OntoWM：一個結合知識本體之資料倉儲探勘系統
OntoWM: An Ontology-Integrated Data Warehouse Mining System

林文揚[1]、吳錦昂[2]、曾明正[3]、江長隆[4]、黃科瑋[5]、洪子翔[6]、陳彥合[7]
[1,6,7]國立高雄大學資訊工程學系，[1]wylin@nuk.edu.tw，[6,7]{a0935541, a0935545}@mail.nuk.edu.tw
[2,3]義守大學資訊工程學系，[2]cwu@csu.edu.tw，[3]clark.tseng@msa.hinet.net
[4,5]國立高雄大學電機工程學系，[4,5]{m0955107, m0955108}@mail.nuk.edu.tw

摘要

本論文提出一個結合知識本體的資料倉儲探勘系統，稱為 OntoWM。此系統將知識本體的概念導入資料倉儲，藉由三個知識本體的建構：(1)描述資料倉儲的綱要組織與屬性關係的倉儲詮釋本體，(2)儲存倉儲資料的相關應用領域專家知識的領域知識本體，以及(3)彙整使用者探勘歷程的探勘歷程本體，定義出更符合使用者進行知識探索的系統架構。我們以多維度關聯規則的探勘為例，說明本系統的架構及設計方式，以及如何搭配運用所建構的知識本體，有效地產生符合使用者需求的規則樣式。

關鍵詞：資料倉儲、資料探勘、知識本體、多維度關聯規則

Abstract

In this paper, we propose an ontology-integrated data warehouse mining system, called OntoWM. Adopting the concept of ontology, this system integrates three different kinds of ontologies: including warehouse meta-ontology, which presents the structural and semantic relationship of warehouse schema, domain ontology, which stores expert knowledge related to the analysis domain of warehouse data, and query history ontology, which consolidates the history of users' mining queries. In this way, our system can provide a more effective and user-friendly environment for knowledge discovery. Taking the discovery of multidimensional association rules as example, we deliberate the design and construction of OntoWM system and how to make use of the ontologies to effectively generate qualified rules that meet user's demand.

Keywords: Data warehouse, data mining, ontology, multidimensional association rule

一、前言

隨著網際網路的興起與各式電子資訊資源的出現，如何由大量的資料庫中發掘出對使用者有意義的知識，已成為目前資訊研究領域中最重要的議題之一，此種概念稱為知識發掘(knowledge discovery)。根據 U.M. Fayyad 等的描繪[4]，知識發掘的過程是一種使用者主導的反覆式交談作業，主要的步驟包括：(1)瞭解應用的領域；(2)產生目標資料集；(3)資料淨化與轉；(4)選擇適當的資料探勘模式與探勘方法；(5)型樣評估與知識呈現。在整個知識發掘的過程中，資料探勘的前置作業，包括資料的淨化、挑選、彙整及格式轉換，最為複雜多樣以至於極耗費人力，因此，目前的知識發掘系統通常導入所謂的資料倉儲(data warehouse)架構[2]，以免除或減

輕使用者在資料前置處理所耗費的時間，而能將精力置於後段的資料探勘與樣式評估過程，此種概念又稱之為資料倉儲探勘(data warehouse mining)。換言之，在整個知識發掘的過程中，資料倉儲扮演了相當關鍵的角色。更具體地說，資料倉儲是整個知識探索系統中最主要的資料儲存所，其供給資料探勘工具以發掘出知識。此種架構基本上已可有效解決前述的資料前置處理的問題，然而在實際應用上目前的資料倉儲探勘仍然存在不少問題，如(1)資料間的組織與語意關係描述貧乏；(2)難以掌握使用者的探勘意圖；(3)缺乏對使用者，特別是非專家與新手的指引與協助；(4)缺乏有效、主動的知識更新能力等，使得使用者在進行知識探索時，往往需花費相當多的時間才能得到真正有用的知識。

為解決上述問題，我們提出一個整合知識本體概念的資料倉儲探勘系統，稱為OntoWM（Ontology-integrated data Warehouse Mining system）。此系統導入三個知識本體：(1) 描述資料倉儲的綱要組織與屬性關係的倉儲詮釋本體，(2)儲存倉儲資料的相關應用領域專家知識的領域知識本體，以及(3)彙整使用者探勘歷程的探勘歷程本體，藉此定義出更符合使用者進行知識探索的系統架構。我們並以多維度關聯規則的探勘為例，探討如何藉助使用者探勘喜好本體的機制，搭配運用本體論概念所建構的資料倉儲的詮釋資料以及領域專家知識，快速地產生符合使用者需求的規則樣式。我們的系統具有下列的優點：

1. 幫助使用者下達正確的探勘查詢，避免發生無法產生適當的規則或樣式的情況。
2. 能在既有的原始資料中，找出可延伸其概念的規則或樣式。
3. 更有效地剪除不可能存在的項目組合，加速找出有意義的關聯規則。

本論文的章節安排如下：首先在第二節中說明我們提出此一系統的構想與整個系統的架構；接著在第三節中細述此系統各主要單元，包括多維度關聯探勘查詢的語法、各知識本體的結構、結合知識本體的探勘方法、以及探勘使用者介面等的設計理念與作法；在第四節中說明過去相關的研究文獻；最後是本論文的總結與未來將繼續的研究工作。

