

行政院國家科學委員會專題研究計畫 成果報告

一個具可適性之 Web 資料倉儲系統之研發與實作(3/3)

計畫類別：個別型計畫

計畫編號：NSC92-2213-E-002-001-

執行期間：92年08月01日至93年07月31日

執行單位：國立臺灣大學電機工程學系暨研究所

計畫主持人：陳銘憲

報告類型：完整報告

處理方式：本計畫可公開查詢

中 華 民 國 93 年 12 月 13 日

行政院國家科學委員會專題研究計畫成果報告

一個具可適性之 Web 資料倉儲系統之研發與實作(3/3)

Design and Implementation of an Adaptive Web Warehousing System

計畫編號：NSC-92-2213-E-002-001-

執行期限：92 年 8 月 1 日至 93 年 7 月 31 日

主持人：陳銘憲教授 臺灣大學電機工程學系暨研究所

一、中文摘要

在網際網路成為新一代媒體的今日，如何從龐大且雜亂無章的瀏覽記錄挖掘出有價值的使用者行為，便成為所有網路服務者亟需面對的課題。本研究之主要目的為建置 Web 環境應用所需的資料倉儲，包含資料轉換程序與資料倉儲系統架設，最後整合資訊勘測的諸項技術，包括相關性、分類與瀏覽路徑等，以尋求隱含於大量使用者資訊中的行為模式，此勘測結果可提供使用者個人化的資訊服務，亦可回授至本倉儲系統以動態的調校資料儲存結構，維持系統最佳化，此一具可適性之特點將特別適用於目前線上的 Web 環境。具體而言，整合以上各項技術成為一適合 Web 的應用環境且具線上分析勘測功能之資料倉儲系統是本計畫的首要目標，因此，我們針對資料倉儲系統之需求而研發資料擷取、儲存與勘測機制及其系統架構。

關鍵詞：網際網路，線上分析處理，資料倉儲，資訊勘測，可適性

英文摘要

As Internet technologies develop rapidly these years, more and more people are attracted by various web applications. How to collect and save the user data efficiently and to find the valuable knowledge from the huge data have become important topics. In this project, the high-efficiency data

collection mechanism and corresponding data processing steps for web environments are first being developed. Second, the OLAP techniques are utilized to build a well-found Web warehousing system. Additionally, data mining capabilities, i.e., association, classification and traversal pattern, are devised to explore the user behaviors hidden in large amount of user data. The mining results offer personal services for Web users, and can be feedback for our system to trigger the dynamic adjustment of storage architectures for optimizing the system. This adaptive characteristic is most suitable for online Web systems. In concrete, integrating the technologies mentioned above to build an adaptive web warehousing system with the OLAM functionality is our objective of this project.

Keywords: Internet, OLAP, Data Warehousing, Data Mining

一、前言

近年來，隨著網際網路的蓬勃發展，越來越多的人們接觸並使用網際網路上的各式應用服務，而目前 Web 所提供的平台由於操作容易且普遍具有圖形介面，因此最為人們廣泛使用，像是電子商務、遠距教學、遠距醫療、數位圖書館/博物館等各式各樣的應用便在 Web 環境中應運而生，

在絕大多數的應用中，均有一個重要的共通點，也就是對於資料的儲存及處理相當的重視，除了妥善的規劃安排伺服器端欲提供給使用者的資料外，相對的使用者也會在瀏覽網站時留下許許多多的資訊，針對這些急速增加的使用者資訊，如何整理及有效的儲存已經成為不可忽視的問題，而如何在這些使用者資訊中尋找出更有價值資訊的方法，提供給網站的系統管理員或是公司以更深入的瞭解使用者行為是更具重要性的一項課題。

