顏色相似性關係進行探討，其探討各文化顏色情緒結果得知紅色為正負向情緒，黑色與灰色通常為負向情緒，白色、藍色與綠色為正向情緒，另外，白色、灰色與黃色顏色較為薄弱，黑色與紅色顏色較為強烈。色彩心理學探討使用者現在心情研究中，**Shugrina et al (2006)** 乃利用人臉辨識以得知使用者表情與筆觸並結合影像色彩以計算表情之情緒，最後參考表情、色彩情緒與筆觸進行影像色彩變更，以得知符合使用者情緒之影像。於圖片色彩研究中，**Wang et al (2006)** 利用粒子群演算法 (Particle Swarm Optimization; PSO) 與模糊 C 均值聚類 (Fuzzy C-Means Clustering) 針對圖片進行色彩分割，以得知影像圖片之主要色彩，最後利用三維情感模型做影片圖片色彩情緒意涵分類。**Yang 與 Peng (2008)** 則以區域分割為基礎，並分析影像圖片以得知色彩長條圖匹配，接著使用者選擇目前心情，則該影像圖片依照使用者目前心情進行色彩微調，進而得知符合使用者心情之影像圖片。其中該研究乃利用心情對應顏色，如強而有力代表紅色，神秘代表紫色，專業代表灰色與黑色，清新代表藍綠色，流行、太陽與友善代表黃色，激烈、大膽與刺激代表紅色，清新與自然代表綠色，平靜與悠閒代表藍色。**Wei et al (2004)** 乃提出一套緊湊集合特徵之色彩情緒分析與次要類型區別進行電影情緒轉換判斷。該研究利用電影場景表示顏色與整部電影所提取情緒，以進行全局調色板差別判斷，進而得知電影情緒轉換。

第三章、多維度網頁文件分類模式

本研究所提出之「多維度網頁文件分類模式」先行探討「網頁設計者」與「網頁使用者」觀點，以得知網頁設計者於設計網頁文件，所表達重要資訊，以及網頁使用者於微感時觀看網頁文件，所注視重點區塊。是故，本研究首先乃開發「網頁文件區塊分佈與標籤區域權重分配模組」，以解析網頁區塊之重點資訊內容，之後即可進行後續多維度網頁文件類別判定。於網頁類別判定方法論中，本研究乃開發「網頁文件意象風格判定模組」以判定該網頁所呈現之情緒類別，以及「網頁文件專業類別判定模組」以判斷該網頁所呈現專業類別，進而獲得多維度網頁文件分類結果。本研究之分析流程與模式架構如圖 3 所示，當中各模組之細節於以下分別敘述。

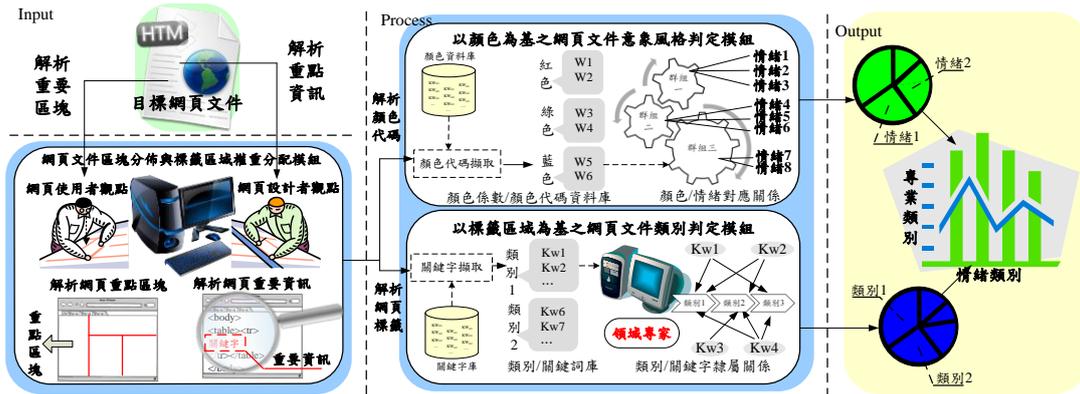


圖 3 多維度網頁文件分類方法論之流程架構

3.1 網頁文件區塊與資訊權重分配模組

本研究所提出之「網頁文件區塊與資訊權重分配模組」乃針對網頁設計者與使用者觀點進行探討，以獲知該網頁在網頁設計者與網頁使用者之創造與瀏覽中，所呈現的網頁重點資訊。以下分別探討設計者與使用者觀點對網頁重點資訊擷取之過程，最後彙整上述網頁使用者與設計者觀點之權重分配，以計算網頁中重點區塊與重要資訊權重，本研究將此模組所使用之符號定義與詳細步驟說明如下：

符號定義

- ALW_i 第 i 個區塊分佈之總權重
- LRW_i 第 i 個由左至右區塊之權重值
- TBW_j 第 i 個由上至下區塊之權重值
- N(LRW) 目標網頁文件中由左至右區塊之權重值總和
- N(TBW) 目標網頁文件中由上至下區塊之權重值總和
- LRT 目標網頁文件中由左至右區塊之權重門檻值
- TBT 目標網頁文件中由上至下區塊之權重門檻
- WD_T 所考量之目標網頁文件
- WD_i 第 i 份網頁文件
- FT_n 第 n 個 frameset 標籤區域所包含之字串與程式編碼
- W_{n,k} 第 n 個 frameset 標籤區域之第 k 個字串
- F_n 第 n 個 frameset 所分割之分割欄列字串
- F_{n,j} 第 n 個 frameset 所分割之第 j 個分割欄列係數
- F'_{n,j} 第 n 個 frameset 所分割之第 j 個轉換分割欄列係數
- FT_{N,FT} 目標網頁文件中最後一個 frameset 標籤區域
- F'_{n,N,F} 第 n 個分割欄列字串中所分割之最後一個轉換分割欄列係數