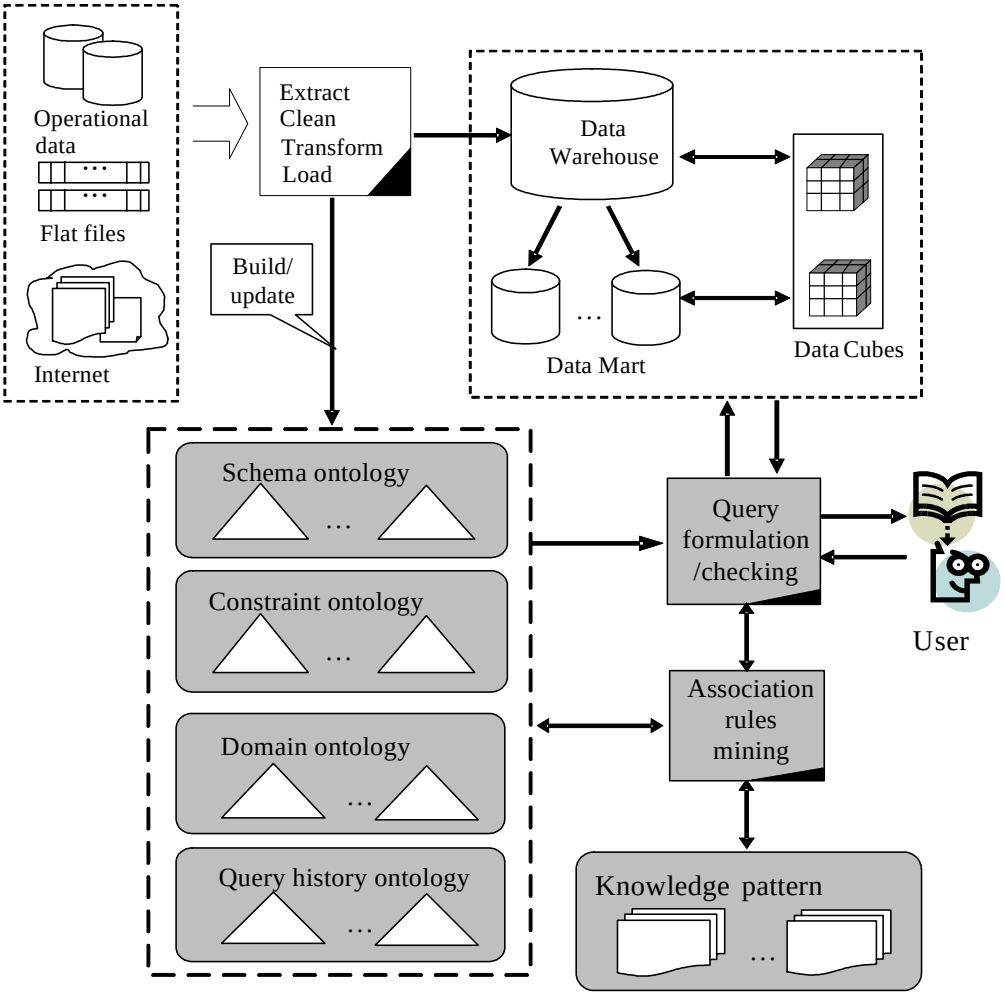
二、研究構想與系統架構

目前的資料倉儲系統大都以關聯式資料模式為主，僅能描繪事實資料表與各維度資料表間的關係[13]，各屬性之間、屬性值相互的關係則付之闕如。若在資料倉儲中能提供資料的組織及語義的相互關係，則能讓使用者在進行資料探勘的過程中，避免許多因不當的資料選取，而產生無意義或難以理解的規則或型樣；而減少此種的錯誤，也就能減少使用者反覆進行資料探勘的次數，及縮短整個探勘的時間。

此外，資料探勘是一項極為主觀的資料處理工作，採掘出的結果及展示的方式往往取決於使用者的喜好。因此，若我們能進一步將使用者的喜好，如探勘的資料範圍、參數的設定、有意義的型樣模式等，也納入資料倉儲探勘系統之中，則能更快過濾對使用者而言不具價值的規則或型樣。

根據上述的概念，我們的構想是定義一個結合知識本體的資料倉儲探勘系統，希望運用知識本體的概念，有效建構出目前關聯式綱要架構無法表達出、但對資料探勘而言卻是極重要的資料關係，無論是語意或結構上的，甚至是其他有用的知識，並整合相關的領域知識本體；另外也能運用知識本體的架構，將使用者的探勘查詢歷程納入系統之中。

為此我們提出如圖一所示的系統架構，其植基於一般常見的資料倉儲系統，並整合三個知識本體：倉儲詮釋本體、領域知識本體、以及探勘歷程本體。使用者可經由一個具智慧引導的探勘交談介面，由系統運用所建構的知識本體，輔助使用者下達正確適當的關聯查詢，經規則探勘模組處理產生符合的關聯規則。