二、研究目的

針對在 Web 環境中的種種應用所蒐集得到的使用者資料，包括使用者提供的個人基本資料或是瀏覽網頁的過程記錄等，這些資料會快速的累積，因此必須透過有效率的方式來儲存更進而處理，這即為資料倉儲(Data Warehousing)的範疇，在經過整理建構成為資料立方體(data cube)後，便可利用線上分析處理(OLAP, On-Line Analytical Processing)的功能來提供系統管理員對於其使用者的行為有整體性的瞭解，一些統計數據更可簡易便利的查詢得到。

更進一步的若要將這些使用者資料轉化成更具價值的知識，便必須利用資訊勘測的技術，因此本計畫的主題即是針對在 Web 環境中的應用所蒐集產生的資料，整合線上分析處理與資訊勘測的技術以進行分析，最後得到隱含在這龐大資料中的寶貴知識。

具體而言，本研究計畫符合了現今資料庫產業發展的一個重要方向 - 建立決策支援系統(Decision Support System)，而資料倉儲與線上分析處理是決策支援系統的主要部份；決策支援系統不但是目前主要的資料庫發展者的開發重點，同時也已

出現了許多的商業產品與服務。然而，支援決策系統對於資料庫的需求，與傳統的線上交易資料庫有著極大的差異。對於此資料庫技術的發展，本研究計畫將著重於資料倉儲與線上分析處理技術的研究，以解決新的需求；而另一著重點在於適用於 Web 應用的資訊勘測技術的整合，我們將研究一系列資訊勘測方法，在不侵犯客戶隱私及能提高服務品質之前提下，我們計畫要勘測的資訊包括：(1) Association (關連性)，即使用者在 Web 環境中瀏覽/交易行為間之關連性、(2) Classification (分類)，即針對使用者的行為模式予以分類、(3) Traversal Patterns (瀏覽路徑)，即使用者在 Web 環境中之瀏覽網頁路徑習慣。這些資訊之勘測將有助於系統資源的規畫與提供更多的加值性服務。

綜上所述，本研究計畫的目的在於將資料倉儲中線上分析處理的技術與資訊勘測加以整合，成為一個線上分析勘測(OLAM, On-Line Analytical Mining)的系統，此線上分析勘測系統其實就是結合資料倉儲中線上分析處理的技術與資訊勘測的技術，把原始資料(raw data)利用線上分析處理後進而作為資訊勘測的資料來源，並發展線上分析勘測系統使得勘測的過程更具互動性且能讓使用者的主觀判斷更能融入其中，以精準地找出符合使用者需求的知識。

近年來，由於網際網路的快速成長，讓各式各樣的應用服務在其上興起，包括有電子商務、遠距教學、遠距醫療、數位圖書館等。資料倉儲與線上分析處理技術可以有效的統整上述應用服務的使用者資訊，而資訊勘測系統可針對上述應用服務的使用者進行行為模式的分析以提供使用者個人化的資訊服務；結合了線上分析處理與資訊勘測的技術更可以使得這些資料

可以結構化的儲存、呈現並轉化為寶貴的知識。針對 Web 的應用環境，我們期望此一研發建構一個整合性的線上分析勘測系統之研究成果期可為國家實驗網路提供同時深具前瞻技術與實用價值之助益。

三、文獻探討

本計畫所應用之多維度資料立方體技術，其理論基礎、建置技術、與相關之效能增進演算法，在許多文獻中被提出與討論。在 [13] 當中，作者提出了多維度資料立方體之理論基礎與運作方式。在其後的研究中 [10]，延伸先前結果發展出在相關的實體化技術與索引選擇的技術，使多維度資料立方體的運作效率更為精進。另在 [7] 的研究中，亦發展出更高效率的的多維度資料處理技術。

而在線上分析處理與資料倉儲部份，亦有許多相關的研究與應用發展。在 [8][17] 的研究中，資料倉儲的查詢及維護成本的最佳化問題得到了初步的解決。而在管理方面，資料倉儲運作參數設定的問題，在 [24] 的研究中提出了一個啟發式的演算法以獲得較好的運作效率。而在討論 OLAP 技術於資料倉儲中之應用研究時，如何選擇影響具系統運作效能之虛擬檢視表以進行實體化為多維度資料立方體之檢視表選擇問題 (problems of view selection)，在一系列的研究中 [5] [9] [18] [23] [25] 中進行了廣泛及深入地探討。

