

結合知識本體與案例式推理之即時廣告推薦 -以線上交談平台的廣告推薦為例

陳榮昌
朝陽科技大學
資訊管理系
rcchen@cyut.edu.tw

蕭朝維
朝陽科技大學
資訊管理系
s9914617@cyut.edu.tw

摘要

隨著人們對網路的使用日趨頻繁，許多企業開始利用網路來宣傳廣告，但是目前的廣告方式通常都只是在特定的主題平台或是以不分主題的隨機方式來打廣告，如此對某一廣告沒有興趣的人們就會排斥甚至遠離，所以廣告的效益將大打折扣。本研究希望能夠藉由使用者的資訊及使用者間聊天的內容的分析，結合知識本體(Ontology)的技術與案例式推理(Case-Base reasoning)的概念來找出使用者比較有興趣的議題分類，再將相關的廣告推薦給使用者，以達到提高即時廣告效益的目的。

關鍵詞：廣告、知識本體、案例式推理、推薦、聊天議題。

1. 前言

由於網際網路的方便性，越來越多人開始透過網路平台從事各式各樣的活動，例如聊天、網路購物、線上討論等，而在從事這些活動的時候，通常都會有一個推薦的服務機制，因為在這些網路平台上的資源都是非常龐大的，使用者往往無法從如此龐大的資源空間裡快速的找到自己想要的資源，因此，資料的過濾與推薦變的非常重要。推薦就是用來幫助使用者找到符合自己需求的資源，利如廣告推薦、文章推薦、產品推薦等。

而另一方面，在這些網路平台的介面上經常可以看到許多的網路廣告，這些網路廣告有各種廣告的方式，利如橫幅廣告(Banner Ads)、按鈕廣告(Button Ads)、插播

式(Interstitial Ads)、文字廣告(Text Links Ads)、贊助式廣告(Sponsored Ads)等[2]，但由於這些廣告方式通常都沒有針對特定主題來進行廣告，而且都是屬於非即時的廣告方式，所謂的非即時的廣告方式指的是平台管理者事先就設定好要放哪些廣告上去。因此，這樣的方式往往沒有辦法達到良好的廣告效益，因為每一個網路平台使用者彼此的興趣都不一樣，所以不一定會對這些廣告感興趣。

雖然目前有許多關於網路廣告推薦的研究，但是其研究重點幾乎都是放在廣告的推薦精準度上，而忽略掉即時的廣告推薦，例如陶幼慧等學者(2005)利用網路探勘應用於電子商務網站做效能的分析，但是並沒有即時推薦的效果[4]，表 1 為即時與一般廣告之優缺點比較。有藉於此，本研究提出將知識本體的技術與案例式推理的概念結合，並應用到網路的交談平台上，從使用者之間的交談內容分析其中的聊天議題，找出聊天議題的相關廣告有哪些，再將相關的廣告推薦給使用者。因為所推薦的廣告與使用者的相關性提高，所以廣告的效率也提昇了。

表 1 即時與一般廣告之比較

	即時推薦	一般推薦
推薦範圍	小	大
客製化	有	無
廣告吸引力	高	低

2. 文獻探討

2.1 知識本體(Ontology)

「知識本體」又可稱為本體論或實體

論，最早來自於哲學的領域，若以哲學的觀點來解釋的話，可解釋為一切存在於世界的本質，學者(Gruber, 1993)定義知識本體是一種概念化的明確描述[13]，而對於人工智慧的領域來說，知識本體被發展為用來提供機器可處理的資訊來源，並供不同的代理人(Agent)之間溝通[3]。具體來講，知識本體是一種利用概念和階層關係來描述知識的方法，透過知識本體加上推論引擎可以促進不同領域之間的知識分享和重覆使用。許多知識本體被發展來幫助知識的分享和再利用(Gruber, 1995)[14]。因此，知識本體普遍被應用在知識工程、網頁探勘等以人工智慧為分支的領域。知識本體的架構，主要可分為實體(Entities)、概念(Concept)、屬性(Attributes)、關聯(Relations)和公理(Axiom)，所謂的實體指的是一種具體的物件且為單一個體，例如一張桌子、一本書等，概念則是由多個相同的實體所組成，屬性指的是一個概念所包含的特徵，例如魚類這個概念包含了鱗片、鰓等特徵，關聯則是代表兩個概念或是兩個實體之間所存在的關係，例如is a relation、a part of 等關聯(如圖 1)，而公理指的是知識本體的推論規則或限制。