RemS	網頁程式碼編碼常用含括字串符號與運算子符號集合，即 $RemS = \{ \text{“}, = \}$
PC_m	第 m 個分割標籤，當 $m=1$ 時， $PC_1 = cols$ ， $m=2$ 時， $PC_2 = rows$
Seg_Set_m	frameset 標籤區域分割標籤，當 $m=1$ 時 Seg_Set_1 為垂直式分割之集合，當 $m=2$ 時 Seg_Set_2 為水平式分割之集合
SegS	網頁程式碼編碼常用符號或係數以進行標籤間分隔之符號集合
FPC	網頁文件之分割方式
Seg_m	網頁文件存在分割標籤之程式碼，當 $m=1$ 時 Seg_1 則說明網頁文件存在 cols 標籤之程式碼，當 $m=2$ 時 Seg_2 則說明網頁文件存在 rows 標籤之程式碼，當 $m=3$ 時 Seg_3 則網頁文件同時存在 cols 標籤與 rows 標籤之程式碼
MAX_m	目標網頁文件解析度面積，當 $m=1$ 則 MAX_1 為面積寬度（以下簡稱寬係數），當 $m=2$ 則 MAX_2 為面積高度（以下簡稱高係數）
$SegS_m$	第 m 個分割欄列係數符號集合，當 $m=1$ 時 $SegS_1$ 為 %， $m=2$ 時 $SegS_2$ 為 *
$N \cdot_{FT}$	目標網頁文件中含 frameset 標籤區域之個數
$N \cdot_F$	分割欄列字串中含分割欄列係數之個數
FL_m	網頁文件中垂直式分割之第 m 塊分割區塊面積
FW_m	網頁文件中水平式分割之第 m 塊分割區塊面積
NR_j	目標網頁文件中第 j 個標籤所包括之標籤區域
DOM_T	目標網頁文件之文件物件模型架構
DOM_i	第 i 個網頁文件之文件物件模型架構
TR	標籤位於網頁空間位置
$NT_{j,k}$	目標網頁文件文件物件模型架構中第 j 個標籤節點位於第 k 個位置
BN	網頁文件程式碼中常用以表示區塊、表單與列表之標籤（以下簡稱區塊標籤），即 $BN = \{ table, tr, td, \dots \}$
BNR_q	目標網頁文件之文件物件模型架構中第 q 個區塊標籤至下一個區塊標籤前所包括之標籤區域（以下簡稱區塊標籤區域）
$N(BNR_q)$	目標網頁文件之文件物件模型架構中第 q 個區塊標籤所包括之標籤區域中標籤節點數量
$N(DOM_T[NT])$	目標網頁文件之文件物件模型架構中標籤節點數量
$DOM_T[BNR_q]$	第 q 個區塊標籤區域中標籤節點數量 $N(BNR_q)$ 占 DOM_T 中標籤節點數量 $N(DOM_T[NT])$ 之比例權重
$BNR_{q,j,k}$	目標網頁文件之文件物件模型架構中第 q 個區塊標籤區域之第 j 個標籤節點位於第 k 個位置
$BNR_q(NT_j)$	第 q 個區塊標籤區域中第 j 個標籤節點與第 1 個標籤節點位置之最小差距值
HB	目標網頁文件之 head 標籤區域
$EB_i(SR, PE)$	第 i 個 CSS 設定區域中選擇器 SR 與屬性設定值 PE
$EB_i(SR)$	第 i 個 CSS 設定區域中選擇器 SR
$EB_i(PE)$	第 i 個 CSS 設定區域中屬性設定值 PE
$N(PE_i)$	第 i 個 CSS 設定區域之屬性設定值 PE 中設定值數量
$W_k[EB_i]$	第 i 個 CSS 設定區域中第 k 個字元
SR_i	第 i 個 CSS 設定區域中選擇器，當 $SR=1$ 時，設定區塊為標籤選擇器； $SR=2$ 時，則設定區塊為 Class 選擇器； $SR=3$ 時，則設定區塊為 ID 選擇器
$PE_{i,j}$	第 i 個 CSS 設定區域之屬性設定值中第 j 個設定值
VA_i	第 i 個 CSS 設定區域中選擇器宣告之變數

$DW_T[BNR_q]$	目標網頁文件中第 q 個區塊標籤區域之修正後權重
ALW_1	第 1 個區塊權重值
TBB_k	目標網頁文件中由上至下第 k 個位置之權重值
$Rnt_1[WD_T]$	目標網頁文件 WD_T 與區塊權重 ALW 之分佈係數
$W_{j,TR}$	分類依據標籤 j 位於網頁空間位置 TR 所分配之權重值，當中 $j \in \{T, B_1, B_2, \dots\}$ ； $TR \in \{TypeIA, TypeIB, TypeIIA, \dots\}$
α_i	含括於標籤 j (網頁空間位置 TR) 之第 i 個外層標籤權重附加值
$Bit_{1,j,t}[WD_T]$	網頁文件中第 1 個重點區塊中標籤 j 之第 t 個外層標籤權重附加值

3.1.1 網頁使用者觀點

本研究參考過去研究建構一套網頁文件區塊分佈權重值設定原則，即 Antti (2005) 指出使用者在瀏覽網頁時，第一眼視線焦點與網頁文件記憶最深刻區域皆為左邊網頁區塊，可由此得知左側區塊重要性；王欽泉 (2007) 從框架型網頁之結構、編排架構與版面設計方式等三方面解析框架型網頁編排設計方式。本設定原則乃根據各區塊分佈位置給予權重設定值，其設定原則分為二部份，分別為從左至右與從上至下順序制訂權重。最後本研究乃判斷目標網頁文件分割方式，若網頁文件分割方式為標籤 frameset 則進入分割判定法則一 (A)。若分割方式為非結構化方式則進行分割判定法則二 (B)。若分割方式為 CSS 語法則進行分割判定法則三 (C)，以獲知網頁文件區塊分佈，當中區塊權重計算公式如公式(1)所示，

$$ALW_1 = \frac{LRT \times \frac{LRW_i}{N(LRW)} + TBT \times \frac{TBW_j}{N(TBW)}}{LRT + TBT} \quad (1)$$

