

從自助旅遊網站的建置實踐自定行程的規則萃取

林青峰* 楊亨利**

國立政治大學資訊管理學系 碩士班研究生*

國立政治大學資訊管理學系 教授**

摘要

隨著日益進步的社會與經濟環境，人們越來越開始重視休閒活動。在以往旅遊資訊較難取得的時候，人們對於如此資訊的需求，往往得求助於朋友、書籍或是旅行業者。這個資訊不對稱的狀況，隨著 Internet 的發展，已經有了明顯的改變。自助旅遊隨著資訊的透明化與旅遊網站的興起，變得有更多人樂於去嘗試了。本研究探討如何建置起一個能夠提供安排自助旅遊資訊的網站，進而討論如何從獲得的自定行程資料中找出一些固定的模式或隱藏於其內的規則。這些模式與規則，可進而幫助經營者提供更符合使用者需求的客製化服務。

關鍵詞：Data Mining、Data Mining、自定行程、自助旅遊

研究背景與目的

根據行政院主計處 92/03/04 的資料，民國 91 年至國家風景區旅遊的民眾超過了 1860 萬人次，這樣的數據較 90 年同期增加了 23%。由國民家庭消費支出結構看來，娛樂教育與文化支出，也由 1985 年的 9.5% 左右增加到 2002 年的 13.3%。這些資料都說明了，隨著社會與經濟環境的日益進步，有越來越多的人重視娛樂與休閒活動。

根據國內自助旅遊推廣者陳德民(1987)所下的定義，自助旅行就是從開始計畫行程到經歷整個旅行過程，一直到旅行結束，完全由自己設計、調配、掌握和實行。想要實行這類的自助旅行，必需先擁有旅行目的地的詳細資訊。這種關鍵性的資訊，以往的旅行者會求助於對目的地熟悉的友人、尋找相關書籍資料或是詢問旅行社。隨著 Internet 的發展，自助旅行的資訊在網路上比比皆是，對於旅行者而言，藉由網路得到旅遊資訊，較舊有的方式來得方便而且經濟。

以目前運作中的旅行網站看來，可以將其分為三類。絕大多數的網站是由旅遊業者(包含旅行社、飯店業者、民宿、娛樂公司或是如 Eztravel 易遊網類型的旅遊入口網站業者)所經營的商業網站，對於旅行者所提供的資訊服務，是較單純的「建議行程」或是建議行程的收集，由於有經營上的考量，各家所提出的建議之中在大同中還藏著小異，如此一來，過多類似的資訊，往往會造成使用者失去決定的焦點。另一種較中立的是由旅遊者所建立的情報交換市集。這類旅遊網站是以情報交換為核心的，主要的經營者可能是因為興趣建立起這個網站，這類網站的參考價值在於它鎖定討論的區域以及參與討論(問／答)的質與量。這種網站雖然會有一些較中立的心得分享，但是隱藏於內的旅遊業者可能會影響到旅行者的決定。最後一種有關旅遊資訊的網站，是政府觀光局或是公園管理處等政府機構所提供的官方網站，這類網站會提供較豐富的交通與旅遊資訊。對於行程安排方面，也有「建議行程」的做法，在規模上雖然較完整，但是因為並未提供關鍵的周邊資訊，還是無法有效的提供旅行者達到自助規劃行程的目的。

上述的情形說明了，日益增加的自助旅行者雖然因為 Internet 的發展而得利，但是也遭遇到了資訊過於散亂的窘境。另一方面，從業者的角度看來，業者也同樣遭遇到一些困難，面對旅行者漸漸採用自助旅行的方式，旅遊業者更需要了解自助旅行者的行為，進而用以改變其經營方式或路線，獲得更高的經營效益。

本研究在這樣的背景之下，試著提出專為為自助旅行者打造的一個旅遊網站模式，這個模式一方面能夠提供旅行者旅遊資訊來協助其進行自訂行程、交通與住宿的安排，補足現有網站的能力；另一方面經由對行程資料做旅遊計劃的資料挖掘，找出旅行者在自助旅行排程的行為模式。旅遊業者可應用找出的規則進行經營方式的改變，而網站經營者則可使用找出的規則進而提供更符合使用者需求的服務。

文獻探討

1.旅遊網站的相關研究

電子旅遊網站的建立是實務界應用 Internet 一個重要領域。近些年國內的旅遊網站如雨後春筍般不斷增多，而國外的旅遊網站也曾幾次的進軍台灣，這顯示旅遊業在國內是有發展潛力的。近些年國內對於旅遊網站相關的研究討論也有慢慢增加的趨勢。

1.旅遊網站的定義

旅遊網站是指透過網際網路，以遊程籌劃、遊程管理、遊程資訊提供等方式，幫助消費者找出不同類型的直接產品提供者，使消費者直接獲取資訊，以及比較分析，並做出購買決策的機構或業者(陳瑞麟，2000)。

2.如何建置一個成功的旅遊網站

對於如何建立起成功的旅遊網站，學者們也提出了許多的討論。黃文生以分析在網際網路上旅遊的動機與消費行為的改變為出發，且運用市場區隔的方式，建議旅遊網站應以單一地理區域及自助旅行做為進入策略。(黃文生，1998)另一方面，吳志伶則是提出旅遊網站經營六個關鍵的成功要素，分別是「低廉的大眾化商品」、「經營體驗型商品的專業知識」、「早期進入者」、「先進的科技能力」、「多元化的產品和服務」與「好的行銷策略」(吳志伶，2001)他並認為網站經營者應建立起一個整合平台，提供一個由消費者百分之百由自行決定行程、團費，成為一個同時具有自由行的彈性以及體驗型商品經營能力的銷售空間。(吳志伶，2001)陳帝仰則是主張網站經營者如依次加強金流、優良商店標章、資訊流及安全性，將可有效的幫助業者取得較大的直接

及間接效益。(陳帝仰, 2003)

蘇惠捷在其對於自助旅行網站的討論中歸納出自助旅遊者所重視的內容與功能分別為為旅遊地的基本資料、旅遊地的進階資料、網站的功能與網站的經營四大類。而他也指出國內目前的旅遊網站總體而言是偏重於網站的功能與網站的經營兩大部分，對於資料的提供仍稍嫌不足。(蘇惠捷, 2002)