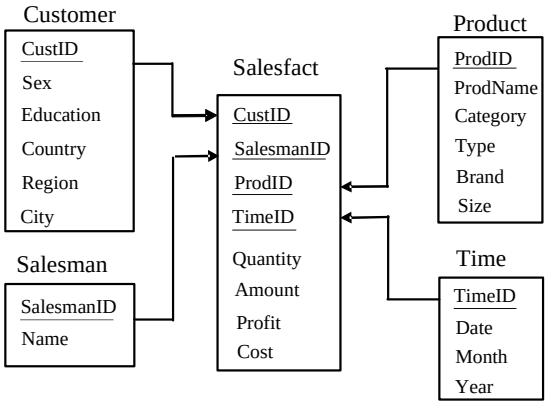


圖一：結合知識本體之資料倉儲探勘系統 OntoWM 的架構

在下節中，我們將細述此系統各主要單元，包括多維度關聯探勘查詢的語法、各知識本體的結構、結合知識本體的探勘方法、以及探勘使用者介面等的設計理念、方法與實作方式。

三、系統設計與實作

為方便說明，我們假設資料倉儲中所儲存的資料為有關電腦及其周邊產品的銷售資料，其綱要結構為如圖二所示的星狀綱要。



圖二：銷售星狀綱要範例

(一)多維度關聯規則探勘語言

首先我們先說明多維度關聯規則的定義，接著描述我們所定義的探勘語法。

所謂的多維度關聯規則可表示成如下

的規則：

$$A_1 = v_1 , A_2 = v_2 , \ldots , A_m = v_m \Rightarrow B_1 = u_1 , B_2 = u_2 , \ldots , B_m = u_n$$

其中 A_i，$1 \le i \le m$，及 B_j，$1 \le j \le n$，皆表示資料的屬性，而 v_i 和 u_j 則分別為屬性 A_i 及 B_j 的值。

根據 J. Han 與其同事及學生的研究[9][12][23]，多維度關聯又可分為維度內關聯(Intra-dimensional association)，維度間關聯(Inter-dimensional association)，及混合式維度關聯(Hybrid dimensional association)。維度內關聯是針對某一維度的屬性，考量其儲存資料間的關聯關係，而維度間關聯指的是不同維度屬性資料間的關聯關係，而混合式維度關聯則是前述兩種的組合。例如考慮圖二的資料倉儲綱要，以下三例分別表示上述的三種多維度關聯規則：

Product.ProdName = ‘Desktop’ ⇒ Product.ProdName = ‘Ink-jet’

Customer.City = ‘Taipei’, Product.ProdName = ‘Desktop’ ⇒ Time.Month = ‘December’

Customer.Country = ‘Taiwan’, Product.ProdName = ‘Desktop’ ⇒ Product.ProdName = ‘Ink-jet’

在資料倉儲中探勘此種多維度關聯規則，首先分析者須決定資料的尺度(granularity)，意即構成每筆交易資料的辨識欄位，以及構成規則的項目欄位為何。

例 1：以圖二的資料倉儲為例，表一即對應辨識欄位為客戶編號、日期，而項目欄位資料表為教育程度、居住城市、產品類別的探勘目標資料表。若是將辨識欄位改成客戶編號，則資料筆數將由 6 筆變成 4 筆，且每筆資料的內容也隨之更動。由此可清楚看出識別欄位直接影響資料的尺度與組成，並進而影響產生的規則。

根據前述的概念，我們設計的多維度關聯規則探勘查詢語法定義如下：

表一：以客戶編號、日期為識別欄位，以教育程度、居住城市、產品類別為項目欄位所產生的探勘目標資料表。

Grouping_ID		**Interested Mining_Items**		
Cust_ID	**Date**	**Education**	**City**	***Category**
C01	2007-02-01	Bachelor	Taipei	B,C,D,E
C02	2007-02-03	Bachelor	Tainan	A,D,E
C03	2007-02-10	31-40	Taichung	C,D
C01	2007-02-10	21-30	Taipei	A,E
C05	2007-02-12	PhD	Kaohsiung	A,B,C
C03	2007-02-15	Master	Taichung	A,E
*A: Printer B: Laptop C: Desktop D: Memory E: Hard Disk				

Mining multidimensional associations
[**Match metarule** <metarule_format>]
[**Grouping_ID** <attribute list>]
Mining_item <attribute list>
From <DW_Name [,Domain_Ontology]>
[**Where** <where conditions>]
[**Having** <having conditions>]
Threshold <ms, mc>

其中位於括號’[]’中的部分表示可省略的選項，位於括號’<>’中表示查詢子句的內容需額外再定義。“Match metarule”子句是供分析者定義規則的樣式，其定義方式是仿造[12]的作法。“Grouping_ID”子句是定義構成交易資料的辨識欄位，而“Mining_item”子句則是定義構成規則的項目欄位。當 Grouping_ID 子句省略時，即表示此探勘的辨識欄位沿用來源資料倉儲的事實資料表的關鍵欄位。“Where”子句是用以設定探勘資料的限制條件，即符合此條件的資料才會被選取。“Having”子句則是用以設定過濾經過群組後的資料的條件式。Threshold 子句中的 ms 及 mc 分別代表最小支持度及最小信賴度門檻。

例 2：下列查詢表示，在日本市場的每個客戶每天的交易資料中，客戶購買 HP 印表機的行為是否與其教育程度、居住城市有關聯，而規則的支持度至少為 30%、信賴度至少須達 50%。

Mining multidimensional associations
Metarule Education = ?, City = ? ⇒ Category = ‘Printer’, Brand = ‘HP’
Grouping_ID CustID, Date
Mining_Item Education, City, Category, Brand