分割判定法則一 (A)：網頁文件 frameset 標籤之分割方式

本法則乃先行擷取網頁文件之 frameset 標籤區域，以解析 frameset 標籤區域中存在分割標籤之字串，並去除當中分割標籤與運算子之符號進而獲知分割欄列字串，之後以分隔符號做分隔進而獲得分割欄列係數，如公式(2)所示。其次判斷標籤區域當中分割方式，如公式(3)所示。其次依照分割方式與分割欄列係數進行轉換，其轉換方式乃先擷取網頁文件中解析度面積之係數，並參照網頁文件之分割欄列係數 (固定值、符號或字元)、frameset 標籤區域存在之集合及其對應之解析度面積，以進行分割欄列係數之轉換，如公式(4)至公式(6)所示，最後判斷網頁文件 frameset 標籤區域數量，以評估修正其解析度面積，進而計算下一個 frameset 標籤區域，如公式(7)所示。待完成係數轉換後即可計算區塊面積，其面積計算乃先行判斷 frameset 標籤區域存在之位置，進而評估分割欄列係數是否去除或修訂解析度面積，以便乘上相對應之解析度面積，因此若標籤區域位至非最後一個區域則進入公式(8)以修訂相對應之解析度面積；反之則進入公式(9)乘上對應之解析度面積。

$$WD_T = \{FT_1, FT_2, \dots, FT_{N_{FT}}\} \quad FT_n = \{W_{n,1}, W_{n,2}, \dots, W_{n,k}, \dots\}$$

$$PC_m \text{ exist in } W_{n,k} \quad \forall k \text{ then } FT_n \in Seg_Set_m \text{ and } F_n = W_{n,k} - PC_m - RemS \quad \forall m = 1, 2 \quad (2)$$

$$F_n = \{F_{n,1}, F_{n,2}, \dots, F_{n,N_f}\}$$

$$IF WD_T \cap Seg_Set_1 \neq \phi \text{ and } WD_T \cap Seg_Set_2 \neq \phi \text{ then } FPC = Seg_3 \quad (3)$$

$$\text{otherwise IF } WD_T \cap Seg_Set_m \neq \phi \text{ then } FPC = Seg_m \quad \forall m = 1, 2$$

$$SegS_1 \text{ and } SegS_2 \text{ not exist in } F_{n,j} \quad \forall j \text{ then } F'_{n,j} = F_{n,j} \quad (4)$$

$$SegS_1 \text{ exist in } F_{n,j} \quad \forall j \text{ and } FT_n \text{ exist in } Seg_Set_m \text{ then } F'_{n,j} = F_{n,j} \times Max_m \quad \forall m = 1, 2 \quad (5)$$

$$\text{SegS}_2 = F_{n,j} \forall j \text{ and } FT_n \text{ exist in Seg_Set}_m \text{ then } F'_{n,j} = \text{Max}_m - \sum_{\text{all } j} F'_{n,j} \forall m=1,2 \quad (6)$$

$$N_{\text{FT}} > 1 \text{ and } FT_n \text{ exist in Seg_Set}_m \text{ then } \text{Max}_m = F_{n,N_{\text{F}}} \forall m=1,2 \quad (7)$$

$$FL_m = \text{MAX}_1 \times F'_{n,j} \forall j \text{ where } n \neq N_{\text{FT}} \text{ and } j \neq N_{\text{F}} \text{ and } FT_n \text{ exist in Seg_Set}_2$$

$$\text{MAX}_1 = F'_{n,j} \text{ where } j = N_{\text{F}} \text{ and } FT_n \text{ exist in Seg_Set}_2 \quad (8)$$

$$FW_m = \text{MAX}_2 \times F'_{n,j} \forall j \text{ where } n \neq N_{\text{FT}} \text{ and } j \neq N_{\text{F}} \text{ and } FT_n \text{ exist in Seg_Set}_1$$

$$\text{MAX}_2 = F'_{n,j} \text{ where } j = N_{\text{F}} \text{ and } FT_n \text{ exist in Seg_Set}_1$$

$$FT_n \text{ exist in Seg_Set}_m \text{ and } n = N_{\text{FT}} \text{ then } FW_m = \text{MAX}_m \times F'_{n,j} \forall j, \forall m=1,2 \quad (9)$$

分割判定法則二 (B)：網頁文件非結構化之分割方式

本法則乃解析網頁文件非結構化之分割方式。DOM Tree 架構 (Document Object Model Tree 即網頁文件之標籤區域以樹狀結構表示) 由多個區塊標籤區域所構成。因此本法則乃先行擷取網頁文件之標籤區域，並利用標籤區域權重分配原則以判定標籤之網頁空間位置，進而參考標籤空間位置關係將網頁文件轉化為 DOM Tree 架構。待完成 DOM 轉換後，接著擷取當中區塊標籤至下一個區塊標籤前 (區塊標籤彙整表如表 1)，以獲知區塊標籤所包括之標籤區域，進而計算區塊標籤區域之標籤節點數量所對應權重，如公式(10)所示。最後本法則計算區塊標籤區域中標籤節點位置與 DOM 標籤節點位置之差距值，以取得其中最小差距值，如公式(11)所示。當中 $\text{NTD}(\text{NT}_{j,k}, \text{NT}_{1,1})$ 為第 j 個標籤節點位於第 k 個位置與第 1 個標籤節點位於第 1 個位置之差距值。