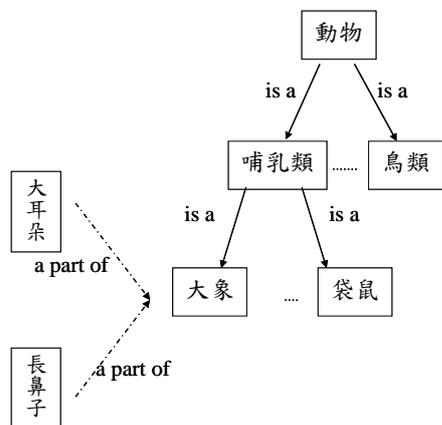


圖 1 本體論的架構-以動物的分類為例

綜合知識本體的特性其優點主要如下所述：

- (1)具有語意的描述能力，可以推論出相關的類別。
- (2)知識本體的屬性可以當作特定資料的特徵，以進行分群與分類。

(3)可以幫助知識的重新利用。

(4)可以擴充不同領域的知識達到知識的共享。這些特性可以用來幫助我們從聊天的重要關鍵字去正確地推論出聊天的議題。

2.2 網路廣告(Display Advertisement)

根據資策會 FIND 對於國內的網路人口調查，2010 的網路人口數達到近 1100 萬人。自從網際網路出現後，無所不在的優勢，使得網路變成廣告主的最愛。另一方面，網路廣告也可以讓廣告主知道自己的廣告的點閱次數，有效的衡量廣告效益。網路廣告市場因此得以快速擴張，是整個廣告產業中成長最快的一環(天下雜誌, 357 期, 2006)，而根據台北市網際網路廣告暨媒體經營協會(IAMA)的統計，在 2010 年台灣整體網路廣告營收市場高達新台幣 78.81 億元，如圖 2。但是，由於網路廣告幾乎都是以隨機的方式來呈現，所以造成點閱率普遍不高的情況，另外，根據網路調查公司 Webtrends(2009)的調查報告，Facebook 的廣告點閱率有漸漸下滑的趨勢，2009 年的點閱率為 0.06%，到了 2010 年的時候則下降成 0.05%，這樣的點閱率遠低於業界中的平均值 0.10%。



圖 2 2007-2010 台灣網路廣告成長趨勢 (圖片來源: IAMA, 2010.1)

而由於網路廣告的盛行，許多學者開始進行關於網路廣告的研究，Fain 等學者(2006)對於贊助式廣告做了調查的相關研究，指出將廣告的內容與使用者意圖做比

較是很重要的[8]，所謂的贊助搜尋意指不按照網路排名而是依據廣告提供者所出的價錢來進行廣告的宣傳，Hung-Chi Huang等學者(2008)提出以分類樹應用到交談內容的分析並用在廣告的推薦上[9]，但是此研究所的方法只能針對英文的交談內容做分析，所以無法處理中文的交談內容。本研究試圖利用Yahoo! directory，自行建立中文的分類樹，並以其為基礎來輔助廣告的推薦。

2.3 案例式推理(Case-Based Reasoning, CBR)

案例式推理是源自人類決策時的一種思考模式，人類的大腦可以視為一個案例式資料庫，內部存在著之前所有已經解決的案例和解決的方法[1]。舉例來說，學生在考試的時候，遇到某個問題會先思考以前是否有寫過相同或是類似的題目，若有就將其解決方法直接拿來使用或是經過修改再使用。而案例式推理的優點是能夠透過歷史記錄快速的找到解決方法並且能夠將方法改進以供未來使用。圖3為案例式推理的循環過程[7]，當我們遇到一個問題的時候，會先到存放案例的資料庫查詢類似的案例，然後進行比對的動作，確定是與目前問題相關之後，就重新利用此案例去找出解決方法，並修改原來的解決方法以解決目前的問題，最後再放入資料庫裡擴充資料庫裡面的案例。