在另一方面，本計畫相關之資訊勘測的技術，如關連性規則探勘演算法 (Association Rules Mining) 在 [1] 及 [2] 的研究中首先提出，並在接下來數年間針對各式不同的應用環境與效能增進技術的研究上，有十分徹底及豐富的研究成果。而其衍生之分析 Web 環境中瀏覽網頁路徑習慣之瀏覽路徑探勘 (Traversal Patterns) 在

[6] 的研究中，提出了具體而完整的解決方法。

基於以上文獻所提供之理論基礎與研究成果，本計畫結合多維度資料立方體、資料倉儲管理及資訊勘測等資訊技術，進行整合性的研究，以建立線上分析勘測系統。

四、研究方法

本研究計畫發展整合資料倉儲及資訊勘測等相關關鍵技術，並針對 Web 環境的應用服務建立線上分析勘測系統。對此目標，我們首先建立 Web 環境資料搜集的機制，以進行使用者行為分析。其後研究重點著重於研發線上分析處理的技術以有效的將所搜集的資料予以結構化，並研發相關資訊勘測技術用以尋求使用者在 Web 的應用環境之行為模式。最後藉由整合前三項之成果，建立一整合性的 Web 資料倉儲及勘測系統。

首先，建立有效且正確的使用者資訊搜集機制是主要重點，預期可能遭遇的困難在於現有之 HTTP 協定下，Web 伺服器上所搜集到的存取記錄 (access log) 大多不能有效的界定單一使用者之存取行為；為摒除代理快取伺服器 (proxy server) 所提供之快取服務造成使用者的存取記錄皆只記錄相對應的代理快取伺服器此資訊短缺問題，我們藉由先前研究所實作出 APCS (Access Pattern Collection Server) 的技術 [14]，在不侵犯使用者個人隱私及能提高服務品質之前提下，進行有效且正確地使用使用者端資料搜集任務；亦即在首部份的研究方法為利用瀏覽器的 cookie 機制或代理人 (agent) 技術來延伸與加強 APCS 原本的架構與功能，以針對不同的 Web 應用服務，建立相關且完整的 Web 環境資料搜集機制。

在使用者行為分析的研究中，我們藉

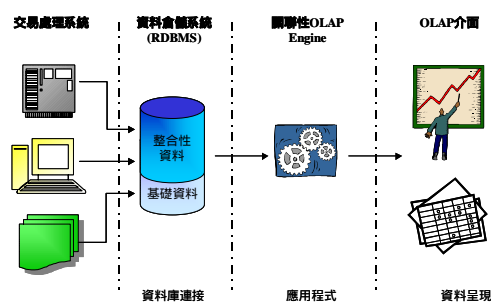
由 Association (關連性)以分析使用者在 Web 環境中瀏覽/交易行為間之關連性，並使用 Classification (分類)技術對使用者的行為模式予以分類。最後應用 Traversal Patterns (瀏覽路徑)之資料技術，探勘使用者在 Web 環境中之瀏覽網頁路徑習慣。此研究成果將有助於系統資源的規畫與提供更多的加值性服務。

在計畫的核心部份，研發線上分析處理與資訊勘測的技術是主要重點。藉由線上分析處理是此軟體技術，可快速地透徹資料內容所代表的資訊，並且可以選擇各種不同的觀點(view)來解釋原始的交易資料。有別於傳統的資料庫系統對於線上交易處理 (OLTP , On-Line Transaction Processing)的處理能力，支援決策系統需有更進一步的資訊系技術。因此我們研究線上分析處理相關技術以處理(1)具歷史性的資料並預估未來的趨勢、(2)合併許多不同異質的資料來源，包括不同的資料格式、資料品質、(3)相關的資料組織技術、存取方法及資料呈現技術。