表 1 區塊標籤彙整表

統一標籤名稱	標籤類型	HTML 標籤
區塊標籤 BN	區塊式標籤	<TABLE>、<TBODY>、<CAPTION>、<DIV>、、<IFRAME>
	表單式標籤	<FORM>、<FIELDSET>、<EGEND>、<INPUT>、<TEXTAREA>、<SELECT>、<OPTION>、<LABEL>
	列表式標籤	、、、<DL>、<DT>、<DD>

$$\text{WD}_T = \{BNR_1, BNR_2, \dots, BNR_q, \dots\}$$

$$\text{DOM}_T[BNR_q] = \frac{N(BNR_q)}{N(\text{DOM}_T[NT])} \quad (10)$$

$$BNR_q = \left\{ \begin{array}{l} NT_{1,1}, NT_{1,2}, NT_{1,3}, \dots, NT_{1,k}, \dots \\ NT_{2,1}, NT_{2,2}, NT_{2,3}, \dots, NT_{2,k}, \dots \\ \dots, \dots, \dots, \dots, \dots, \dots \\ NT_{j,1}, NT_{j,2}, NT_{j,3}, \dots, NT_{j,k}, \dots \end{array} \right\}$$

$$BNR_q(\text{NT}_j) = \underset{\text{All } k}{\text{MIN}} (\text{NTD}(\text{NT}_{j,k}, \text{NT}_{1,1})) \quad (11)$$

$$BNR_{q,j,k} = BNR_q(\text{NT}_j)$$

分割判定法則三 (C)：網頁文件 CSS 語法之分割方式

本法則首先擷取網頁文件之 head 與 body 標籤區域，其中 head 標籤區域含 CSS 設定區域乃作為選擇器判斷與屬性設定值擷取，如公式(12)所示。由於 CSS (Cascading Style Sheets 即串樣式列表) 提供選擇器作為宣告方式與屬性設定值設定其網頁文件樣式，因此，本法則其次擷取 CSS 設定區域中選擇器宣告變數所對應之標籤，以計算屬性設定值與設定值數量，進而得知標籤區域之對應設定值數量，如公式(13)所示。另外本法則將 body 標籤區域轉化成 DOM 架構，以判別各個標籤節點與區塊標籤區域，接著利用 CSS 設定區域中選擇器宣告變數比對各個標籤節點，進而獲知此標籤節點對應之 CSS 設定區域中設定值數量，最後修正區塊標籤區域中標籤節點數量占 DOM 中標籤節點數量之比例權重，即得知每個區塊標籤區域之總權重，如公式(14)所示，並參考分割判定法則二

中公式(10)，以得知此區塊標籤區域中標籤節點與每個標籤節點位置之最小差距值，即得知區塊分佈與各區塊所含標籤數量。

$$HB = \{EB_1(SR, PE), EB_2(SR, PE), \dots, EB_i(SR, PE), \dots\}$$

$$SR_i = \begin{cases} 1 & \text{IF } W_1[EB_i] = '.' \\ 2 & \text{IF } W_1[EB_i] = '#' \\ 3 & \text{otherwise} \end{cases} \quad (12)$$

$$VA_i = \begin{cases} EB_i(SR) - W_1[EB_i] & \text{IF } SR_i = 1 \text{ and } SR_i = 2 \\ EB_i(SR) & \text{IF } SR_i = 3 \end{cases} \quad (13)$$

$$EB_i(PE) = \{PE_{i,1}, PE_{i,2}, \dots, PE_{i,j}, \dots\}$$

$$DW_T[BNR_q] = \begin{cases} DOM_T[BNR_q] \times N(PE_i) & \text{IF } VA_i \cap NT_{j,k} \neq \phi \forall i, \forall j \\ DOM_T[BNR_q] & \text{otherwise} \end{cases} \quad (14)$$

3.1.2 網頁設計者觀點

本研究之網頁分類模組是以網頁中文字型資料為分析依據，並考量當中文字語調加強標籤，以作為本研究分類依據之標籤，當中所需探討之網頁標籤彙整如表 2。待分類依據標籤擷取完成後，本研究須進一步探討標籤區域於網頁空間上關係。(許琇娟 2000)指出常見之網頁文件空間規劃係劃分為三種：(1)平行式空間規劃由於標籤區域皆為獨立存在，彼此間不互相影響，因此於此規劃下之標籤區域之權重設定僅參照各標籤重要性即可。(2)巢狀式空間規劃由於一個標籤區域中（稱為外層標籤區域），其所含括之內容除了字彙外，尚包含其它標籤區域（稱為內層標籤區域），導致部分文件內容可能會同時由一個外層與多個內層標籤區域所包括，因此重疊區域中之網頁內容，主要以內層標籤區域權重最高。(3)混合式空間規劃此空間規劃係同時運用平行式與巢狀式標籤區域處理原則，因此於權重計算方式與巢狀式空間規劃相同。

表 2 分類依據標籤列表

標籤類型	標籤名稱	重點文字標籤	
網頁頭部標籤	<title>	標題文字設定標籤 (T)	<h1>、<h2>、<h3>、<h4>、<h5>、<h6>
網頁主體標籤	<body>	粗體文字標籤 (B1)	、
		斜體文字標籤 (B2)	<address>、<cite>、<dfn>、、<i>
		列表法文字標籤 (B3)	、、、<dl>、<dt>、<dd>
超連結標籤	<a href >	導引網路節點標籤 (AH)	

3.1.3 觀點下權重分配

本研究彙整網頁使用者與設計者觀點所得知重點區塊與重要資訊權重，以獲知網重要區塊資訊。首先本研究從使用者觀點中得知區塊分佈權重、區塊分佈位置與其它重要網頁資料，並根據分割法則不同以進行不同計算方式，如公式(15)至公式(17)所示。待得知區塊位於網頁文件之權重後則進入公式(18)，以將網頁文件之各區塊權重正規化。

其次本研究從網頁設計者觀點中得知重疊區域中之網頁內容，主要是以最內層標籤區域之標籤特性呈現，但內層標籤區域為多層標籤區域所包括，故此內層標籤區域的意義將會被適當地增強，因此內層區域亦需附加外層區域之權重值，以加強表現此標籤區域字詞之重要性，是故此區域之權重計算方式如公式(19)與表 3 所示，最後本研究將藉由使用者觀點區塊權重所得知重點區塊權重，與設計者觀點資訊權重所得知重要資訊權重，並且參照重點區塊與重要資訊權重，以得知重要區塊資訊如公式(20)所示，公式當中網頁使用者門檻值 $IbTh$ 與網頁設計者門檻值 $InTh$ 乃制定網頁權重著重之觀點。