From PC_Sales_Star
Where Country='Japan'
Threshold ms=30%, mc=50%

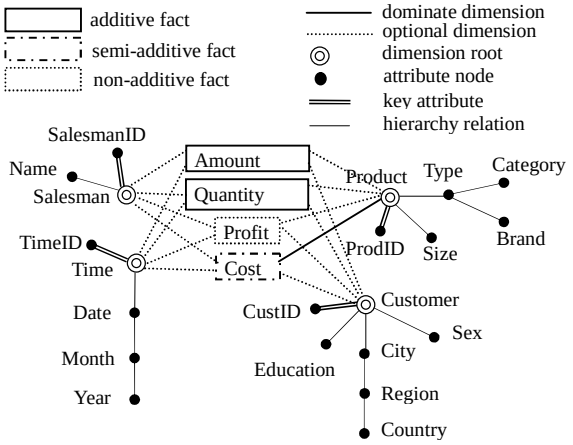
(二)倉儲詮釋本體

多維度關聯雖然可深入表達資料間的各種可能的關聯關係，但其複雜度會隨著維度與屬性的增加而成指數成長，故分析者更需要瞭解資料間的組織結構及語意上的關係，才能針對相關的資料欄位進行探勘，以減少不必要的錯誤嘗試。例如若使用者瞭解產品的型號與類別間具有階層的關係，即由產品的型號即可決定其所屬的類別，則就不會嘗試去發掘使用者所購買的產品資料中，不同的型號與類別間是否具有關聯關係，而期望得到如下的關聯規則

Product.Type = HP 3325 ⇒ Product.Category = Ink-jet

除了上述的階層關係會影響關聯探勘外，其實只要屬性之間存在所謂的功能相依關係—即對兩個不同的屬性 A_1、A_2 及任意兩筆資料 t_1、t_2，若 $t_1.A_1 = t_2.A_1$，則 $t_1.A_2 = t_2.A_2$—就會產生類似的結果。例如，客戶所購買的產品廠牌往往與其居住的國家有絕對的關係。因此，若資料倉儲中有描繪出屬性間的功能相依關係，則也能避免因使用者不當的資料選取，而產生無意義或了無新意的關聯規則。因此，我們導入倉儲詮釋本體，此知識本體又可分為兩部分：倉儲綱要詮釋本體，以及屬性關聯限制本體。

倉儲綱要本體是用以描述資料倉儲的詮釋資料，包括綱要架構、維度階層關係、屬性與屬性值的相依關係、以及量測屬性與維度屬性間的關係等。圖三顯示的是以圖二為例的詮釋本體的概念示意圖，其中表達了維度結構，包括產品維度(Product)、客戶維度(Customer)、時間維度(Time)、以及銷售人員維度(Salesman)；各維度的階層關係；量測屬性的累加特性，分為(1)可累加性(additive)，意指此量測值可在任意的維度組合下進行加總運算，如銷售量(Quantity)與銷售金額(Amount)，(2)半累加性(semi-additive)，指此量測值必須搭配某些維度其結果才有意義，如成本(Cost)必須搭配含有產品的維度組合才有意義，以及(3)不可累加性(non-additive)，指此量測值不具累加性，必須經額外的計算才能獲得，如利潤(Profit)是由銷售金額減去成本而得之。



圖三：對應圖二之倉儲綱要詮釋本體。

屬性關聯限制本體則是用以描繪屬性間的各種限制關係，更明確地說，是指此知識的運用可有效避免使用者在探勘的過程中，因不當的資料選取，而產生無意義或難以理解的規則或型樣。針對多維度關聯規則的探勘，我們目前定義了七種屬性限制關係：

1. 互斥(Exclusive)：此關係表示兩屬性應避免同時被選取作為探勘的屬性，例如產品編號(ProdID)與產品名稱(ProdName)皆可辨識某一產品，故在探勘時，應避免分析者同時選取此兩個屬性，可減少不必要的計算。
2. 唯項(ItemOnly)：此表示某一屬性只適合作為探勘項目的屬性，例如產品大小(Size)就其意義而言不適合作為構成決定資料尺度(granularity)的群組屬性。
3. 決定(Decide)：此關係即傳統的功能相依關係，例如產品編號可決定產品的種類、大小、型式、廠牌等。

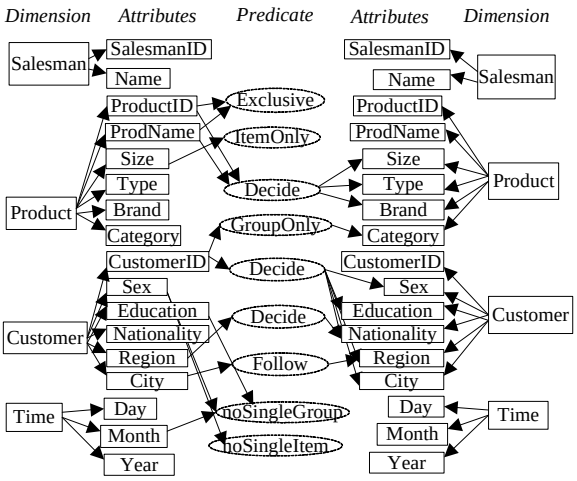
4. 伴隨(Follow)：此表示某些屬性必須伴隨其他屬性方能正確識別資料的差異。例如，城市(City)不能單獨被選取，至少必須伴隨區域(Region)，此乃因不同區域內可能有相同的城市名。