3. 旅行網站的經營方法

在過去的文獻中，以討論旅行網站的商業經營方法最多，大概可區分為兩類，一類是針對於網站行銷方式加以討論的，例如：旅行社網際網路之行銷策略研究(陳姿雯, 2000)、旅行業者對網路行銷績效認知之研究(黃正一, 2001)。另一類則是針對旅行網站中的客戶關係來做討論，例如：網站服務品質與顧客忠誠度關聯性之探討 - 以專業旅遊網站為例(黃忠國, 2003)、旅遊網站前端功能與電子化顧客關係管理績效之關係研究(陳帝仰, 2003)。

2. Data Mining

隨著資料收集科技與網路技術的進展，大量使用者所產生的資料被公司或組織有效率的收集起來。但是面對如此巨大的資料量，組織往往無法分析，或是僅能做粗淺的統計資料計算。這樣的現象引發了近年來 Data Mining 這個領域的發展。

資料掘取(Data Mining)或被稱為資料庫裏的知識發現(Knowledge Discovery in Databases, KDD)是將隱性的、顯然還不知道且具有有用潛力的資訊經由一種不簡單(Nontrivial)的方式由資料中萃取出來的過程。而這個知識發掘的過程，會隨著資料庫數量的成長而達到超越人類處理能力的地步。(Frawley et al., 1991)

Data Mining 領域在九十年代興起，對於所能取得的規則和作法有許多精彩的討論(Frawley et al., 1991; Han, et al., 1992; Agrawal and Srikant, 1994; Fayyad 1996)。

Han 的概念樹學習法(Han, Cai and Cercone, 1992)是最近很受到重視

的一種資料掘取演算法。該演算法主要精神是經由各領域專家所建立概念階層(Concept Hierarchy)開始，這個概念階層代表資料意義的抽象層級。最高的概念階層是以 ANY 代表的全部概念，其下會有一至多層的概念層來表達一項資料的抽象層級概念，本文中圖 3 到圖 10 就是概念階層的例子。在資料掘取的過程中，就是依造這種概念樹來不斷的向上累計來進行一般化(Generalized)，一直到一般化的個數滿足預設的一個門檻值為止。最後再整理成人類能了解的邏輯規則。這種方法因為整合了領域的專家，所以找出的規則並不是單純經由資料比對出來的零碎資料，而是真正與領域有結合的知識。

概念樹學習法經由不斷的發展，已有十分豐富的文獻討論。利用這些方法，可以找出的種類豐富的規則。茲將類型與說明整理如下：

- a. 關聯規則：當資料為一個集合時，該規則代表在某個 item 出現的條件下，也同時會出現的 item 的狀況。例如「有 75% 的客戶在購買牛奶之時，又會連帶的購買麵包。」(Han and Fu, 1995)
- b. 特性規則：某特定資料所具備的特性。例如「有 90% 的感冒患者，會出現發高燒、頭痛與流鼻水的症狀。」(Han, Cai and Cercone, 1993)
- c. 區別規則：能明顯分辨目標群體和對照群體所具有的不同特性。例如「區分感冒和水痘的病癥是流鼻水和全身發癢。」(Agrawal, et al., 1992; Agrawal, Imielinski and Swan, 1993)
- d. 演化規則：具有特性與區別規則兩種混合的屬性，一方面，演化規則會討論各階段資料具有什麼樣的特性；另一方面，演化規則也會討論現在和過去的資料如何區別。(Han, Koperski and Adhikary, 1996)
- f. 分群規則：資料庫中依選定的特性做分群。(Chen, Han and Yu, 1996)
- e. 連續模式：能找出計劃中連續動作的固定模式。例如「飛機航班之中所降落機場的規模是先降落幾次的小機場，之後再降落一次大機場的模式」(Han, Yang and Kim, 1999)

客製化旅遊網站的建立

1. 應該的參與建立者

以目前現有的網站經營者來看，建立者大約可分成三類人物：業界經營者、有能力的同好者或是政府機構。這三個角色對於旅遊產業來說，都是有影響力的關鍵人物。但是如果考量可以中立的提供旅遊資訊服務來說，業界經營者很難把持得住中立性的服務指標。若從將資料做分析來了解旅行者的行為看來，對於有能力的同好者來說，又沒有必要性。綜合這二個角度看來，在建置初期，由政府機構成為這個提供服務又兼促進產業發展的角色，是比由另外二種角色建立來的理想。但是，以經營這樣的旅遊網站的立場看來，另外二種角色，也是不可或缺的；如果失去了業界經營者的支持，這樣網站就無法提供如飛機訂位、旅館訂位等整合性質的服務。相對的，一個旅遊網站如果缺乏有能力的使用者參與的話，為了持續維持資訊內容的豐富性，網站經營者將會花上很多額外的力量。所以，要建置和營運一個能夠提供旅行者完整資訊的網站，是需要這三類使用者相互配合與利益共享的。

2. 建立動態排程服務的架構

1. 自助旅行者會遭遇到的問題

對自助旅行的人們來說，最麻煩的事情是要安排整個旅程的行程。這個包含幾點出發、幾點會到達目的地、用什麼方式到達、在一個景點要停留多久的行程表(如表 1)，是自助旅行的旅行者很難處理的問題之一。行程安排者產生困難的地方往往在於：(1)不知道移動到下個預計行程將花費多少時間。(2)不知道這個行程應該合理的分配多少時間。(3)對於這個景點沒有完整的旅遊、歷史、交通、餐宿方面資訊。(4)幾乎沒有費用相關的資訊。

表 1 一般自助旅行的行程表

第一天行程表							
預計到達時間	預計離開時間	行程名稱	行程描述	預計花費		移動到下一個行程的	
				項目	金額	交通方式	預計花費時間
07:30	08:10	集合、出發 (布袋港)	團員們由各地至布袋港集合，搭船前向馬公	船費	1100	船運	90 分鐘
09:40	10:15	CheckIn	完成 checkin 與大件行李的放置，開始正式行程	住宿費	500	機車	30 分鐘
10:45	11:50	山水沙灘	山水沙灘有著名的美麗海灣。人們又稱它豬落水，相傳在古早的時候...	-	0	步行	10 分鐘
12:00	13:00	午餐	至附近的餐廳吃飯	餐費	150	機車	20 分鐘
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

對於一個能夠幫助旅行者自行排程的網站來說，為了能有效率的幫助使用者克服上述的困難，應該提供(a)協助使用者順利安排行程的動態排程功能，並能線上儲存以利後續修改或直接列印出行程表。(b)提供已排定行程中可能會用到有關住宿、交通的聯絡或連結資訊，以協助使用者與業者接觸。(c)提供估計的預算花費範圍，以協助使用者了解支出範圍。如此一來，才能真正的符合使用者的需求。

