

# 行政院國家科學委員會專題研究計畫成果報告

## 電子商務系統之使用機制與資訊搜尋之設計與實作

Design and Implementation of Access Mechanism and Data Mining in Electronic  
Commerce

計畫編號：NSC 87-2213-E-002-101

執行期限：86 年 8 月 1 日至 87 年 7 月 31 日

主持人：陳銘憲 臺灣大學電信中心

### 一、中文摘要

在此一計畫中，我們設計並完成了電子商務系統之使用機制，並對於以網路交易資料為主的商品相關性與消費者行為的相關研究。電子商務是指消費者可以透過網際網路來進行消費的網路交易機制。根據文獻，以往提出的資訊搜尋方法，皆是分析傳統交易方式下所產生的交易資料。我們發現此類型的搜尋方法難以分析電子商務服務所衍生的新型態客戶交易行為，即客戶在網路上的交易路徑，交易網站與交易商品的相關性。因此，我們在此一研究計畫中針對此問題加以深入研究，並提出一個架構來模擬現有的電子商務使用機制以產生完整的消費者交易資料，並設計一系列的演算法來分析此交易資料來對商品相關性與消費者行為進行資訊搜尋。

關鍵詞：網際網路，資訊搜尋，電子商務

### Abstract

In this report, we explore a new data mining capability which involves mining Web-based association rules generated from electronic commerce services in a distributed commodities-provided environment where customers can seek for commodities of interest, travel from one node (i.e., Web page,

whose location is unique in the Internet) to another node via the corresponding facilities (i.e., hyperlinks) provided, and make transactions immediately. It is important to note that, since customers are traveling the commodities-providing Web sites to search for desired commodities and make transactions in the corresponding Web nodes, Web traversal sequences and transaction records are both utilized to mine the Web-based association rule and customer behavior, showing that this problem is very different from the traversal pattern problem and others which are mainly based on customer transaction items.

**Keywords:** Data mining, Electronic Commerce, Internet, World Wide Web

### 二、緣由與目的

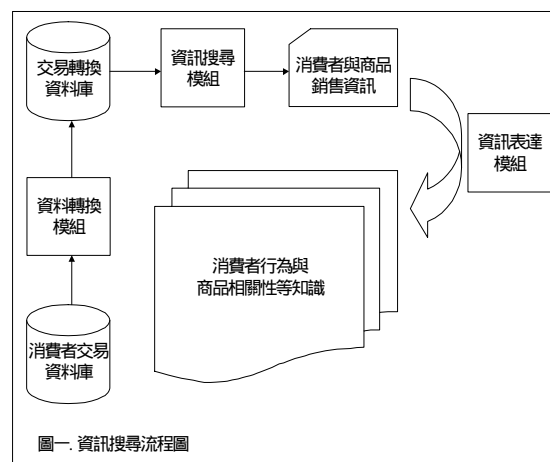
近年來，配合相關技術的研發，WWW 已使得全世界各區域的網路及個人 PC 互相聯結在一起，以免費或極低廉的通信費用相互分享電腦的資源。這些發展已使得網路上的使用人口更是快速提升。國際上許多研發單位已將近兩年網際網路之發展定位為工業革命。在 WWW 應用日益普及的今日，網路購物將成為極競爭之市場。如何在此新興媒體上，建立安全便捷的電

電子商務系統 (Electronic Commerce) 是電子、資訊學界及產業界共同關注的重要課題。而未來的電子商務系統設計勢必由消費者的消費行為所主導。如此一來，分析瞭解電子商務消費者的行為模式已成為極重要的課題。資訊搜尋意指由許多處理過後的資料去尋找新的資訊或學習新的知識。其主要的應用之一便是由消費者的交易資料來分析消費者行為模式。由資訊搜尋所獲得的資訊經分析整理後將可以用以作為系統改進或是輔助商業決策以開發新的商機。舉例而言，瞭解何種人傾向於透過那些網站來購買何種飲料將有助於網路行銷。

此一計畫的主要目的在於設計電子商務之使用機制與設計資訊搜尋演算法以分析商品相關性與消費者行為。舉實際應用而言，我們可以了解那些網站是一般消費者的熱門購物網站，那些網頁上販售的商品具有相關性等。檢視當今已被提出的資訊搜尋方法，我們發現尚未有研究針對這個問題進行探討，現有的方法都是以根據傳統交易資料來設計資訊搜尋演算法。使用此類方法雖然可以獲得消費者資訊。然而，商務應用上之網站設計與商品擺設的分析則極為困難。因此，我們所提出的方法所獲得的資訊將可充分表達電子商務交易的消費者行為與網站商品相關性。