5. 唯群(GroupOnly)：此表示某些屬性只適合作為探勘查詢的群組屬性，例如客戶編號(CustID)。

6. 非獨群(NoSingleGroup)：此表示某些屬性不適合單獨作為探勘查詢的群組屬性，例如性別(Sex)。此乃因這些屬性的值域(Domain)太小，如性別只有兩種，若是單獨作為群組屬性，則資料筆數過少(以性別而言，只有兩筆)，難以產生有意義的規則。

7. 非獨項(NoSingleItem)：此表示某些屬性不適合單獨作為探勘查詢的項目屬性，如性別。其原因與前述的非獨群類似，若是單獨作為項目屬性，則構成每筆資料的項目太少，以性別而言只有兩種，同樣無法產生有意義的規則。

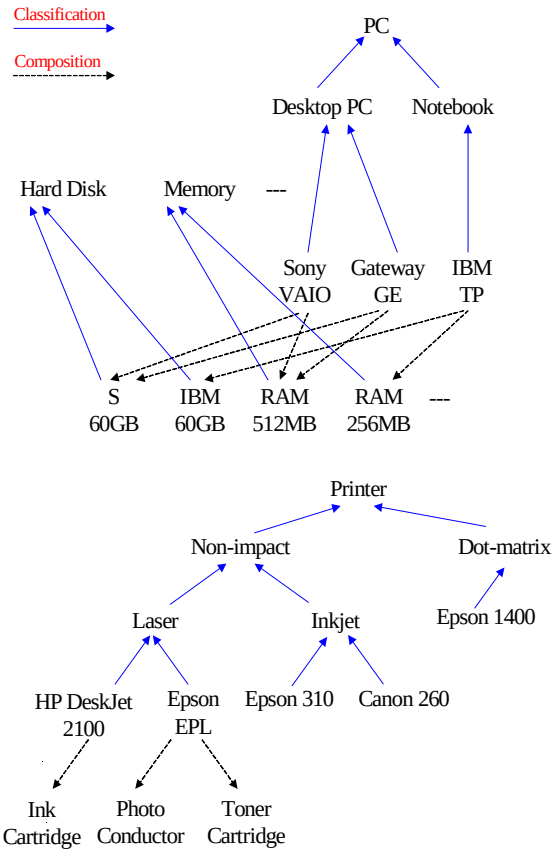
圖四所描繪的即為對應圖二的倉儲資料的屬性關聯限制本體。



圖四：對應圖二之屬性關聯限制本體。

(三)領域知識本體

所謂的領域知識本體是用以儲存與資料倉儲所儲存的資料主題相關的專家知識。例如若資料倉儲的分析主題為銷售資料，則此領域的知識本體將用以描繪如商品的概念階層架構、商品零組件的關係等。以電腦相關產品為例，圖五所描繪的為有關產品的分類關係與零件間的組成關係的領域知識。



圖五：以電腦與周邊相關產品為例之領域知識本體，包括產品類別與組成元件關係。

導入領域知識本體的主要原因之一是能在既有的原始資料中，找出可延伸其概念的規則或樣式。如前所述，目前的資料倉儲系統在描繪其所儲存的資料組織關係時，通常無法完全反映資料記錄彼此之間可能具有的衍伸概念關係。例如以圖三的為例，產品維度的屬性階層僅將產品分類或區分不同廠牌，未必能反映其他產品間可能存在的概念階層。由圖五中我們可看出 Printer 及 PC 雖為產品的兩個類別，但是型號至類別之間仍然有其他的類別關係並未反映在綱要結構上，而且不同的產品的概念階層關係亦不相同。因此，若能結合此領域知識本體，則能找出其他有價值

的規則，如下例

Customer.City = ‘Taipei’ ⇒ Product.Category = ‘Notebook’ with ‘IBM 60GB HD’

此規則表示居住台北的客戶傾向買所安裝的硬碟為 IBM 60GB 的筆記型電腦。

(四)探勘喜好本體

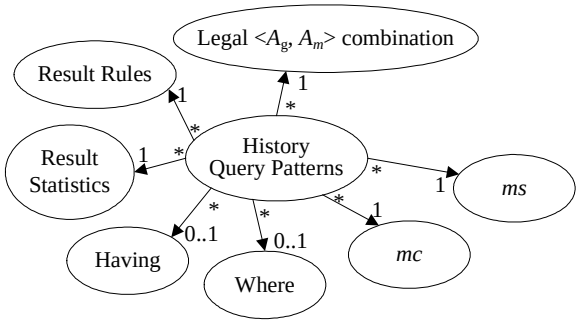
此知識本體旨在收集使用者在資料探勘中所下達的查詢樣式，如資料範圍、參數設定、過濾條件等，建構出使用者的探勘喜好本體。建構此知識本體的動機之一是希望藉由累積使用者過去的探勘歷程，由中產生有用的資訊，可以對後續使用者在設定查詢的重要參數，如支持度與可信度門檻值時，自動給予適當的初始建議值。如此可減輕許多使用者不知如何設定這些參數的困擾，同時也減少反覆探勘的次數與時間。

此探勘歷程知識本體的結構如圖六所示。我們以類似星狀綱要結構的方式來匯整使用者曾經下達、且滿意其結果的關聯查詢。除了紀錄其語法，包括識別屬性與項目屬性(<A_g, A_m>)、Where 及 Having 條件、以及支持度與可信度門檻之外，並紀錄其探勘的結果，即頻繁項目集與關聯規則，以及探勘結果的統計資料，包括：