表 3 標籤區域權重設定彙整表

標籤代碼(TR)	平行式空間規劃(Type I)		巢狀式空間規劃(Type II)		混合式空間規劃(Type III)		
	A	B	A	B	...	B	...
T	W _{Ti,TypeIA} ，當中<Title>僅出現於 Type I 之 A 區域，故僅考慮此區塊之權重						
B0	W _{B0,TypeIA}	W _{B0,TypeIB}	W _{B0,TypeIIA}	W _{B0,TypeIIB}	...	W _{B0,TypeIIIB}	...
B1	W _{B1,TypeIA}	W _{B1,TypeIB}	W _{B1,TypeIIA}	W _{B1,TypeIIB}	...	W _{B1,TypeIIIB}	...
...
AH	W _{AH,TypeIA}	W _{AH,TypeIB}	W _{AH,TypeIIA}	W _{AH,TypeIIB}	...	W _{AH,TypeIIIB}	...

$$ALW_i = \frac{\frac{FW_m}{MAX_1 \times MAX_2} \times LRW_i}{N(LRW)} + \frac{\frac{FL_m}{MAX_1 \times MAX_2} \times TBW_j}{N(TBW)} \quad \forall m \quad (15)$$

$$TBB_k = TBT \times \frac{TBW_j}{N(TBW)} \quad \forall j \quad (16)$$

$$ALW_i = \frac{LRT \times \frac{BNR_{q,j,k} \times DOM[BNR_q] \times LRW_i}{N(LRW)} + TBB_k}{LRT + TBT} \quad \forall j \quad (17)$$

$$TBB_k = TBT \times \frac{TBW_j}{N(TBW)} \quad \forall j \quad (17)$$

$$ALW_i = \frac{LRT \times \frac{BNR_{q,j,k} \times DW[BNR_q] \times LRW_i}{N(LRW)} + TBB_k}{LRT + TBT} \quad \forall j \quad (18)$$

$$Rnt_l[WD_T] = \frac{ALW_l}{\sum_{ALL\ l} ALW_l} \quad (18)$$

$$W_{j,TR} = \left[1 + \sum_{all\ i} (\alpha_i) \right] \cdot W_j \quad \text{where } 0 \leq \alpha_i \leq 1 \quad (19)$$

$$Bit_{l,j,t}[WD_T] = IbTh \times Rnt_l[WD_T] + InTh \times W_{j,TR} \quad (20)$$

3.2 網頁文件意象風格判定模組

本研究所提出之「網頁文件意象風格判定模組」乃先行蒐集各顏色對應顏色代碼資料，並從標籤區域擷取當中顏色標籤，之後將顏色標籤轉化為 RGB 代碼，以得知顏色代碼與主要分析顏色（主要分析顏色如表 4 所示）之分析係數，以判定顏色接近於特定主要分析顏色；反之，則進行目標顏色代碼轉換。首先此模組所使用之符號定義如下：

表 4 主要分析顏色（參考文獻顏色情緒彙整表）

顏色		16 進位制			對應情緒
編號	名稱	紅	綠	藍	
MC ₁	紅	255	0	0	活躍、興奮
MC ₂	綠	0	255	0	平靜
MC ₃	藍	0	0	255	放鬆
MC ₄	黃	255	255	0	歡樂
MC ₅	青	0	255	255	平靜
MC ₆	紫	255	0	255	神秘
MC ₇	黑	0	0	0	失落、低沉
MC ₈	白	255	255	255	樸素

符號定義

D_i	既有訓練網頁文件庫中之第 i 份網頁文件
C_T	目標顏色代碼
MC_n	第 k 種主要分析顏色
WD_i	第 i 份網頁文件
WD_T	所考量之目標框架型網頁文件
$TA(C_T, MC_n)$	目標顏色代碼 C_T 與主要分析顏色 MC_n 之分析係數
$TC(C_T, DC_r)$	目標顏色代碼 C_T 與顏色 DC_r 之轉換係數
$TC(C_T, DC_g)$	目標顏色代碼 C_T 與顏色 DC_g 之轉換係數
$TC(C_T, DC_b)$	目標顏色代碼 C_T 與顏色 DC_b 之轉換係數
$Rct_n[WD_T]$	目標網頁文件 WD_T 與主要分析顏色 MC_n 之加權後顏色類別分析係數
MAX_r	三原色當中 DC_r 之最大值 255
MAX_g	三原色當中 DC_g 之最大值 255
MAX_b	三原色當中 DC_b 之最大值 255
DC_r	三原色之紅色
DC_g	三原色之綠色
DC_b	三原色之藍色

步驟 (D1): 顏色代碼擷取技術與轉成 RGB 代碼

本步驟乃以網頁標籤區域為分析基礎，首先，本研究建立顏色代碼資料庫，以進行顏色代碼判斷，接著擷取目標網頁文件中顏色代碼，並進行顏色代碼資料庫比對，以判斷此顏色代碼是否為 RGB 代碼，若為 RGB 代碼則進行截斷，反之則進行轉換，最後 RGB 代碼依序截斷後，即可得知紅綠藍之轉換係數，其轉換公式為公式(21)所示。

$$\begin{aligned}
C_T &= 'W_1' + 'W_2' + 'W_3' + 'W_4' + 'W_5' + 'W_6' \\
DC_r &= 'W_1' + 'W_2' \quad DC_g = 'W_3' + 'W_4' \quad DC_b = 'W_5' + 'W_6' \\
TC(C_T, DC_r) &= V[W_1] \cdot 16^1 + V[W_2] \cdot 16^0 \\
TC(C_T, DC_g) &= V[W_3] \cdot 16^1 + V[W_4] \cdot 16^0 \\
TC(C_T, DC_b) &= V[W_5] \cdot 16^1 + V[W_6] \cdot 16^0
\end{aligned}
\tag{21}$$

$$\text{where IF } W_i = \begin{cases} A \text{ THEN } V[W_i] = 10 \\ B \text{ THEN } V[W_i] = 11 \\ C \text{ THEN } V[W_i] = 12 \\ D \text{ THEN } V[W_i] = 13 \\ E \text{ THEN } V[W_i] = 14 \\ F \text{ THEN } V[W_i] = 15 \end{cases}, \forall i = 1, 2, 3, 4, 5, 6$$