2. 構念架構的建立

為了實踐這樣的功能，如圖 1 所示，本研究設計了一個客製化行程網站的構念架構。

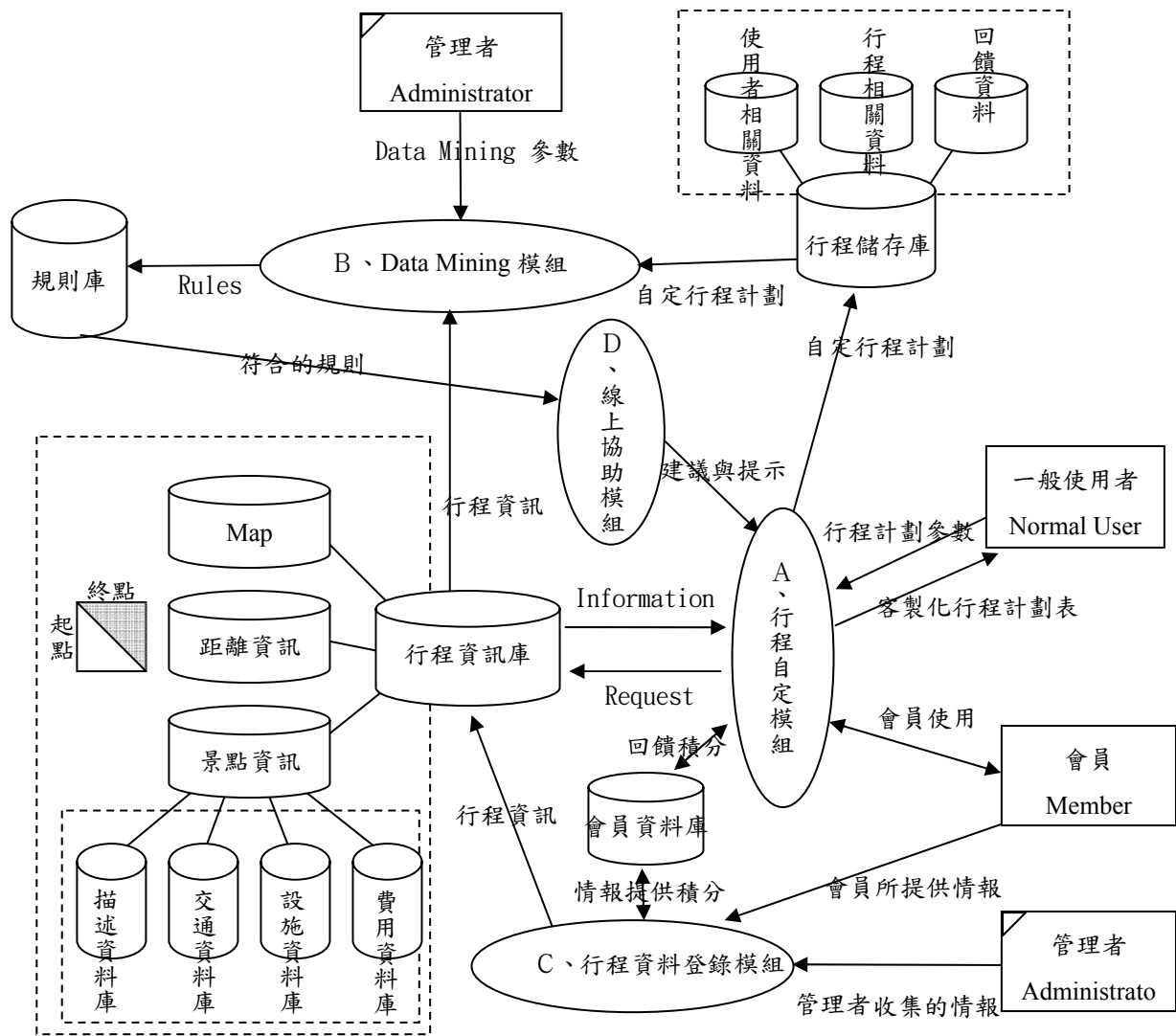


圖 1 客製化行程網站的概念架構

以下依圖中各符號代號，分別就功能與構成解釋之。

A.參與者：在圖 1 中是以方框表示，可區分成以下三類：

I.一般使用者：對於本系統有興趣的一般使用者，不希望參與討論或提供情報，只希望能夠使用本系統協助取得資訊。對於這樣的使用者，系統還是會在其自定行程時對其個人的資訊做較模糊的資料收集。在本架構中，並不期待這類的使用者會進行回饋或情報提供的動作。

II.會員：對於本系統不僅有興趣使用，而且會因為某些原因能夠為系統提供資訊。與一般使用者不同的，會員會提供個人詳細的資料，

也可能會參與情報提供的討論動作；以及對已儲存的自定行程，會員會提供回饋資訊來得到累積得分。會員的組成除了是進階的使用者之外，還可能是旅遊產業的經營者。

III. 管理者：是本系統的經營人員，需要輸入大部份的正規情報；對於會員所提供的情報，管理者會做適當的過濾才使其正式登錄。會員管理以及其它系統層面的管理事宜都需要管理者來處理。Data Mining 模組的使用，可以幫助管理者了解使用者的行為，進而提供改進系統的想法。

B. 資料庫：在圖 1 中是以立體圓柱圖表示，可區分成四大資料庫：

I. 行程資料庫：是主要的內容資料庫，其中包含：

i. Map：用以儲存一區域的地圖，以及各景點在地圖上的位置，是用來實作點選地圖景點位置即能連接到景點資訊的功能。

ii. 距離資訊：用來儲存一區域中各景點到各景點之間的距離，這是一個上三角矩陣形式的資料區塊，當該區域中有 N 個景點時，本資料區塊將應該有 $N(N-1)/2$ 筆距離資料。這個資料，是用以對景點與景點間移動的時間做預估之用。

iii. 景點資料庫：這是旅遊資料庫的核心。其中的描述資料庫，對該景點的名勝、特點或歷史風情有文字及多媒體的介紹展示，對於提供使用者了解該景點有直接的幫助。而交通資料庫，可以幫助使用者了解該景點的交通狀況，利用什麼方式可以到達該景點，以及交通業者的聯絡方式和交通資訊聯結(如火車時刻表的聯結)。設施資料庫則是儲存某景點有提供什麼旅遊服務與其聯絡方式，例如一般的餐飲、住宿提供，或是如露營地提供、賞鯨之旅之類的服務。費用資料庫則會配合設施資料庫與交通資料庫對於各項旅遊項目有個粗略的報價，這個功能可以提供使用者了解大略的預估費用。