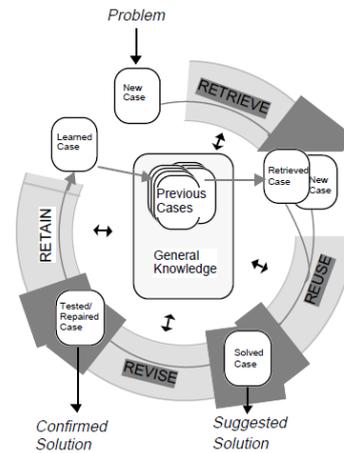


圖3 案例式推理之循環圖[7]

3. 研究方法

3.1 研究目的與系統架構

本研究的目的是將案例式推理的概念與知識本體結合，然後應用到廣告的即時推薦上，而研究範圍是以購物的領域為範疇，並且以線上交談平台做為應用例子，透過知識本體的 is a relation、a part of 等特性來輔助推論出相關的廣告。我們也將 CBR 的概念融入到廣告推薦上，用來分析哪些廣告比較吸引使用者，並將比較符合使用者需求的廣告呈現給使用者。

在圖4的系統架構中，我們每隔一段時間擷取使用者的交談內容，並做斷詞處理、過濾處理、詞彙分析，然後透過知識本體的分類架構找出聊天議題的延伸概念組，最後結合過去的點閱紀錄，以案例式推理及相似度分析來找出與聊天議題較接近的廣告，並將其推薦給使用者。