步驟 (D2): 計算目標顏色代碼與主要分析顏色之分析係數

此步驟將轉換之紅綠藍係數代入表 5 群組一公式算得目標顏色代碼接近於主要分析顏色紅綠藍，若最大分析係數不超過門檻值 (66%)，則目標顏色代碼可能隸屬主要分析顏色黃紫青，故代入表 5 群組二公式。另外若三顏色係數平等則代入表 5 群組三公式。本研究利用表 5 已初步判定該目標網頁文件接近於主要分析顏色之分析係數。最後參照重要區塊資訊中，顏色代碼標籤位於重要區塊之權重值與分析係數，進而計算加權後顏色類別分析係數 $Rct_n[WD_T]$ ，如公式(22)所示。最後將目標網頁文件中各群組之分析係數彙整並正規化，如公式(23)所示，正規化過後可得顏色類別分析係數 $Rct'_n[WD_T]$ 。顏色類別分析係數 $Rct'_n[WD_T]$ 越大則代表網頁文件越趨近主要分析顏色所對應之情緒。

表 5 目標顏色代碼 C_T 與顏色紅 DC_r 、綠 DC_g 、藍 DC_b 之公式表

顏色 群組	分析 顏色	目標顏色代碼	計算公式
群組一	紅	$TA(C_T, MC_1)$	$\frac{TC(C_T, DC_r)}{TC(C_T, DC_r) \cdot TC(C_T, DC_g) \cdot TC(C_T, DC_b)}$
	綠	$TA(C_T, MC_2)$	$\frac{TC(C_T, DC_g)}{TC(C_T, DC_r) \cdot TC(C_T, DC_g) \cdot TC(C_T, DC_b)}$
	藍	$TA(C_T, MC_3)$	$\frac{TC(C_T, DC_b)}{TC(C_T, DC_r) \cdot TC(C_T, DC_g) \cdot TC(C_T, DC_b)}$
群組二	黃	$TA(C_T, MC_4)$	$\frac{TC(C_T, DC_r) \cdot TC(C_T, DC_g)}{TC(C_T, DC_r) \cdot TC(C_T, DC_g) \cdot TC(C_T, DC_b)}$
	紫	$TA(C_T, MC_5)$	$\frac{TC(C_T, DC_g) \cdot TC(C_T, DC_b)}{TC(C_T, DC_r) \cdot TC(C_T, DC_g) \cdot TC(C_T, DC_b)}$
	青	$TA(C_T, MC_6)$	$\frac{TC(C_T, DC_r) \cdot TC(C_T, DC_b)}{TC(C_T, DC_r) \cdot TC(C_T, DC_g) \cdot TC(C_T, DC_b)}$
群組三	黑	$TA(C_T, MC_7)$	$1 - \frac{TC(C_T, DC_r) + TC(C_T, DC_g) + TC(C_T, DC_b)}{MAX_r + MAX_g + MAX_b}$
	白	$TA(C_T, MC_8)$	$\frac{TC(C_T, DC_r) + TC(C_T, DC_g) + TC(C_T, DC_b)}{MAX_r + MAX_g + MAX_b}$

$$Rct_n[WD_T] = \sum_{all\ i} TA(C_i, MC_n) \times Rnt_i[WD_T] \quad (22)$$

$$Rct'_n[WD_T] = \frac{Rct_n[WD_T]}{\sum_{all\ n} Rct_n[WD_T]} \quad (23)$$

3.3 網頁文件專業類別判定模組

本模組乃以網頁標籤區域為分析基礎，先行蒐集網頁文件當中字串，並擷取關鍵字詞，以進行網頁文件類別判定，若判定結果不明確，則擷取其鏈結網頁，以做進一步判定。故本研究乃提出二階段之網頁文件類別判定，分別為「階段一：網頁文件專業類別判定」與「階段二：鏈結網頁關聯程度推導」，其此模組所使用之符號說明如下：

符號定義

D_i	既有訓練網頁文件庫中之第 i 份網頁文件
$F(N_j, KW_i)$	標籤 j 所包括標籤區域中發生關鍵字 KW_i 之次數
G_k	第 k 種專業類別
KW_i	經關鍵字整併後，關鍵字集合之第 i 個關鍵字
$N(D_j, KW_i)$	訓練網頁文件 D_j 中發生關鍵字 KW_i 之次數
N_j	標籤 j 所包括之標籤區域，當中 $j \in \{T, B_0, B_1, B_2, \dots\}$
$R(G_k, KW_i)$	關鍵字 KW_i 與專業類別 G_k 之隸屬係數
$Rnt'_k[WD_T]$	目標網頁文件 WD_T 與專業類別 G_k 之關係係數
$Rnt_k[WD_T]$	目標網頁文件 WD_T 與專業類別 G_k 之類別隸屬係數
WD_i	第 i 份網頁文件
WD_T	所考量之目標網頁文件

β_j	相關程度遞減排序後之第 j 個鏈結網頁的修正權重係數
LWD_t	目標網頁文件之第 t 個鏈結網頁
$M[LWD_t]$	目標網頁文件 WD_T 與鏈結網頁 LWD_t 之關聯係數
$MRlt_k[WD_T]$	目標網頁文件 WD_T 與專業類別 G_k 之修正類別隸屬係數
$N(WD_T[TITLE])$	目標網頁文件 WD_T 之<title>標籤區域所包含之字數
$Rlt_k[WD_j]$	為選擇之前 S 個鏈結網頁中，排序為第 j 個鏈結網頁 WD_j 與專業類別 G_k 之類別隸屬係數
WD_j	依關聯係數 $M[LWD_t]$ 遞減排序後之第 j 個鏈結網頁
$N(WD_T[TITLE] \cap LWD_t[TITLE])$	目標網頁與第 t 個鏈結網頁之標題文字重複字數比例