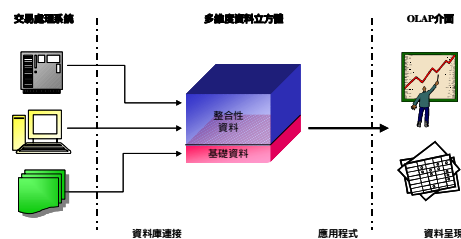
對於上述增進線上分析複雜的查詢及資料呈現能力，本計畫以多維度資料立方體來建立並實作此線上分析處理模型；舉例建立一個銷售量的資料倉儲會包含了以下維度：時間、銷售地區、業務員及產品等以解釋一筆交易資料的內容，通常這些維度內會包含層級從屬關係 (concept hierarchy)，例如時間維度會包含了年、季、月、日，地區維度會包含了縣、市、區，以多維度資料立方體處理線上分析查詢的好處在於其內容已包含了所有的維度構成的資料空間內所有的值，及對各維度的整合(aggregation)運算結果，系統運作上，我們指定了各資料維度的值(value of each dimension)，就可以準確地查詢出該筆交易內容，且可以對於任一或多個維度進行整

合性的查詢，也可進行跨層級的交叉查詢。

對於線上分析處理的技術，本計畫研究以下三種架構：(1)關聯性線上分析處理 (ROLAP , Relational OLAP)、(2)多維度線上分析處理(MOLAP , Multi-Dimensional OLAP)、(3)整合型線上分析處理 (HOLAP , Hybrid OLAP)；ROLAP 的查詢是即時產生，可查詢到最詳細的資料層級，但每次分析查詢都直接存取資料庫，因此效率較低；而 MOLAP 是將預先計算好的多維度資料立方體以可快速查詢的特殊資料結構直接儲存於資料庫，每次就直接對此資料結構進行查詢，因此查詢效率高，但是可查詢的項目是固定的且佔有相當的空間，較不具即時性。而 HOLAP 綜合兩者的特性，將資料層級較高以 MOLAP 的方式便利較大規模的查詢，而對於細部查詢則採用 ROLAP 的方式來即時對資料庫進行查詢。針對不同 Web 應用服務的特性，本計畫研究不同的線上分析處理模型，以應用於即時性的需求及不同使用者人數環境等。線上處理分析的系統架構及多維度線上分析處理的系統架構所圖一及圖二所示。



圖一 關聯性線上分析處理的系統架構



圖二 多維度線上分析處理的系統架構

在資訊勘測的技術與 Web 資料蒐集機制與線上分析處理技術整合方面，計畫以資料倉儲中所建立的資料立方體為基準，在其上進行各種資訊勘測；並將線上分析勘測的過程分為六個步驟：

- (1) 整合(APCS)模組與資料倉儲系統，達成資料搜集與資料載入資料倉儲自動化。
- (2) 建立索引/虛擬化資料表具體化機置。
- (3) 建立與應用目標相符的資料立方體，且所有勘測過程中會用到的維度均會包含於此資料立體塊內，並連結資料倉儲系統與線上分析處理模組。
- (4) 建立完善的線上分析處理的前端使用者介面，提供完整的資料立方體操作功能與資訊勘測功能。
- (5) 進行全系統測試；將資訊勘測結果與 OLAP 模組及 APCS 模組的資料進行驗證、比對，以測試線上分析勘測系統的完整性。
- (6) 系統最佳化；由線上分析處理前端的查詢與資訊勘測結果，對於系統操作參數進行調整，包括資料倉儲的索引/虛擬資料表進行具體化的調整 對於 MOLAP 與 ROLAP 的資料層級調整，及對於資訊勘測演算法的相關參數進行調整，並傳回相關模組進行修正，以達到最佳的分析效率。可使線上分析勘測系統真正成為「可適性」的系統。