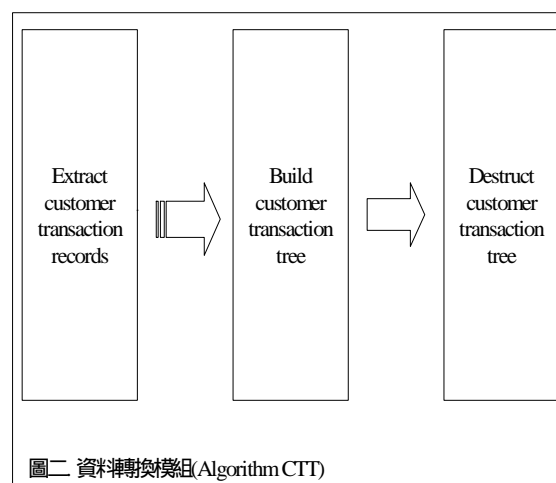
### 三、結果與討論

我們所提出的電子商務系統之使用機制與資訊搜尋之設計可以圖一來表示。消費者交易資料庫內所存的資料乃是根據我們設計的電子商務使用機制所模擬出來的消費者交易資料。根據這張流程圖，要分析此類型的消費者交易資料總共有三個步驟，分別由三個模組來進行：資料轉換模組，資訊搜尋模組與資訊表達模組。



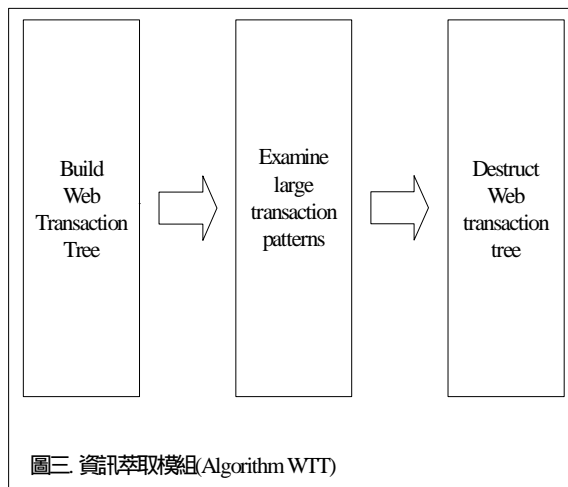
圖一：資訊搜尋流程圖

首先，我們設計了資料轉換模組，其內部演算法稱之為 Customer Transaction Tree (CTT)，將消費者的每一筆網路交易記錄轉換成為另一種資料型態：Web transaction record (WTR)，並將其存入交易轉換資料庫中。當我們將所有的網路交易記錄都轉換成為 WTR 之後，資訊搜尋模組會自交易轉換資料庫中將所有的 WTR 讀出。我們所設計資訊搜尋模組內部的演算法 Web Transaction Tree (WTT) 能自所有的 WTR 中獲得消費者與商品銷售資訊：Large Transaction Patterns (LTP)。當所有的 LTP 被獲得後，資訊表達模組會萃取出 LTP 中所有 rules 的組合並經由 support 與 confidence 的計算加以確認並正確的表達消費者行為與商品相關性等知識。



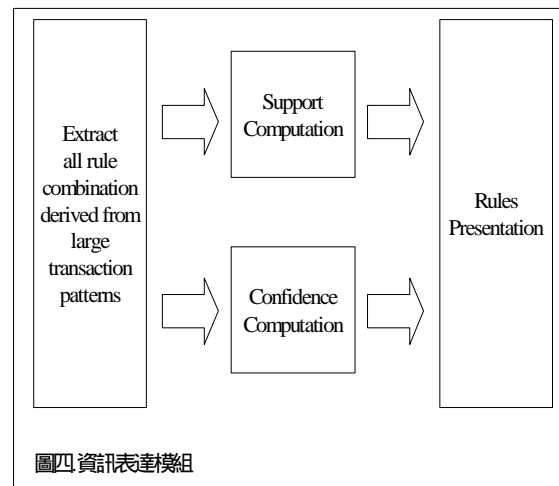
圖二：資料轉換模組 (Algorithm CTT)

我們所設計的資料轉換模組可以圖二來表示。此模組中，主要由三個子模組所構成。首先，子模組一會自消費者交易資料庫中讀出每一筆原始的消費者交易記錄(WT)，自其中萃取出消費者的 Maximal Transaction Sequence (MTS)。然後，子模組二會將每一筆WT的每一個MTS用以建構 Customer transaction tree。當所有的 MTS 都被讀出後，而且 Customer Transaction Tree 也被建構完成之後，子模組三會用 Depth First Traversal (DFS) 的方式以解構 Customer Transaction Tree 來獲取每一筆轉換後的 WTR。