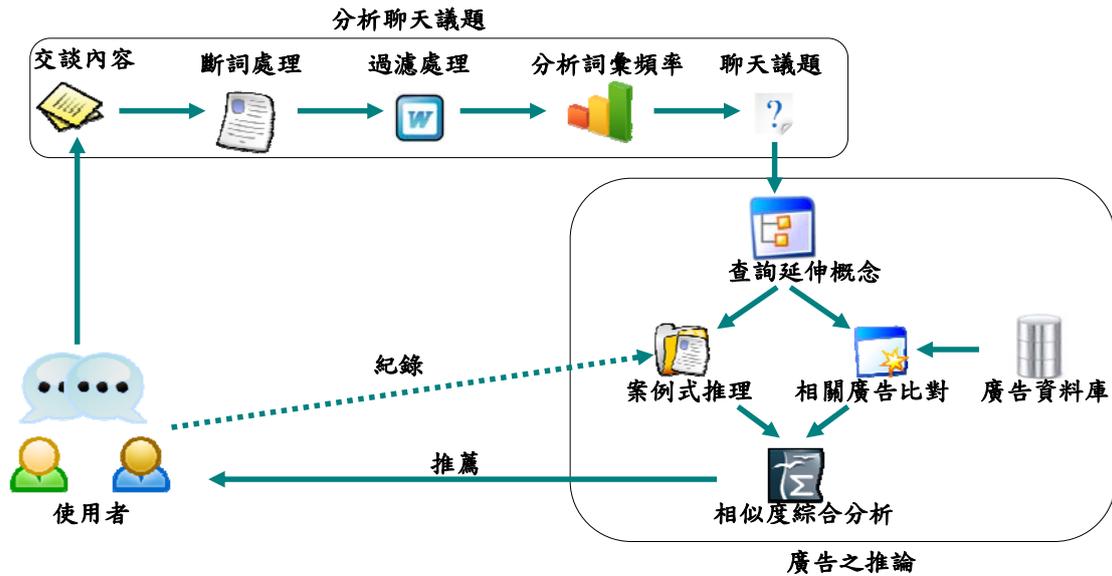


圖 4 系統架構流程圖

3.2 知識本體之建置

知識本體的建置方式主要可分為手動建置和半自動建置，其中手動建置因為是自行建置，所以會產生過於主觀的缺點，而半自動建構的方式有很多種方法，如以字典為基礎的方式、以文字分群為基礎等，但都還是有其優缺點。本研究選擇專家知識本體來做為推論的依據，因為專家知識本體已被多位專家或使用者驗證過，因此，專家定義的知識本體是一個較完整且可靠的知識本體，所以可信度較為其他方式所建立的知識本體高[6]。目前已有許多現成的專家知識本體，例如Yahoo! directory、維基百科分類索引、ODP(Open Directory Project)等，學者Garofalakis和Giannakoudi (2009)利用ODP做為專家知識本體並且用來達到個人化網頁排序的目的[10]，Stuart E. Middleton等學者(2009)在研究中所實驗的推薦系統，Quickstep和Foxtrot分別是基於ODP和CORA兩個專家知識本體所建置而成[12]，本研究則是採用Yahoo! Directory做為我們的專家知識本體，圖 5 為截取自Yahoo! Directory[15]的一部分。我們考量到範圍訂的太大將使得聊天議題的種類落差太大，造成推薦上的困難，因此選擇以「購物」分類的三個階層做為知識本體的領域範圍，並以每個分類底下的網站做為要推薦的廣告，圖 6 為

Yahoo! Directory-分類底下所包含的網站[15]。我們使用由史丹佛大學所開發的Protégé為建置工具，因為其優點包含了簡單易用、支援多種之識本體語言、支援多種儲存格式等[5]。圖 7 是本研究所建置的專家知識本體。



圖 5 Yahoo! Directory[15]



圖 6 Yahoo! Directory-分類底下所包含的網站[15]

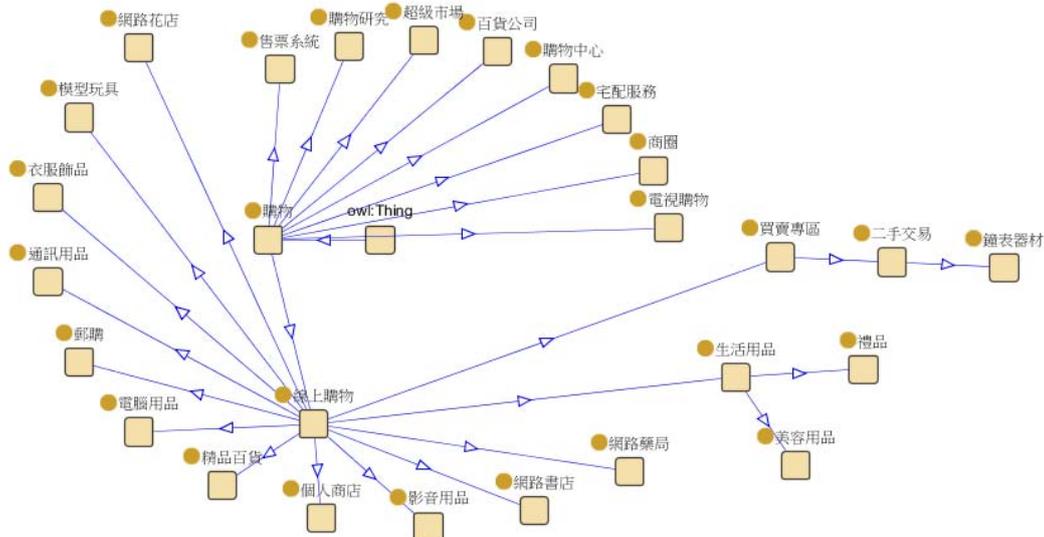


圖 7 專家知識本體之階層圖

3.3 分析聊天議題

本研究在線上交談平台方面是以聊天室做為分析的平台，在研究假設方面是限制使用者在一對一的情況之下，因為若同時有太多的使用者將造成聊天議題過多，使得系統必須推薦太多的廣告，這將造成廣告的推薦失去焦點，而且變的更為複雜，另外也將知識本體的領域定在購物的範圍。在聊天議題分析的部份，主要可以分成以下四個步驟：