藉由本研究計畫的執行，我們研發上述之相關關鍵技術，針對 Web 環境的應用服務建立線上分析勘測系統。

五、結果與討論

本計畫依工作重點及執行時間分為以下三個階段

- (1) 建立在 Web 伺服器上搜集使用者資訊的機制與規劃適用於 Web 環境之資訊勘測技術
- (2) 建置 Web 環境應用所需的資料倉儲
- (3) 整合完整的 Web 資料倉儲系統及系統最佳化

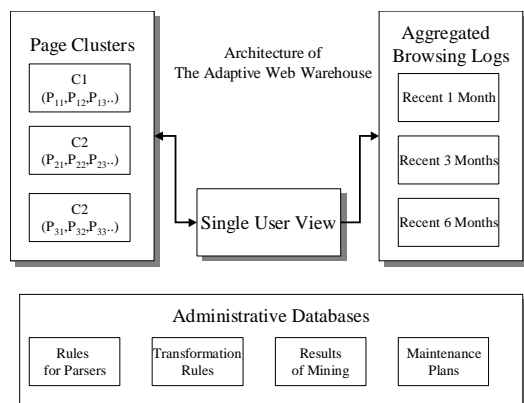
在第一階段計畫中，我們已研發了能整理且有效儲存這些資料的機制，並從這些使用者資訊中尋找出更有價值的知識，以提供給網站的系統管理者參考，以更深入的瞭解使用者行為。相對於傳統媒體，而網際網路卻可在某種程度上識別使用者。若網站實施會員制，更可在使用者登入時，即時認證個人身份。這個差異使得網站服務提供者可針對「網頁存取記錄」，應用新一代的資料儲存技術—資料倉儲(Data Warehousing)與資料分析技術—資訊勘測(Data Mining)，將存取記錄從資料轉化成可精確瞭解使用者行為的寶貴知識。針對在 Web 環境中的種種應用所搜集得到的使用者資料，透過有效率的方式來儲存更進而處理。在經過整理建構成為資料立方體後，再利用線上分析處理的功能來提供系統管理者對於其使用者的行為有整體性的瞭解，一些統計數據更可簡易便利的查詢得到。更進一步的若要將這些使用者資料轉化成更具價值的知識，便必須利用資訊勘測的技術，因此本計畫的主題即是針對在 Web 環境中的應用，研發有效率的資料搜集機制，配合資料清理(data cleaning)、資料轉換(data transformation)與資料整合(data integration)等技巧來建構一 Web 資料倉儲，進而整合線上分析處理與資訊勘測的技術以進行分析，最後得到隱含在這龐大資料中的寶貴知識。

在第二階段研究成果中，以延續第一階段的研究成果與建置資 Web 資料倉儲為主。首先的研究成果為設計 Web 資料倉儲

的實體資料架構。此部份之設計包括

- (1) 單一使用者基本資料，內容為彙整所有使用者的個人資料、聯絡資料、接觸記錄等資料，建立單一使用者的基本資料。此部分為整體資料倉儲的核心。
- (2) 彙整後瀏覽記錄(Aggregated Browsing Log)：根據不同時間區間有不同的記錄項目與彙整方法，產生不同時間的瀏覽資料倉儲。越近期的瀏覽行為，所記錄的資料也越詳細。本階段研究此一倉儲之區間間隔為一個月、三個月、六個月。
- (3) 網頁所屬組群資料(Clusters of pages)：將網頁之關係以樹狀描述並歸類，以進行分析瀏覽行為。本網頁會根據目前最近點選次數自動決定其分類樹狀深度，最近越少點選者其分類越粗略。
- (4) 系統管理資料庫(Administrative databases)：包含前端資訊擷取/處理/轉換模組所使用之規則資料表(rules tables)與其他系統維護及管理所需之資料庫。