II. 行程儲存庫：為了能夠了解使用者自定行程的行為，系統會將使用者或會員所創作的自定行程儲存在行程儲存庫中，最後這個儲存

庫的資料，將成為實施 Data Mining 的原料。行程儲存庫包含三個部份：

- i. 使用者相關資料：這部份的資料是針對儲存非會員的一般使用者資訊所設計的。這個部份要收集的資料有：使用者的性別、居住地、教育程度、職業以及月收入狀況。在表 2 中，ProcessId 是主鍵，用以對每一個自訂行程排班做編號。

表 2 使用者相關資料 Table

ProcessId	性別	居住地	教育程度	職業	月收入
P000001	Male	台北市	大專	資訊相關	50,000
P000002	Female	桃園市	研究所	教育相關	70,000

- ii. 行程相關資料：這部份是一個旅遊計劃真正的資料內容，也就是系統經由使用者一連串的設定之後，所產生出的旅遊計劃實際儲存格式。實體的資料表格設計如下：(在表 3 Table B 中所表示的 EvenId，是對每個排定的行程做編號，其內含包括 ProcessId+Day(天數)+Order(順序數)，如 P000002 的第一天第三個景點的 EvenId 為 P0000020103)

表 3 行程相關資料 Table

Table A: 行程基本資料

ProcessId	出發地	會員 ID	區域別	類別	人數	日期	天數
P000001	台北市	-	太平山	家庭	4	2003/8/1	3
P000002	桃園市	-	阿里山	個人	1	2003/8/5	4
P000003	台北市	M00001	澎湖	團體	25	2003/9/8	4

Table B: 行程計劃資料

EvenId	ProcessId	P000001	P000001	P000001	P000001	P000001	P000001
	Day	01	01	02	02	03	03
	Order	01	02	01	02	01	02
景點 ID		TP0001	TP0004	TP0003	TP0008	TP0015	TP0099
到達時間		20:20	22:50	07:30	09:00	09:00	11:30
離開時間		20:50	23:00	08:00	23:00	09:30	-
述敘		準備前往 太平山	CheckIn	起床	太平山 露營	準備返回	返家
移動時間		120	-	60	-	120	-
移動方式		自行開車	-	步行	-	自行開車	-

Table C: 行程預算資料

CostID	EvenId	CostItem	CostValue
C0000000000001	P0000010101	油資(來回)	500
C0000000000002	P0000010102	住宿費	1200
C0000000000003	P0000010202	租露營地費用	1000
C0000000000004	P0000010202	露營其他費用	1000

iii. 回饋相關資料：這部份的資料，是由會員所提供，會員對於實行計劃後，對於事後行程的設計評估資料。

表 4 回饋相關資料 Table

ProcessId	最好 景點	最好 天數	最壞 景點	最壞 天數	自我 評分	WhyGood	WhyBad
P000003	PH00038	02	PH00012	NA	75
P000008	AL00005	03	NA	NA	90
S000101	NA	NA	SC00003	ALL	15

III. 會員資料庫：用以儲存會員的個人資料以及其供獻度的資料庫。

IV. 規則庫：用以儲存利用 Data Mining 所找出來的規則。除了供管理者了解使用者的行為之外，還可以供線上協助模組提供建議給使用者。

C. 模組：在圖 1 是以橢圓形的方式來表示，包含三大模組：

I. 行程自定模組：對於使用者(或會員)所提出的自定行程，行程自定模組會採用對話填入或圖像選擇的方示來引導使用者建立他的行程。

II. Data Mining 模組：對於已儲存的使用者自定行程，Data Mining 模組將會找到其中存在的規則。

III. 行程資料登錄模組：對於要輸入情報的管理者，情報登錄模組是一個直接的輸入管理介面；而對於要提供情報或要提出疑問的會員而言，這個模組具有留言版的功能，管理者對於會員所提供的情報資訊，會進行資訊的審核及過濾。

IV. 線上協助模組：對於線上排程的使用者而言，線上協助模組將會

參考規則庫內不斷增加的規則對使用者提出排班的建議。

3.3 情境說明

為了明確表達出這個概念架構能為旅行者提供什麼服務，在這裏我們假設了二個使用者：一個是一般的使用者，另一個是已參與會員的使用者。在這裏，將說明這個概念架構會提供這二類使用者什麼樣的服務。

3.3.1 一般使用者的情境

非會員的一般使用者不用登入，即可使用自助旅行的排程系統。首先使用者要先對自我的基本資料做填選(表 2)。接下來使用者要對這次的自助旅行做基本項目的選擇(表 3)，利用出發地、旅遊地區與出發時間的選項，在系統內部將會產生第一天第一個行程(由出發地到目的地)，使用者只要給定出發時間、前往方式，就可以開始進行正式的旅遊行程規劃。旅遊行程排定系統是由三個子頁面所構成的，一個是目前已排定的行程表；另一個是顯示出目前位置與可參觀景點位置的地圖頁面；最後一個是隨景點點選而變動的景點資訊展示頁面。資訊展示頁面由於是綜合了多個子資料庫的資訊，所以會有景點的敘述、歷史由來、周邊的設施、交通狀況的說明、以及到達所需時間(依交通模式不同)與花費的預估值。使用者可以藉由點選地圖頁面的景點標示，並參造資訊展示頁面的說明來考慮是否將點選的景點及參加的設施項目加入行程表中。當使用者決定加入該景點時，系統會詢問使用者交通模式，來決定到達時間。(如果使用者勾選同一交通模式，系統將不會詢問，而當天的所有行程將與第一個景點相同。)當使用者每加入一個景點時，系統就會在行程表頁面增加上一個行程，並且加總行程的預算。當選擇要新增一個景點時，系統會啟動線上協助模助，對於將要新增的景點做出交通、時間、食宿或安排上的建議。當使用者認為安排完一天的行程，使用者可以使用 Call-It-A-Day 的按鍵完結這一天的排班。(或當是行程表派班已超過該天晚上 24:00 時，系統將自動進行完結的工作。)在經過幾天同樣步驟的排班之後，使用者就可以完成排班作業。完成排班作業後，使用者可以直接在線上列印，或是使用線上回寄的功能將排班表寄回指定的信箱裏。