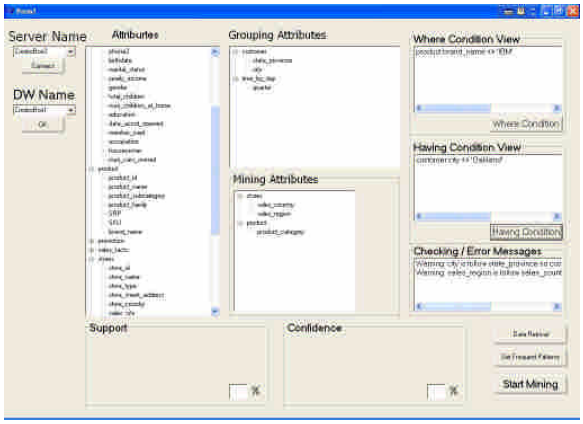
- 頻繁項目集的數目
- 項目集的平均長度與平均支持度
- 最大的項目集的長度與支持度
- 最小的項目集的長度與支持度
- 規則的數目
- 規則最大的可信度、最小可信度與平均可信度

(五)探勘使用者介面

由於多維度探勘的查詢比單純的為減輕使用者操作上的困難，我們採用類似圖形化舉例式查詢(Graphical Query By Example，簡稱 GQBE)的設計理念[18]，以 GUI 的方式呈現各查詢語法的子句設定，引導使用者雖然不熟悉查詢的正確語法，也能順利表達所欲進行的探勘查詢。



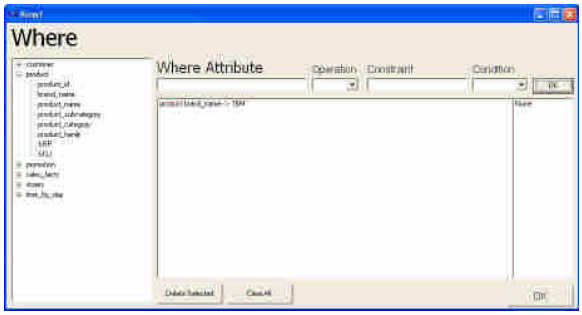
圖六：探勘歷程知識本體的結構。



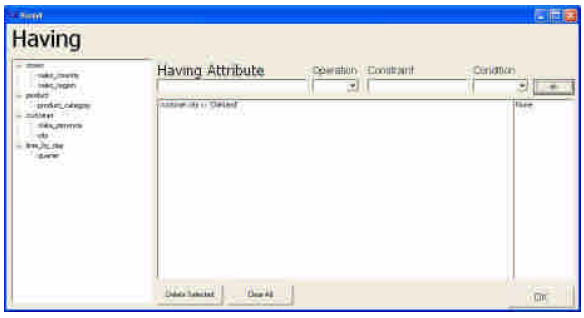
圖七：OntoWM 系統使用者操作介面。

如圖七所顯示，使用者先選取倉儲伺服器與資料倉儲名稱後，系統即顯示所有的資料表及其屬性。接著使用者即可以拖曳的方式，選取適當的屬性當作群組識別欄位(Grouping Attributes)與探勘項目欄位(Mining Attributes)，這過程中若是選取的屬性組合不當或不合宜，則系統會即時產生警告或錯誤訊息，提醒使用者更改其設定；若是前述的設定無問題，則系統會由探勘喜好本體中，自動計算產生適當的支持度與可信度門檻；若是需要，可點選 Where 與 Having 條件設定按鈕，進入如圖八與圖九的畫面進行設定，為 OntoWM 系統的操作介面。同樣地，此部分的設定方式也是以拖曳的方式，逐個選取屬性，並以 QBE 的方式完成邏輯判斷式的設定。最後，按下 Start Mining 按鈕，產生對應的關聯查詢與符合的規則。此外，對於初次

使用本系統的使用者，我們也設計了一個逐步引導使用者完成查詢設定與探勘工作的精靈。



圖八：Where 條件設定子畫面。



圖九：Having 條件設定子畫面。

(六)探勘處理流程與方法

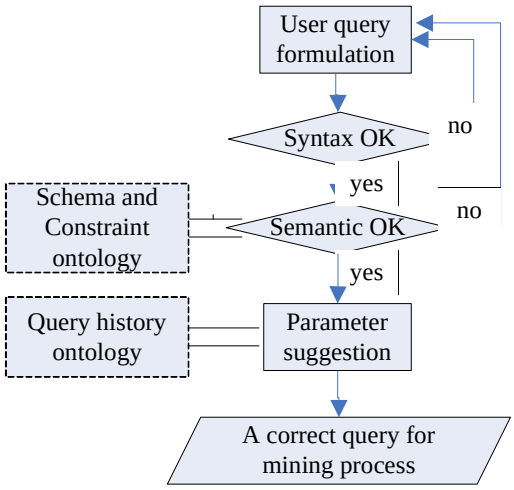
在這一小節中，我們將說明 OntoWM 的核心：使用者探勘查詢的處理流程與方法。

整個查詢的過程大致可分為四個主要步驟：

1. 即時驗證查詢語法與語意的正確性；
2. 由資料倉儲擷取資料並進行格式轉換；
3. 自目標資料集中探勘符合的頻繁項目集；
4. 由頻繁項目集組成關聯規則。

接下來說明各主要步驟的處理方式及其與各知識本體間的存取關係。

圖十所示為步驟 1 的資料流程。在使用者設定查詢的過程中，此系統模組會即時且持續地檢查其語法的正確性，接著運用資料倉儲詮釋本體，檢查語意上是否有違反或不適宜的設定，最後再根據探勘歷程本體的內容，尋找與比對接近的查詢，自動計算產生支持度與可信度門檻。