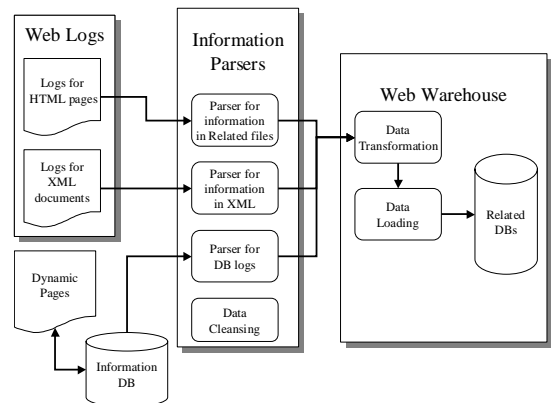
本階段研究設計的 Web 資料倉儲實體架構如圖三所示：



圖三、Web 資料倉儲實體架構

除 Web 資料倉儲的實體資料架構設計外，第二階段研究成果亦包含 Web 記錄檔

與資料倉儲資料處理程序之設計。設計的 Web 資料倉儲處理程序如圖四所示：



圖四、Web 資料倉儲處理程序

Web 資料倉儲處理程序主要分為以下項目：

目：

- (1) 資料來源：

Web 資料倉儲的資料來源主要可分為兩部分，分別為網站所擷取之使用者瀏覽記錄(file access log)與相關網頁向資料庫進行查詢(Query)之記錄。為使資料倉儲系統可適用於各類型網站，系統需具備轉換此形態記錄成為索取資訊記錄(requested information log)之功能。亦即使用者之瀏覽記錄必須加工處理，使之成為結構性資料(structured data)，才可瞭解標籤中資料的意義並存入資料倉儲。

- (2) 資料意義展現與資料萃取：

此部份程序包含資訊剖析及資訊樹狀描述模組。此部份研究根據不同的原始資料型態，如非結構化的 HTML 頁面與半結構化的 XML 文件，設計對應的資訊剖析模組。並將剖析後的資訊瀏覽記錄，描述成樹狀關係，再利用相關描述技術如 Root Labeled Tree(RLB)描述來源資訊。

- (3) 資料處理：

此部份包含資料萃取、轉換及載入模組。各模組根據所需選取所需之資料

來源，並建立規則資料庫，記載所有來源資料所承載的資訊，與對應的目的地表格欄位之關係。為節省儲存空間與增進日後查詢效率，名稱相關資訊皆以對應表(Mapping Tables)進行編碼轉換後進行儲存。最後資料載入程序可將 Web 瀏覽資訊轉入結構化的關連式資料庫 (RDBMS) 中。透過前述之各程序，瀏覽行為所對應之相關資訊，可存入相關之瀏覽記錄資料表與單一使用者資料庫。各頁面相關之資訊樹狀描述及統計資料，也一併進行載入與更新。

而前述各模組可透過資料庫系統軟體包裝成單一程序組 (Transformation Packages)，並與資料庫管理系統之管理與排程功能整合，以達成資料倉儲之定期載入、轉換與更新之自動化。

在第三階段的計畫中，工作重點為研發適用於 web 的資訊勘測技術、建置線上分析處理系統與系統整合。基於先前所研發之線上分析處理系統技術，整合相關資料庫及處理程序，並提供使用者各種不同的觀點(view)來解釋原始的交易資料，以達成快速地透徹資料內容所代表資訊之目的。最後整合完整的 Web 資料倉儲系統並測試調整以確保系統之最佳化。此階段之研究成果主要分為以下幾部份：

(1) 建立分析模型

建立以整合型線上分析處理 (HOLAP) 為技術背景的分析模型，規畫在前端線上分析處理系統中合適的資料維度、維度間的層級從屬關係、充足的度量維度(measure dimension)。並討論在 HOLAP 的環境下，MOLAP 與 ROLAP 間轉換的時間及機制。