3.3.2 會員的使用情境

會員使用排班系統除了不用填寫個人基本資料之外。(申請會員時資料庫中已有會員的基本資料。)比一般使用者方便的還有線上暫存功能，會員可以將目前的排班表暫存在每人一組的暫存區中，待下次登入後再繼續排班。除了一般使用者可使用的排班系統外，會員還可參與討論區的討論，會員可以提出問題詢問網站經營者或是其它的會員。而回答問題以及有情報要提供的會員也可以經由累積點數的功能來增加自己的點數。點數的累積對於會員的好處除了可經由排名累積人氣指數外，當網站與業者有進行優惠活動時，具有累積點數的會員也可以優先換得。會員還有的一個功能是填寫回饋表的功能。這個功能也可以累積點數。在填寫回饋表(表 4)時會員可以選擇排定行程但未成行的選項，這個功能是為了得到正確的資料，對於選擇未成行的會員，累積點數也會很微量的增加。

Data Mining

1.Data Mining 的目標與過程

本研究所使用的 Data Mining 方式，主要是以 Han 在 1992 年所發展的概念樹學習法。本研究為了發掘發用者的行為模式，用來提供旅行業者或是旅行者在行程排班的建議。採用以下的運作流程來對全體或特定的資料做知識的掘取。

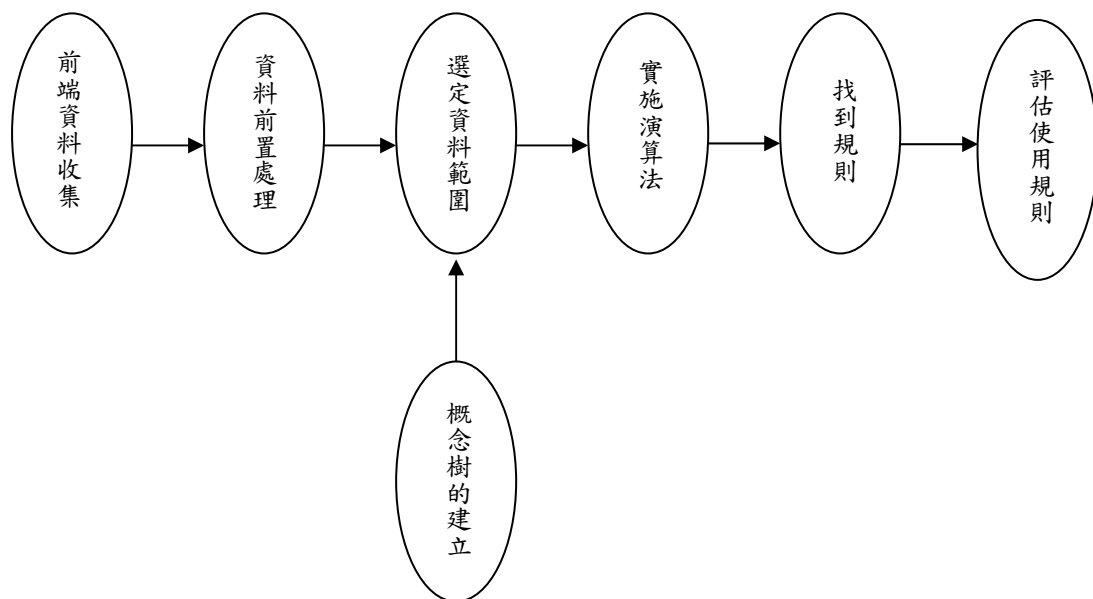


圖 2 本研究 Data Mining 的運作流程

以下，將分別說明各個流程的做法。

2. 資料的前置處理

經由具有動態行程排班功能的前端系統，我們可以直接取得自訂行程及與使用者有關的資料。但這些收集到的資料並不一定都有意義；空值或故意輸入錯誤的資料對於知識發掘一直是一個存在已久的大問題。為了避免這樣的問題過於干擾知識發掘的結果，我們採取三種策略來收集以及處理原始資料：

(1) 利用防錯機制與好的使用者介面確保原始資料不出現邏輯錯誤

資料的邏輯錯誤是常見的一種資料錯誤，而這類的邏輯錯誤，往往是資料收集的過程發生系統錯誤所造成的；例如說收集到的資料中：同一天行程的資料，第一個行程的到達時間，比第二個行程的到達時間還要來的晚。收集資料不合理的現象，是可以利用較好的使用者介面或是防錯機制來預防的。要達成如此的保證，除了系統發展者需要完整的測試之外，應該還要在實際收集資料前先請使用者試用並提供使用者錯誤回報的機制來確保系統可以達到更正確的地步。

(2) 只收集需要、模糊的資料，以降低使用者反彈

另一種難以處理的資料錯誤是使用者故意輸入錯誤的部份。因為使用者填入的資料格式是正確的，但是內容不是事實，這樣的現象造成研究者拿沒意義的資料去做分析。從使用者的角度來看，為什麼輸入錯誤資料的原因不外乎(a)不希望提供較隱私的資訊給它人。(b)題目或問項太多或類似問題重覆，令使用者過於厭煩，導致使用者用亂數給答。對於這樣的問題，除了對於使用者說明收集到的資料僅會用來做增進提供服務的分析之外，另外也應該調整資料的收集項目，不重要的資料就不要收集，牽涉到太過敏感的資料，也應該做較模糊的收集即可。如此一來，可以降低錯誤資料被輸入的狀況。

(3) 對於明顯錯誤的資料，就完全不採用

就算是經由上述二個策略調整過的系統，還是難免會收集到錯誤的資料。故意的錯誤也許是難以發現的，但是為了讓研究的準確性提高，系統發展者會訂定的邏輯關係或規則判斷，一但經判斷出某一筆資料是錯誤資料，對於該資料，本研究是採以較嚴格的直接刪除方式來處