我們所設計的資訊搜尋模組可以圖三來表示。此模組中，同樣由三個子模組所構成。首先，子模組一會自交易轉換資料庫中讀出所有轉換後的消費者交易記錄(WTR)，自其中萃取出所有消費者的 Maximal Transaction Record (MTR)。然後，子模組二會以所有的 MTR 當 input，用以建構 Web Transaction Tree。當所有的 MTR 被讀出後，而且 Web Transaction Tree 也被建構完成之後，子模組三也會用 Depth First Traversal (DFS) 的方式以解構 Web Transaction Tree 來獲取所有的消費者與商品銷售資訊: Large Transaction

Patterns(LTP)。



圖四所表示的為資訊表達模組。此模組中，我們針對每一個 LTP 的資訊，分析所有可能存在的 rules，藉由 support 和 confidence 的計算，獲得此一 LTP 所表達消費者行為與商品相關性等知識。其中，support 所代表的意義是此 LTP 出現在所有的網路交易中的頻率。Support 越高代表此 LTP 出現在網路交易中的次數越多。Confidence 所代表的意義是指在此 LTP 之網站上所賣的物品之相關性的信心指數。Confidence 越高代表其相關性就越高。

#### 四、計畫成果自評

在這個研究中，我們觀察到此一研究之根本問題為網際網路上消費者行為模式的資訊搜尋。針對此一根本問題，我們設計了一電子商務系統之使用機制以模擬並產生出消費者網路交易的資料。然後，我們開發出一新的資訊搜尋技術，可應用於電子商務的消費者交易行為分析與商品相關性分析。此一搜尋技術已經透過程式模擬，數據分析與數理推導的方式加以驗證，並確認其可行性與正確性。我們確信這種搜尋技術可以獲得更完整的電子商務消費者行為與商品相關性資訊，而這些都

是現有的搜尋技術所無法獲取的資訊。

電腦科技與網際網路已將人類生活帶入一嶄新的新紀元，新的工具或是新的技術將對電子商務的經濟型態與消費者行為分析研究的進行有十分重大的影響。因此，電子商務資訊搜尋技術的研究無論在實用上或學理上都具有其價值，我們將繼續利用我們所開發的電子商務交易之使用機制來加以設計新的資訊搜尋演算法以對此類課題加以探討。

## 五、參考文獻

- [1] E. Caramel, S. Crawford, and H. Chen, "Browsing in Hypertext: A Cognitive Study," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 22, no. 5, pp. 865-883, Sept. 1992.
- [2] T. R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association rules in Large Databases," *Proc. 20<sup>th</sup> Int'l Conf. Very Large Data Bases*, pp. 478-499, Sept. 1994.
- [3] J.-S. Park, M.-S. Chen, and P.S. Yu, "Efficient Parallel Data Mining for Association rules," *Proc. Fourth Int'l Conf. Information and Knowledge Management*, Baltimore, Nov. 1995.
- [4] M.-S. Chen, J. Han, and P. S. Yu. "Data Mining: An Overview from a Database Perspective." *IEEE Transactions on Knowledge and Data Engineering*, Vol. 8, No. 6, pp. 866-883, December 1996.
- [5] Patiwat Panurach. "Money in Electronic Commerce: Digital Cash, Electronic Fund Transfer, and Ecash." *Communication of the ACM*, vol. 39, no. 6, pp 45-50, June 1996.
- [6] Christoph Schlueter and Michael J. Shaw. "A Strategic Framework for Developing Electronic Commerce." *IEEE Internet Computing*, vol. 1, no. 6, pp 20-28, Nov./Dec., 1997.
- [7] I.-Y. Lin and M.-S. Chen. "On Methodology for Client-Based User Access Pattern Collection in the Web." *Proceeding of the 11th Conference on Information Networking*, Jan 1997.
- [8] I.-Y. Lin and M.-S. Chen, "Design and Implementation of an Access Pattern Collection Server for Web Data Mining," *Proc. of the Intern'l Symposium on Internet Technology*, April 24-26, 1998.
- [9] J.-S. Park, M.-S. Chen, and P.S. Yu, "Using A Hash-Based Method with Transaction Trimming for Mining Association rules," *IEEE Transactions on Knowledge and Data Engineering*, vol. 9, no. 5, pp. 813-825, Sept./Oct. 1997.
- [10] M.-S. Chen, J.-S. Park, and P.S. Yu, "Efficient Data Mining for Path Traversal Patterns," *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 2, pp. 209-221, Mar./Apr. 1998.