圖十：探勘查詢驗證的資料流程。

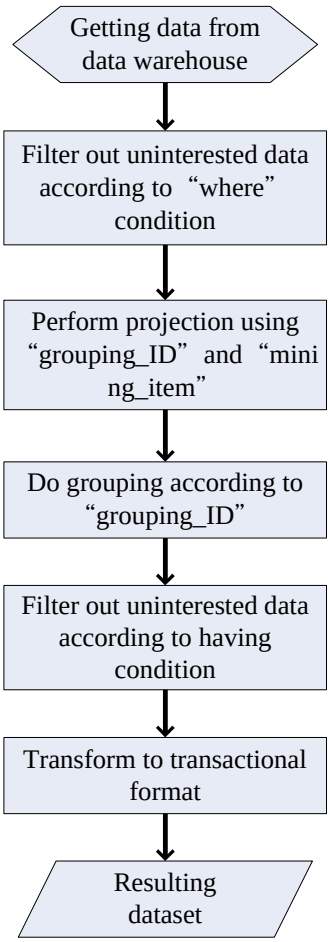
圖十一所示為步驟 2 的資料流程。這個步驟主要在擷取資料及格式轉換，故毋須存取任何知識本體。

步驟 3 在探勘頻繁項目集的過程中，需要存取的知識本體為領域知識本體及探勘歷程本體；前者是用來產生延伸的概念項目集，而後者則是可利用其儲存的頻繁項目集避免不必要的計算以加速執行的時間。有關此步驟的探勘方法是採用我們發展的 AROC 演算法，限於篇幅無法在此詳細描述，其細節可參考[22]。

最後在步驟 4 產生關聯規則的階段，則是需要存取領域知識本體，加入項目間的類別與組合的語意，以強化規則的意義。

(七)系統實作

本系統是以 SQL Server2005 作為資料倉儲的資料庫平台，除了儲存資料倉儲的資料外，也以關聯式資料表的方式儲存倉儲詮釋本體與探勘記錄本體。我們之所以採用關聯式資料來儲存這兩個知識本體的主要考慮是資料存取的效率。然而領域知識本體的建構則是採 OWL 格式[24]，使用史丹佛大學所發展的 Protégé[25]來進行編輯，再轉存為 OWL 格式檔案。之所以



圖十一所示為步驟 2 的資料流程。

採行 OWL 的資料格式而非關聯式資料的原因在於考慮此領域知識本體具有共通性，將來可與其他現有的領域知識結合。至於程式介面則是以 Borland C Builder 來開發，目前已完成系統的雛型，也初步驗證其正確性與效能。

四、相關文獻探討

根據我們所蒐集的文獻顯示，目前國內外尚無直接與本計畫所研究的議題相似或雷同的研究計畫或成果。因此，我們將就部分相關的研究或技術作說明，大致分為二方面：(1)知識本體論在資料探勘的應用；(2)資料倉儲探勘。

(一)知識本體在資料探勘的應用

目前知識本體應用在知識發掘領域如資料倉儲、資料探勘的研究仍屬少數。不過若是將概念階層(或分類學，taxonomy)視為是知識本體的一種，則有關知識本體在資料探勘方面的研究可追溯至 1995 年，Han & Fu[7]以及 Srikant & Agrawal[19]分別提出如何結合概念階層知識，探勘所謂的階層式關聯規則以及一般化關聯規則。不過這些研究的重點都在演算法的設計，而非知識本體的組織架構以及對資料採掘的助益。之後 Taylor 等人[21]利用知識呈現系統 ParkaDB 整合了知識本體和資料庫，並利用資料探勘技術歸納出高階概念之分類規則呈現給使用者。Bernstein 等人[1]則探討如何利用知識本體來評估挑選適當的分類方法。

(二)資料倉儲探勘

雖然資料倉儲系統與資料探勘技術的結合已是目前知識採掘的主要平台架構，但有關此二者如何更緊密結合的研究仍然是知識採掘領域中主要的議題之一。早期研究資料探勘與資料倉儲相結合的學者，大多專注在如何利用原本作為 OLAP 使用的資料方體或多維度資料庫進行資料採掘。以加拿大華裔學者 J. Han 為首的研究團隊[5][12][23]是最早投入此主題的研究單位；他們提出了一個結合 OLAP 與資料探勘的系統(稱為 DBMiner)[6]，發展出利用資料方體中的資料進行各種資料採掘的分析處理，包括關聯式規則、分類、預測與分群。

除了 J. Han 等有系統的投注在這方面的研究以外，其他則是一些零星的相關研究。在[10]中，作者亦提到可利用 OLAP 資料方體進行資料採掘的概念，不過其重點卻是在如何利用平行處理技術建立 OLAP 的資料方體。在[3]中，作者提出一個應用於電子商務的分散式 OLAP 架構，文中特別提到如何建置供資料採掘使用的資料方體及其架構只是該論文著重在電子商務的應用層面，未能深入建立資料採掘的資料方體的完整架構。