理。

3.概念樹的建立

概念階層提供了控制一般化過程必備的背景資料。(Han, Cai and Cercone, 1992) 在本研究中，我們建立了以下的概念樹：

(1)使用者基本資料相關的概念階層

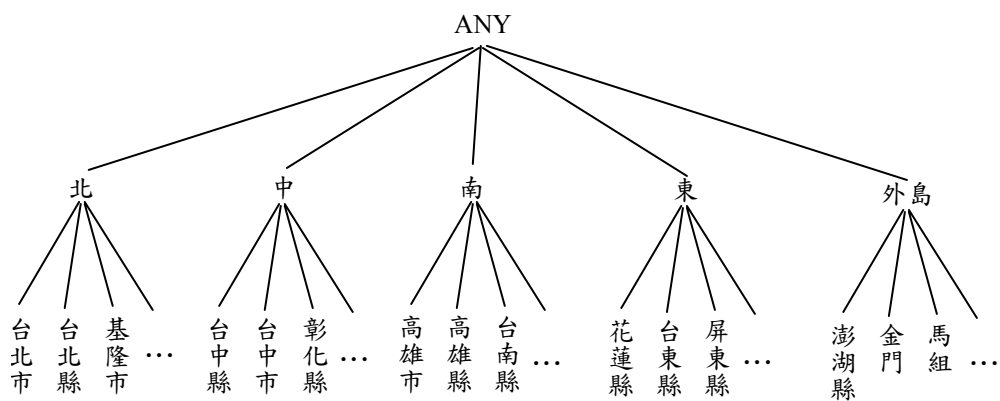


圖 3 台灣地域的概念階層

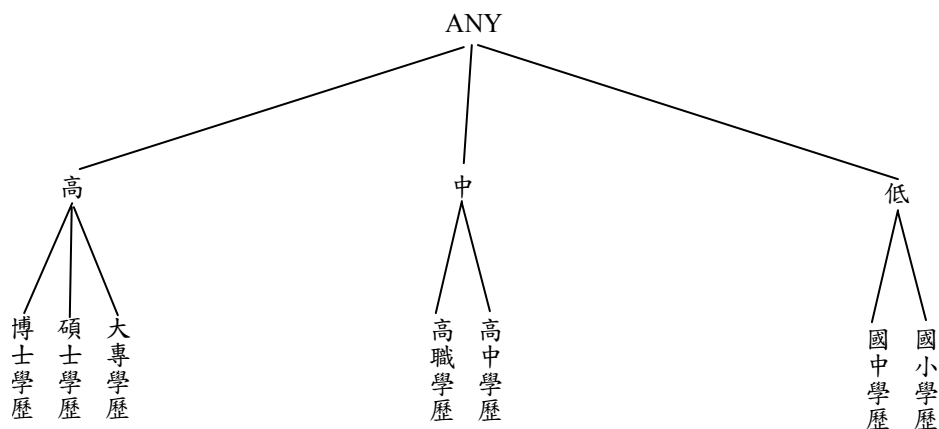


圖 4 使用者教育程度的概念階層

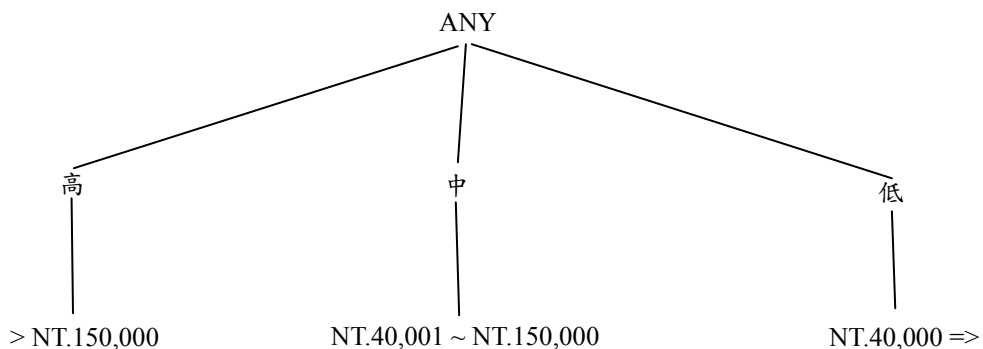


圖 5 使用者家庭月收入的概念階層

(2)行程資料相關的概念階層

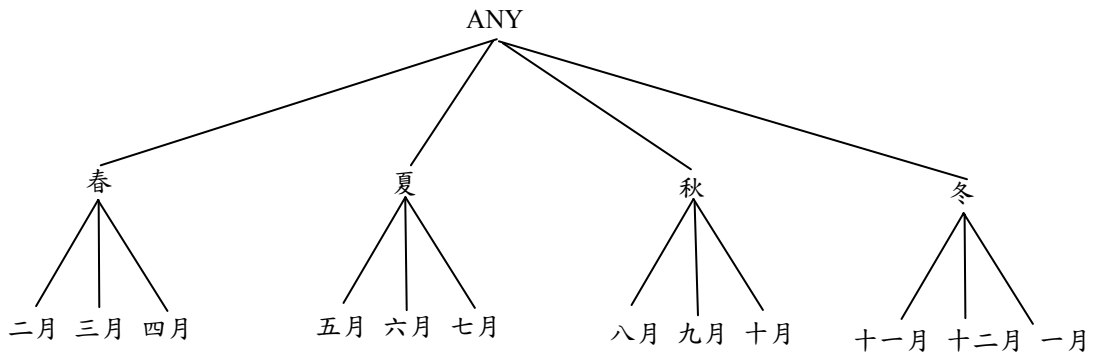


圖 6 日期的概念階層

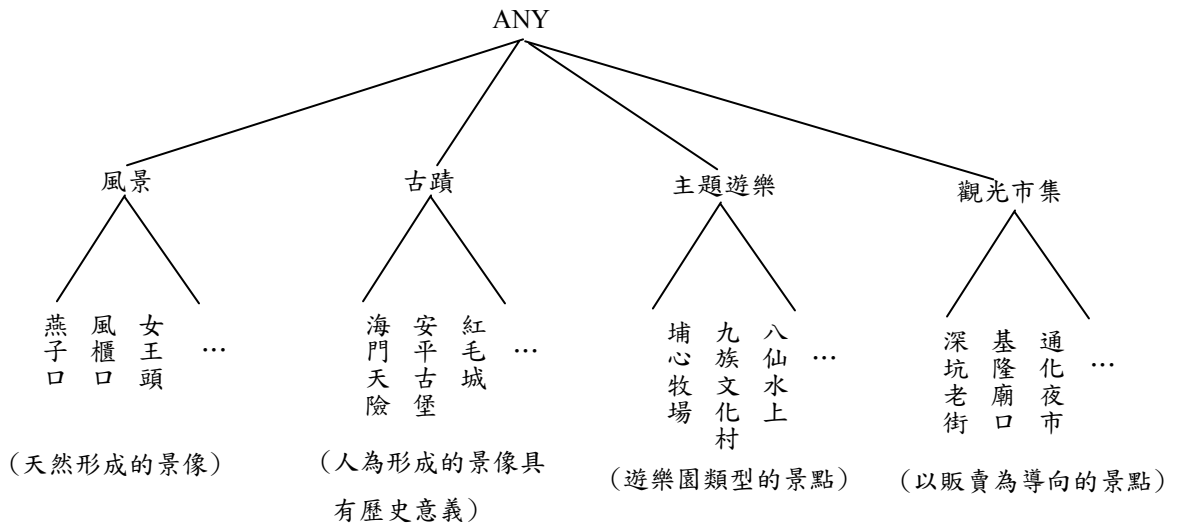


圖 7 景點類別(以旅行性質來分)的概念階層

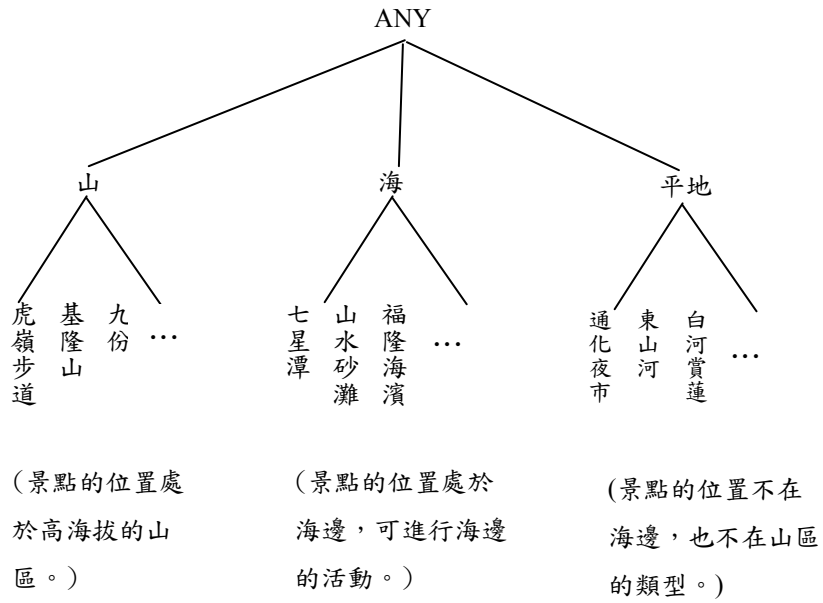


圖 8 景點類別(以景點位置來分)的概念階層

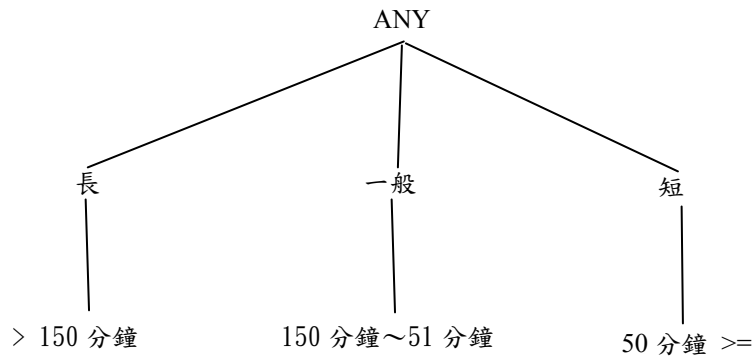