階段一：網頁文件專業類別判定

本研究乃參考分類依據標籤列表所界定標籤，將網頁文件資料先行以各網頁標籤進行區分，因此可將網頁文件 WD_j 乃劃分多個網頁標籤區域合併而成，如公式(24)所示。接著運用(孫銘聰、侯建良 2002)關鍵字擷取方法，擷取網頁文件 WD_T 中各網頁標籤區域所包含之關鍵字。其次乃利用網頁文件關鍵字出現頻率、領域關鍵字與專業類別關係之訓練資料庫，以及參照重要區塊資訊，進而計算網頁文件 WD_T 與各類別 G_k 之關係係數，以判定此網頁文件類別偏向，如公式(25)所示。最後因網頁文件之關係係數總和不為 1，因此將網頁文件 WD_T 與專業類別 G_k 之關係係數予以正規化，如公式(26)所示。

$$WD_j = \{N_T, N_{B_0}, N_{B_1}, \dots, N_{B_4}\} \quad (24)$$

$$Rlt'_k[WD_T] = \frac{\sum_{alli} \sum_{allj} \sum_{alll} R(G_k, KW_i) \cdot F(N_j, KW_i) \cdot Bit_{l,j,i}[WD_T]}{\sum_{alli} \sum_{allj} \sum_{alll} F(N_j, KW_i) \cdot Bit_{l,j,i}[WD_T]} \text{ where } j \in \{T, B_0, B_1, B_2, \dots\} \quad (25)$$

$$Rlt_k[WD_T] = \frac{Rlt'_k[WD_T]}{\sum_{allk} Rlt'_k[WD_T]} \quad (26)$$

階段二：鏈結網頁關聯程度推導

本研究乃以鏈結標籤為基礎，並計算網頁 WD_T 與其第 t 個鏈結網頁 LWD_t 之標題文字重複字數比例並參照重要區塊資訊權重值，以計算鏈結網頁之關聯係數 $M[LWD_t]$ ，如公式(27)所示。依據關聯係數之計算結果予以排序，依此排序結果可制定預選等級。最後選定前 S 個鏈結網頁後，由關聯係數最高之鏈結網頁指定權重為 β_s 。本研究乃依此權重修正網頁與各類別之隸屬係數，以求得修正後類別隸屬係數，如公式(28)所示。

$$M[LWD_t] = \frac{N(WD_T[TITLE] \cap LWD_t[TITLE])}{N(WD_T[TITLE])} \cdot W_{AH,TR} \quad (27)$$

$$MRlt_k[WD_T] = \frac{Rlt_k[WD_T] + \sum_{j=1}^S \beta_j \cdot (Rlt_k[WD_j])}{1 + \sum_{j=1}^S \beta_j} \text{ where } 0 \leq \beta_s \leq \beta_{s-1} \leq \dots \leq \beta_1 \leq 1 \quad (28)$$

3.4 多維度網頁文件類別判定

本研究在進行多維度網頁文件類別判定前，需進行「網頁文件區塊與資訊權重分配模組」以得知重要資訊與重點區塊，接著進行「網頁文件意象風格判定模組」與「網頁文件專業類別判定模組」之判定，以得知其專業類別與情緒類別，當中專業類別與情緒類別乃以「網頁文件區塊與資訊權重分配模組」為基礎進行判定，故判定出類別符合網頁文件設計者與使用者所注重之觀點，另外，使用者亦可從多維度網頁文件類別判定結

果得知網頁文件歸屬領域與帶給使用者之情緒意涵；因此，使用者能從多維度網頁文件類別中得知目標網頁文件與專業類別、情緒類別之關係，同時設計者亦能從判定結果得知設計專業領域網頁文件時更適合表達情緒之顏色。因此本研究可藉由專業類別與情緒類別所判定網頁文件多維度類別當中看出多維度意涵，假設專業類別「科技」與情緒類別「紫色」之關係，其判定結果為紫色科技網頁，同時可得知多維度網頁文件類別意義為「令使用者感到神秘的科技型網頁」。

第四章、系統架構與規劃

依據本研究所發展多維度網頁文件分類之方法論，本研究乃於網際網路環境下建構一套多維度網頁文件分類系統（系統運作架構如圖 4 所示）。由於本系統乃架構於網際網路環境下，故可允許多位網頁查詢使用者使用其分類功能。其中使用者可分為網頁查詢使用者與系統管理者，依權限不同分別可執行不同功能：首先網頁查詢使用者可透過使用者資料維護模組維護個人基本資料（如圖 4 之 (A) 流程所示），其次網頁查詢使用者透過網頁文件維護模組上傳欲分類之網頁文件，並可透過網頁文件查詢功能查詢以判定完成之網頁文件多維度類別（如圖 4 之 (B) 流程所示）。而系統管理者除有權限操作與網頁查詢使用者相同之功能外，亦可藉由系統權重模組來修改系統各個參數資料與修訂權重、以及透過關鍵字維護模組、網頁文字/顏色標籤維護模組修改分類依據之關鍵字、標籤、顏色代碼資料（如圖 4 之 (C) 與 (D) 流程所示），將以上模組之資料建立、修訂後即可作為多維度網頁文件分類模式之判定依據，並透過網頁區塊分佈與標籤區域權重分配模組、網頁文件意象風格判定模組與網頁文件專業類別判定模組，以判別網頁文件多維度類別（如圖 4 之 (E) 流程所示）。