步驟一、交談內容擷取

對於交談內容的擷取，我們考量到每一個使用者的打字速度並不相同，而且打字速度的快慢將影響交談內容的數量，所以將以交談時間與交談內容的文字數量做為判斷是否擷取的依據，而關於實驗參數的設定則必須經過更多次的測試才能得知，因此，在交談時間方面，我們暫時設為 1 分鐘，而交談數量將設為 100 個字。

步驟二、斷詞處理

由於中文字詞的數量非常的龐大，而且並不像英文詞彙都有空白來區分字詞，因此，必須先分析出其交談內容包是由哪些詞彙所組成，才能做更進一步的處理。對於斷詞，我們採用由中央研究院所開發的中文斷詞系統 CKIP(Chinese Knowledge Information Processing Group)。CKIP 必須

自行撰寫 Socket 程式然後將欲處理的內容傳送到伺服器端進行斷詞處理，處理完之後客戶端所接收的資料不僅是斷詞之後的結果也包含了每一個詞彙的詞性。

步驟三、過濾處理

經過斷詞的處理之後，我們可以得到交談內容所包含的每一個詞彙，並且能從斷詞的結果了解每一個詞彙的詞性為何，接下來就開始依據詞性和贅字對照表過濾出可能的聊天議題，所謂的贅字指的是沒有意義的文字，其依詞性可主要分為語助詞、代名詞等，如表 2。

表 2 贅字對照表之舉例

語助詞	例如「嗎」、「呀」、「哈」、「呵」等。
代名詞	例如「你」、「我」、「他」等。
問候詞	例如「嗨」、「哈囉」、「早安」、「午安」、「晚安」等。
介系詞	例如「的」、「個」、「之」等。
量詞	例如「條」、「尾」、「片」、「件」等。

步驟四、詞彙頻率之分析

在經過贅字的過濾之後，所留下的字通常都是比較有意義的字詞，因此，我們將這些字詞當作是聊天議題的關鍵字，而接下來，則會開始先分析每一個聊天議題的關鍵字是否有包含在專家知識本體裡面，因為不包含在知識本體裡面主要有兩種可能性，第一種可能性為比較新穎的字

詞，第二種可能性為超過本研究領域範圍的字詞。

當可能聊天議題的關鍵字不包含在知識本體裡面的時候，我們會將此議題移除掉，只留下有包含在知識本體裡面的關鍵字，而接下來，我們則會開始透過 TF(Term Frequency) 計算這些關鍵字的出現頻率，並且將頻率最高的前 3 個關鍵字當作交談內容的聊天議題，然後以此聊天議題到知識本體裡面進行推論與查詢。

3.4 廣告之推論

在知識本體的推論部分，本研究參考蔡旺典在碩士論文中對於知識本體的推論方式[6]，該論文將關鍵字組加上其在知識本體中的父概念、鄰概念和子概念彙整成延伸概念向量(Extended Concept Vector)。延伸概念向量因為涵蓋的觀念更細，比原來的關鍵字組更能夠表達某段文章的涵義。在本研究中，因為鄰概念對議題的影響較小，所以我們只取關鍵字本身及其父概念及子概念而形成延伸概念組。當系統得到聊天內容的關鍵字組後，我們會根據這些關鍵字組從知識本體的階層架構找出直屬的父概念與子概念然後加上原來的關鍵字組，共同組成聊天議題的延伸概念組 $T = \{t_1, t_2, \dots, t_n\}$ 。