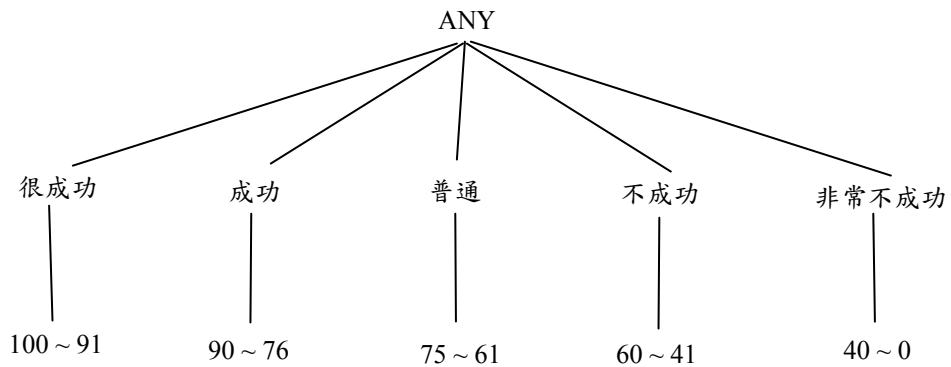


圖 9 景點停留時間長短的概念階層

(3) 回饋相關的概念階層



4. 選定資料範圍

在運行 Data Mining 演算法前，研究者應該對自己想要找出怎麼樣的資訊有一個大略的目標。例如說，我要研究在春、夏兩季間旅行者的景點停留時間長短模式。在這樣的一個問題中，研究者應該先選出滿足在「春、夏兩季」出遊的資料，再交由演算法找出規則。

5. Data Mining 能找出的規則

在一個這樣的架構之下，運用 Han 的概念樹學習法。研究者可以找出許多類型的規則，以下對於可以找出的規則類型和用處做一個介紹：

(1) 行程之間依循的模式

經由分析所給的特定或不特定的資料，我們可以參考圖 7、圖 8 或圖 9 任一類的概念階層來一般化，得到對應的行程模式。例如說：我們參造圖 9 來對所有人的旅行計劃進行一般化，經由 Han 的 Plan Mining 演算概念，可以找出如「有 75% 的人所排的行程，會有『[短]-長』類型的出現；而有 25% 的人所排的行程，會有『中-[長]』類型的出現。」這樣的資訊，這種由一個個順序的景點排程中所找出的規則，可當做之後提供使用者動態建議行程服務的基本資料來源，對於增強網站的能力有正面的幫助。