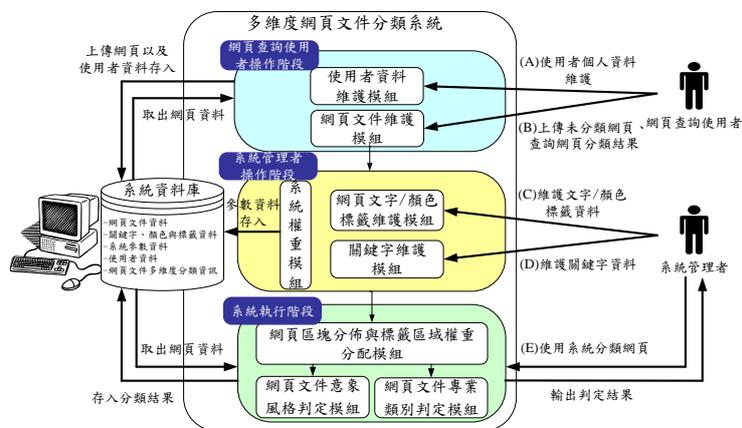


圖 4 多維度網頁文件分類系統運作架構

網頁查詢使用者可透過網頁文件維護模組上傳欲分類之網頁文件，即圖 5 所示使用者乃上傳網頁文件名為「絕地反攻 微軟 window8 開放下載」、類別為「科技」、關鍵字與摘要等網頁基本資料，並瀏覽上傳網頁檔名為「絕地反攻 微軟 window8 開放下載」之檔案，並可檢視所上傳之網頁文件與多維度分類結果。而系統管理者則可透過系統權重設定模組事先設定系統權重值，即設定設定網頁類別判定門檻值「0.5」與鏈結預選等級「2」（如圖 6 所示）等權重參數，接著系統管理者即可藉由多維度網頁文件分類系統，將欲分類之網頁文件勾選並送出後，系統乃先行計算目標網頁中各區塊重點區塊權重以及各標籤區域所包括文字之重要資訊權重，即網頁文件「絕地反攻 微軟 window8 開放下載」當中「網頁文件區塊 1」之重點區塊權重「46%」（如圖 7 所示）與標籤區域「b」所包括字詞「平板電腦」之重要資訊權重「12%」（如圖 8 所示），接著進入不同維度類別判定，若進入網頁文件專業類別判定模組即計算網頁文件之類別隸屬係數，即網頁文件與專業類別「科技」之隸屬係數「45%」，以獲知網頁文件類別「科技」，

模式為基礎乃開發一套「多維度網頁類別判定系統」，讓知識提供者透過多維度網頁類別判定系統，以取得網頁文件多維度類別，進而讓提供者提供網頁予以需求者之參考。

參考文獻

1. 王欽泉，2005，框架型網頁編排設計與意象之研究，國立臺灣師範大學圖文傳播學系碩士論文。
2. 孫銘聰，2002，啟發式電子化文件權限推論模式與技術構建，國立清華大學工業工程與工程管理學系碩士論文。
3. 許琇娟，2000，以漸進式標籤區域分析為基礎之網頁分類器，淡江大學資訊工程學系碩士論文。
4. Adams, F. M., and Osgood, C. E. "A cross-cultural study of the affective meanings of color." *Journal of Cross-Cultural Psychology* (4:2), 1973, pp 135-156.
5. Antti, O., Lari, K., and Jari, L. "Expectations and memory in link search," *Computers in Human Behavior* (21:5), 2005, pp. 773-789.
6. Chen, H., Liu, H., Han, J., Yin, X., and He, J. "Exploring optimization of semantic relationship graph for multi-relational Bayesian classification," *Decision Support Systems* (48:1), 2009, pp. 112-121.
7. Kim, H. J., Kim, J. U., and Ra, Y. G. "Boosting Naïve Bayes text classification using uncertainty-based selective sampling," *Neurocomputing* (67), 2005, pp. 403-410.
8. Lim, C. S., Lee, K. J., and Kim, G. C. "Multiple sets of features for automatic genre classification of web documents," *Information Processing & Management* (41:5), 2005, pp. 1263-1276.
9. Oulasvirta, A., Kärkkäinen, L., and Laarni, J. "Expectations and memory in link search," *Computers in Human Behavior* (21:5), 2005, pp. 773-789.
10. Pearson, R., and Schaik, P. V. "The effect of spatial layout of and link colour in web pages on performance in a visual search task and an interactive search task," *International Journal of Human-Computer Studies* (59:3), 2003, pp. 327-353.
11. Park, H. S., and Noh, S. J. "Enhancement of web design quality through the QFD approach," *Total Quality Management* (13:3), 2002, pp. 393-401.
12. Schiessl, M., Duda, S., Thölke, A. and Fischer, R., "Eye tracking and its application in usability and media research," *Mensch-Maschine-Interaktion* (6), 2003, pp. 41-50.
13. Shugrina, M., Betke, M., and Collomosse, J. "Empathic painting: Interactive stylization through observed emotional state," *Non-photorealistic Animation and Rendering* (6), 2006, pp. 87-96.
14. Wang, J. C., and Day, R. F. "The effects of attention inertia on advertisements on the WWW," *Computers in Human Behavior* (23:3), 2007, pp. 1390-1407.
15. Wang, Y., Phillips, I. T., and Haralick, R. M. "Document zone content classification and its performance evaluation," *Pattern Recognition* (39:1), 2006, pp. 57-73.
16. Wei, C. Y., Dimitrova, N. and Chang, S. F., "Color-mood analysis of films based on syntactic and psychological models," *IEEE International Conference on Multimedia and Expo* (2), 2004, pp. 831-834.
17. Wexner, L. B. "The degree to which colors (hues) are associated with mood-tones," *Journal of Applied Psychology* (38:6), 1954, pp. 432-435.
18. Yang, C. K., and Peng, L. K. "Automatic mood-transferring between color images," *IEEE Computer Graphics and Applications* (28:2), 2008, pp. 52-61.
19. Youn, E., and Jeong, M. K. "Class dependent feature scaling method using naive Bayes classifier for text datamining," *Pattern Recognition Letters* (30:5), 2009, pp. 477-485.
20. Zhang, M. L., Peña, J. M., and Robles, V. "Feature selection for multi-label naive Bayes classification," *Information Sciences* (179:19), 2009, pp. 3218-3229.