另外，所有由廣告商提供的廣告也可以找到其相對應的關鍵字組，我們也依同樣的方法透過知識本體找出直屬的父概念與子概念然後加上原來的關鍵字組，共同組成廣告的延伸概念組 $A = \{a_1, a_2, \dots, a_n\}$ 。有了聊天議題與廣告的延伸概念組(T 和 A)，我們就可以使用相似度運算來找出其相似的程度，我們定義 T 和 A 的相似度為

$$sim(T, A) = \frac{|T \cap A|}{|T \cup A|} \quad (1)$$

因此，我們就可以將與聊天議題相似度最高的廣告推薦給使用者。

另外，我們也將案例式推理的概念融入廣告的推薦上。如果某一個聊天議題(TA)被

談論時，我們推薦的廣告 A 被接受，那就代表談論與 TA 類似的聊天議題時，廣告 A 容易被接受。所以，我們就比對目前的聊天議題 A 與曾經接受的廣告所相對應的議題 TA，如果他相似度很高，也列入推薦的依據。

我們定義其相似度為

$$sim(T, TA) = \frac{|T \cap TA|}{|T \cup TA|} \quad (2)$$

如果某一個廣告非常熱門，則他所相對應的聊天議題可能會不只一個議題(假設其分別為)，那麼，我們可以定義由案例式推理所得到的推薦分數為其最大相似度的值，即

$$sim(T, TA) = \max(sim(T, TA_1), sim(T, TA_2), \dots, sim(T, TA_m)) \quad (3)$$

如果某一個廣告是新廣告或沒有被推薦過的廣告，我們可以定義 $sim(T, TA) = 0$ 。

如此，我們就可以同時考慮聊天議題相似度與案例式推理所得到的相似度來做為推薦的依據。我們以滿足這兩者之一的最大值的廣告來當作我們推薦的對象，即廣告 A 的推薦分數

$$Grade(T, A) = \max(sim(T, A), sim(T, TA)) \quad (4)$$

其中，T 為聊天議題，TA 為廣告 A 曾被推薦成功時的聊天議題。

因此，我們只要將目前得到的聊天議題 T 對每個廣告打分數，得到 $Grade(T, A)$ ，就能夠找到分數較高的幾個廣告，以這些廣告來推薦給使用者。

3.5 廣告推薦流程說明例

此節是在說明廣告的推薦流程，當使用者聊天的時間達到 1 分鐘或是聊天內容達到 100 個字之後，系統會將最近的對話內容擷取起來並連線至 CKIP 做斷詞處理，緊接著會將斷詞結果進行過濾處理，根據贅字對照表將標點符號、贅字等去除掉，然後只留下比較有意義的字詞，再依據頻率最高的關鍵字推論出延伸概念，最

後再經由詞彙頻率的分析找出對話內容的聊天議題。假設判斷出聊天議題包含「生活用品」和「精品百貨」，然後經由知識本體的推論之後，就能得到延伸概念組， $T=\{\text{生活用品, 購物, 線上購物, 美容用品, 禮品, 精品百貨}\}$ 。

接下來，我們會與每一個廣告做相似度的比較，假設有 A1、A2、A3 和 A4 四個廣告， $A1=\{\text{精品百貨, 線上購物, 購物}\}$ 、 $A2=\{\text{百貨公司, 購物}\}$ 、 $A3=\{\text{二手交易, 買賣專區, 線上購物, 購物}\}$ 、 $A4=\{\text{網路藥局, 線上購物, 購物}\}$ ，經過公式(1)的比對之後就能得知延伸概念組 T 和每一個廣告的相似度。分別為 0.5、0.14、0.25 和 0.14。

此外，我們也會比對之前的推薦紀錄，假設有廣告 A1、A2 及 A3 曾經被推薦過，其相對應的聊天議題為 $TA1=\{\text{禮品, 生活用品, 線上購物, 購物}\}$ 、 $TA2=\{\text{美容用品, 線上購物, 購物}\}$ 及 $TA3=\{\text{電視購物, 購物}\}$ ，經過公式(2)的比對之後就能知道延伸概念組 T 和 TA1、TA2 及 TA3 的相似度分別為 0.66、0.5、0.14，那麼，透過公式(4)的計算，四個廣告的推薦值分別為 $\text{Grade}(T, TA1) = \max(0.5, 0.66) = 0.66$ ， $\text{Grade}(T, TA2) = \max(0.14, 0.5) = 0.5$ ， $\text{Grade}(T, TA3) = \max(0.25, 0.14) = 0.14$ ， $\text{Grade}(T, TA4) = \max(0.14, 0) = 0.14$ 。因此，如果推薦值的門檻值為 0.5 的話，我們就可以推薦廣告 A1 和 A2。