(2) 建立整合型線上分析處理系統

以現有 RDBMS，設計相關資料查詢與維護指令(DML & DDL)，並透過與

前端線上分析處理之軟體整合，建立 Web 應用分析的多維度資料立方體。並開發加強型查詢與資料整合指令，整合為整合型線上分析處理系統。

(3) 改進前端線上分析處理的互動式使用者介面

透過介面設計與執行批次檔撰寫，改進線上分析處理軟體之介面功能，協助使用者進行線上分析處理的基本操作，包括 pivot、roll-up、drill-down、slice-and-dice 等。讓使用者有最大的操作自由度，並能保持分析查詢的彈性。

(4) 完成資料勘測技術與線上分析處理整合

藉由前階段所開發之資料轉換模組，轉換 Web 資料倉儲存之各項資訊，提供資料勘測技術完整且資料品質高之分析資料源，充份地利用資料倉儲及線上分析處理對於將原始資料同質化及結構化後的優勢，進行使用者瀏覽行為勘測，並再回饋至資料轉換與載入模組，儲存於倉儲系統供後續分析使用。

(5) 資料驗證與系統維護

分段進行資料倉儲架構修正及勘測結果驗證。並藉由所建立系統維護機制，持續進行資料搜集、資訊整合、與資訊勘測三部份工作，調整各資料處理模組與分析模組，定期進行系統效能調整。

此階段針對資料倉儲、線上分析處理及資訊勘測三項技術進行系統整合之研究，建立包括線上分析處理技術融入資訊勘測程序、線上分析處理之資料立方體與資料倉儲系統整合、資料倉儲系統提供具良好品質之資料源供資訊與分析結果回存資料倉儲等程序，讓各系統模組間資訊可有效率地傳輸、運用。且最後建立之系統維護機制可持續進行資料搜集、資訊整合

與資訊勘測三部份工作，確保系統維持良好之運作效率。

參與本計畫的研究人員透過這個計畫各階段之研究，學習 Web 環境中的資料倉儲建構技術，設計具理論基礎及考量實際運作需要之系統架構與分析模組，充份應用所學相關資料技術於實用。並透過系統建置過程，增加開發系統軟體的系統分析、系統開發、測試驗證等實務經驗。由研究發展中所得到的經驗和成果可以提供線上資料分析儲存的相關資訊系統參考使用。本計畫對於如何整合應用新一代的資料儲存技術—資料倉儲與資料分析技術—資訊勘測，將存取記錄從資料轉化成可精確瞭解使用者行為的寶貴知識，可提供實務經驗供相關資訊系統參考使用。我們可以預期 Web 資料倉儲系統之重要性必將與日俱增，而藉由本計畫之研究成果，期可為國家實驗網路提供同時深具前瞻技術與實用價值之助益。