(2) 旅行者的特性規則

經由分析給定或全部的使用者資料，我們可以找出給定資料的旅行者特性。例如說：我們可以找出「有 60% 使用排程系統的旅行者，具有中等的收入，而且居住在北部；有 20% 使用排程系統的旅行者，具有高等收入，而且居住在南部；30% 的受高等教育的使用者，會在夏季安排去外島遊玩。」這樣的資訊除了可以了解特定一群使用者的狀況之外，對於旅遊業者可以針對使用者的特性進行商品的改良以求更適應使用者的需求。對於網站經營者，這部份的資訊可以用來建置客製

化的主動資訊提供，例如說由上述第三條的規則，當有一個受高等教育的使用者登入網站之後，系統在夏季將會主動提供外島遊玩的相關資訊給使用者。

(3) 景點／旅遊行程的特性規則

經由分析給定或全部的行程與使用者資料，我們可以找出一種景點或旅遊行程所具備的特性。例如說：我們可以找出「有 50%的澎湖自助行程，他的企劃者具有中等的教育程度；而有 95%有澎湖自助行程，會選擇在夏季出發。」這樣的資訊可以提供網站管理者建置建議性的提示服務，例如說有一個使用者指定了澎湖的自助旅行，但是並不是選擇在夏季出發，這時網站可以對使用者提出冬季並不是澎湖的觀光季節如此的警訊。

(4) 旅遊者行程項目的關聯規則

經由分析給定或全部的行程與使用者資料，我們可以找出旅遊者旅遊行為的關聯性。例如說：我們可以找出「有 50%預計前往阿里山的使用者，當他有排定晨起觀日出的景點時，他也同時有興趣搭小火車及進行森林步道」這部份的資訊，可當做之後提供使用者動態建議行程服務的基本資料來源，對於增強網站的能力有正面的幫助。

(5) 旅遊者特性的分群規則

經由分析給定或全部的行程與使用者資料，我們可以將旅遊者分別歸類成幾個不同的分群。例如說：我們可以找出「預計前往阿里山的使用者，有 50%是有受過中等教育且居住在北部的，30%是有高收入的，另外 20%是其它類的」這部份的資訊對於業者而言可以更明白他們潛在的使用者狀況與分佈是什麼。對於網站來說，也可以當作客製化主動提供資訊的資訊來源。

(6) 混合規則

經由混合應用各項規則的資料分析，我們可以找出多變化的混合規則。例如說：我們可以先找出這樣的關聯規則「有 50%預計前往阿里山的使用者，當他有排定晨起觀日出的景點時，他也同時有興趣搭小火車」第二階段，我們再針對符合這種關聯規則的資料做特性規則分析；也就是說第二階段完畢後，我們會找到的規則將是「符合以上關

聯規則的使用者，它具有怎麼樣的特性？」這個問題的答案。利用多階段的應用模式，我們將可以找到十分多樣化的規則。

研究結論與貢獻

1. 研究結論

本研究試著建立起一個幫助旅遊者排定和運行旅行計劃的網站概念架構。並且能夠進一步對其前端所收集的資料做知識掘取的動作。藉由找出來的知識，網站經營者可以更進一步的依使用者類型提供客製化的資訊服務。也可以對使用者在動態排班上過程中，提出一些建議與提示。如此一來，網站將更能引吸住使用者。

對於業者而言，藉由與如此的網站同盟，除了一方向可以擁有站上的客戶之外，另一方面也可以對自助旅行者的行為做更深入的了解。對於自身商品或路線的改良也可以做為改進的依據，不會再是以往由業者方的主觀排定旅程。

這樣網站建立，自助旅行者可以很容易建立起自己的行程排班，對於食、宿、交通、花費等旅遊相關資訊也可很方便的取得。集中而透明資訊和更符合需求的建議，在這個概念模式下，對使用者而言當然是一大利多。

2. 貢獻

在實務上，本研究提出了一個的自助旅遊網站的概念式架構。對於有意建立起服務自助旅行者的業者或機構來說，這個配合知識掘取的架構可以對網站經營者、業者與使用者三者都提供更好的服務。

對於在學術上，本研究探討了一般旅遊網站討論所較不會探討知識掘取技術，並且應用其中 Plan Mining 的想法，從客戶端所取得的自助旅遊者排班記錄中找出行程排班間規則。這對於後續學者在研究自助旅遊者的行為上，提供了一個可以繼續探討的方向。

參考文獻

1. 交通部觀光局統計資料，2002年3月11日。
2. 吳智伶(2001)，旅遊網站經營模式，國立臺灣大學商學研究所碩士論文
3. 黃文生(1998)，在網際網路的旅遊業經營型態，國立台灣大學商學研究所碩士論文。
4. 陳瑞麟(2000)，以消費者為基的網際網路策略—我國旅遊網站之實証研究，國立成功大學企業管理學系碩士論文。
5. 陳帝仰(2003)，旅遊網站前端功能與電子化顧客關係管理績效之關係研究，長庚大學企業管理研究所碩士論文。
6. 蘇慧捷(2001)，從自助旅遊資訊行為探討旅遊網站內容之規劃，國立臺灣大學圖書資訊學研究所碩士論文。
7. 陳姿雯(2000)，旅行社網際網路之行銷策略研究，國立交通大學經營管理研究所碩士論文。
8. 黃正一(2001)，旅行業者對網路行銷績效認知之研究，中國文化大學觀光事業研究所碩士論文。
9. 黃忠國(2003)，網站服務品質與顧客忠誠度關聯性之探討 -以專業旅遊網站為例，佛光人文社會學院資訊學研究所碩士論文。
10. Agrawal, R., Imielinski, T., and Swami, A., "Database Mining: A Performance Perspective," *IEEE Transactions on Knowledge*, Vol.5, No.6 December 1993, pp.914-925.
11. Agrawal, R., and Srikant, R., "Mining Sequential Patterns," *IEEE 11th International Conference on Data Engineering*, Taipei, Taiwan, March 1995.
12. Han, J., Cai, Y., and Cercone, N., "Knowledge Discovery in Databases:

An Attribute-Oriented Approach,” *Proceeding of the 18th VLDB Conference*, Canada, August, 1992, pp.547-549.

13. Han, J., and Fu, Y., “Discovery of Multiple-Level Association Rules from Large Databases,” *Proc. Of 1995 Int’l conf. on Very Large Data Bases (VLDS’95)*, Zurich, Switzerland, September 1995, pp 420-431.
14. Han, J., Yang, Q., and Kim, E., “Plan Mining by Divide-and-Conquer,” *Proc. 1999 SIGMOD’99 Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD’99)* , Philadelphia, PA, May 1999, pp. 8:1-8:6.