4. 結論與未來工作

本研究提出以知識本體結合案例式推理的概念的推薦方式，並將其應用到線上交談平台上，我們分析使用者的交談內容，了解使用者之間的聊天議題，推論出與聊天議題相關的延伸概念組，並依此延伸概念組結合網路廣告的點閱情形，推論出比較符合使用者需求的廣告給使用者。

本研究的即時推薦有別於一般廣告的推薦方式，它可以隨著聊天議題的變化來更換廣告，因此更能達到即時廣告推薦的效果。

在未來的研究中，我們將透過實驗的測試來分析擷取交談內容的最佳時機點，

並比較相關的研究方法，進行效能的分析與評估。

參考文獻

- [1] 吳柏諺，以案例式推理為基礎之高速公路上危險事件預測系統，國立台灣師範大學資訊工程研究所碩士論文，民國 97 年 6 月。
- [2] 周郁菁，在時間限制下，網路廣告促銷方式對網路衝動性購買行為影響之研究，朝陽科技大學企業管理系碩士論文，民國 97 年 5 月。
- [3] 林建宏，正規化概念分析建構電腦病毒特徵之知識本體，國立雲林科技大學資訊管理系碩士論文，民國 95 年 5 月。
- [4] 陶幼慧、莊明達、劉書助，網路探勘應用於電子商務網站之績效實驗，第十一屆資訊管理實務研討會，實踐大學，DEC.2-3，2005。
- [5] 潘紫菁，應用本體論強化軟體技術之知識管理，國立成功大學工程科學研究所碩士論文，民國 95 年 7 月。
- [6] 蔡旺典，建立個人化知識本體來輔助網頁行為探勘 - 以個人化網頁排序為例，朝陽科技大學資訊管理系碩士論文，民國 96 年 1 月。
- [7] A. Aamodt, E. Plaza, "Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches," *AI Communications*. IOS Press, Vol. 7, No. 3, 1994, pp. 39-59.
- [8] D. C. Fain and J. O. Pedersen, "Sponsored Search: A Brief History," *Bulletin of the American Society for Information Science and Technology*, Vol. 32, No. 2, 2006, pp.12-13.
- [9] H. C. Huang, M. S. Lin and H. H. Chen, "Analysis of Intention in Dialogues Using Category Trees and Its Application to Advertisement Recommendation," *The Third International joint conference on Natural Language Processing*, India, 2008, pp. 625-630.
- [10] J. Garofalakis and T. Giannakoudi, "Exploiting Ontologies for Web Search Personalization" *Web Personalization in Intelligent Environments*, Vol. 229, 2009, pp. 49-54.
- [11] P. L. Hsu, P. C. Liu, and Y. S. Chen, "Using Ontology to Map Categories in Blog," *Integrating AI and Data Mining*, Hobart, 2006, pp. 65-72.

- [12] S. E. Middleton, D. D. Roure, and N. R. Shadbolt. "Ontology-based recommender systems," In S. Staab, & R. Studer, *Handbook on Ontologies, Springer-Verlag, Series on Handbooks in Information Systems*, 2009, pp 477-498.
- [13] T. R. Gruber, "A Translation Approach to Portable Ontology Specifications," *IEEE Trans. on Rob.*, Vol. 14, No. 3, 1993, pp. 199-220.
- [14] T. R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," *IEEE Trans. on Rob.*, Vol. 43, No. 5-6, 1995, pp. 907-928.
- [15] Yahoo! directory,
<http://tw.dir.yahoo.com/>.