六、參考文獻

- [1] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules in Large Databases", *Proceedings of 1994 International Conference on Very Large Data Bases*, September 1994, Pages 478 - 499
- [2] R. Agrawal, T. Imielinski and A. Swami, "Mining Association Rules between Sets of Items in Large Databases", *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data*, 1993, Pages 207 - 216
- [3] J. Chang and S. Lee. Query reformulation using materialized views in data warehouse environment. In *Proc. Of the Int'l Workshop on Data Warehousing and OLAP*, pages 54—59, 1998.
- [4] S. Chaudhuri and U. Dayal, "An Overview of Data Warehousing and OLAP Technology", *SIGMOD Record* 26(1): 65-74, 1997
- [5] S. Chaudhuri, R. Krishnamurthy, S. Potamianos, and K. Shim. Optimizing queries with materialized views. In *Proc. Of Int. Conference On Data Engineering*, 1995.
- [6] M.-S. Chen, J.-S. Park and P. S. Yu, "Efficient Data Mining for Path Traversal Patterns," *IEEE Trans. on Knowledge and Data Engineering*, Vol. 10, No. 2, pp. 209-221, April 1998
- [7] H. Gupta. Selection of views to materialize in a data warehouse. In *ICDT*, pages 98—112, 1997.
- [8] H. Gupta and I. S. Mumick. Selection of views to materialize under a maintenance cost constraint. *Lecture Notes in Computer Science*, 1540:453—470, 1999.
- [9] A. Gupta, V. Harinarayan, and D. Quass. Aggregate-query processing in data warehousing environments. In *The VLDB Journal*, pages 358—369, 1995.
- [10] H. Gupta, V. Harinarayan, A. Rajaraman, and J. D. Ullman. Index selection for OLAP. In *Proceedings of ICDE*, pages, 208—219, 1997.
- [11] J. Han, S. Chee, and Y. J. (Chiang) Tam, "Issues for On-Line Analytical Mining of Data Warehouses", *Proceedings of the 1998 SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery (DMKD'98)*, June 1998, Pages 2:1 - 2:5
- [12] J. Han, J. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. *Proceedings of the 2000 ACM-SIGMOD International Conference on Management of Data*, May 2000.
- [13] V. Harinarayan, A. Rajaraman, J. D. Ullman, "Implementing Data Cubes Efficiently", *Proceedings of the 1996 SIGMOD Conference*, 1996
- [14] M. Kamber, J. Han, and Y. J. (Chiang) Tam, "Metarule-Guided Mining of Multi-Dimensional Association Rules Using Data Cubes", *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (KDD'97)*, August 1997, Pages 207 - 210
- [15] H.-Y. Kao, S.-H. Lin, J.-M. Ho and M.-S. Chen, "Mining Web Information

- Structures and Contents based on Entropy Analysis," *accepted by IEEE Trans. on Knowledge and Data Engineering*, 2003.
- [16] H.-Y. Kao, J.-M. Ho, and M.-S. Chen, "Information Clustering on DOM with Multi-Granularity Centroid Converging for Web Information Hierarchy Mining," *Proc. of the IEEE 2003 Intern'l Conf. on Web Intelligence (WI-2003)*, October 13-17, 2003.
- [17] W. J. Labio, D. Quass, and B. Adelberg. Physical database design for data warehouses. In *Proceedings of the International Conference of Data Engineering*, pages 277—288, 1997.
- [18] A. Y. Levy, A. O. Mendelzon, Y. Sagiv, and D. Srivastava. Answering queries using views. In *Proceedings of the 14th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, pages 95—104, San Jose, Calif., 1995.
- [19] C.-C. Lin and M.-S. Chen, "VIPAS: Virtual Link Powered Authority Search in the Web," *Proc. of the 29th Intern'l Conf. on Very Large Data Bases (VLDB-2003)*, September 9-12, 2003.
- [20] C.-H. Lee, C.-R. Lin and M.-S. Chen, "On Mining General Temporal Association Rules in a Publication Database," *Proceedings of the First IEEE International Conference on Data Mining (ICDM-01)*, November 29 - December 2, 2001.
- [21] C.-R. Lin and M.-S. Chen, "A Robust and Efficient Clustering Algorithm based on Cohesion Self-Merging," *Proceedings of the 8th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD-2002)*, July 23-26, 2002.
- [22] C.-R. Lin, C.-H. Lee, M.-S. Chen and P. S. Yu, "Distributed Data Mining in a Chain Store Database of Short Transactions," *Proceedings of the 8th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining (KDD-2002)*, July 23-26, 2002.
- [23] H. Mistry, P. Roy, S. Sudarshan, and K. Ramamritham. Materialized view selection and maintenance using multi-query optimization. In *Proceedings of SIGMOD Intl. Conf. On Management of Data*, 2001.
- [24] D. Theodoratos and T. K. Sellis. Data warehouse configuration. In *The VLDB Journal*, pages 126—135, 1997.
- [25] J. Yang, K. Karlapalem, and Q. Li. Algorithms for materialized view design in data warehousing environment. In *The VLDB Journal*, pages 136—145, 1997.