在[17]中，Psaila 及 Lanzi 則考慮如何

從資料倉儲中的資料進行階層式的關聯規則採掘，提出一套階層式規則的型式與採掘的概念性方法，與 J. Han 的研究最大不同之處是資料直接取自在資料倉儲中以傳統的星狀模式儲存的資料，而非已經過處理的 OLAP 資料方體。在[15]中，Perng 等人也提出一個自關聯式資料庫中採掘關聯規則的探勘空間架構，其概念類似前述 Han 等的多維度探勘的觀念，不過其重點在於探勘搜尋空間的修剪。另外，也有學者[14]研究如何由星狀架構的資料中，不需事先作資料表合併的動作，直接發掘出關聯規則的探勘方法。

五、結論與未來工作

在本論文中，我們提出結合知識本體論的資料倉儲探勘系統的概念，並實際建構一結合知識本體的資料倉儲探勘系統 OntoWM。我們在三方面導入知識本體的概念：(1)資料倉儲的詮釋資料；(2)所儲存的倉儲資料的相關領域知識；以及(3)彙整使用者的探勘紀錄。

目前我們已完成 OntoWM 系統的雛型，未來我們將以此雛型系統為基礎，朝下列幾方面繼續研究並擴充其功能：

1. 變動環境下的探勘機制：目前的 OntoWM 系統是植基於資料來源及各知識本體為固定不變的假設下所開發完成，然實際的情況是這些資料及知識皆隨外界環境變化而改變，因此我們將研究導入在變動環境下的探勘機制，包括資料倉儲內容的更新、各知識本體結構或內容的異動等，如何進行探勘處理。
2. 知識本體的建構與維護：目前的 OntoWM 的知識本體的建構方式都是由系統開發人員利用其他工具建構、存入資料庫中。未來我們希望能將知識本體建構的機制整合到 OntoWM 中，以提供系統管理者較方便的操作機制。
3. 其他探勘技術：目前的 OntoWM 系統只提供關聯規則的探勘，未來我們主要的工作之一將陸續研究開發其他常見的探勘技術[8]，包括分類(Classification)、分群(Clustering)、以及序列型樣(Sequential pattern)。

致謝

本論文為國科會補助之研究計畫成果 NSC 94-2213-E-390-006 以及 NSC 95-2213-E-214-024。

參考文獻

[1] A. Bernstein, F. Provost, and S. Hill, "Intelligent assistance for the data mining process: an ontology-based approach," CeDER Working Paper IS-02-02, Center for Digital Economy Research, Stern School of Business, New York University, 2002.

[2] S. Chaudhuri and U. Dayal, "An overview of data warehouse and OLAP technology," *ACM SIGMOD Record*, Vol. 26, pp. 65-74, 1997.

[3] Q. Chen, U. Dayal, and M. Hsu, "A distributed OLAP infrastructure for e-commerce," *Proc. of IEEE 1999 IFCIS Int. Conf. on Cooperative Information Systems*, pp. 209-220, 1999.

[4] U. Fayyad, P.S. Gregory, and S. Padhraic, "The KDD process for extracting useful knowledge from volumes of data," *Communications of the ACM*, Vol. 39, No. 11, pp. 27–34, 1996.

[5] J. Han, "OLAP mining: An integration of OLAP with data mining," *Proc. of IFIP Conf. on Data Semantics*, pp. 1-11, 1997.

[6] J. Han et al., "DBMiner: A system for mining knowledge in large relational databases," *Proc of Int. Conf. on Data Mining and Knowledge Discovery*, pp. 250-255, 1996.

[7] J. Han and Y. Fu, "Discovery of multi-

ple-level association rules from large databases," *Proc. of 21st Very Large Databases Conference*, pp. 420-431, 1995.

[8] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2001.

[9] J. Han, L.V.S. Lakshmanan, and R.T. Ng, "Constraint-based multidimensional data mining," *IEEE Computer*, Vol. 32, No. 8, pp. 46-50, 1999.

[10] S. Goil and A. Choudhary, "High performance data mining using data cubes on parallel computers," *Proc. of 1st Int. Symposium on Parallel and Distributed Processing*, pp. 548-555, 1998.

[11] T.R. Gruber, "A translation approach to portable ontology specifications," *Knowledge Acquisition*, Vol. 5, pp. 199-220, 1993.

[12] M. Kamber, J. Han, and J.Y. Chiang, "Metarule-guided mining of multi-dimensional association rules using data cubes," *Proc of 3rd Int. Conf. Knowledge Discovery and Data Mining*, pp. 207-210, 1997.

[13] R. Kimball, *The Data Warehouse Toolkit*, John Wiley & Sons, INC. 1996.

[14] E.K.K. Ng, A.W.C. Fu, and K. Wang, "Mining association rules from stars," *Proc. of 2002 Int. Conf. on Data Mining*, pp. 322-331, 2002.

[15] C. Perng, H. Wang, S. Ma, and J.L. Hellerstein, "FARM: A framework for exploring mining spaces with multiple attributes," *Proc. of IEEE Int. Conf. Data Mining*, pp. 449-456, 2001.

[16] T. Priebe and G. Pernul, "Ontology-based integration of OLAP and information retrieval," *Proc. of 14th Int. Workshop on Database and Expert Systems Applications*, pp. 610-614, 2003.

[17] G. Psaila and P.L. Lanzi, "Hierarchy-based mining of association rules in data warehouses," *Proc. of ACM Symposium on Applied Computing*, pp. 307-312, 2000.

[18] A. Silberschatz, H.F. Korth, and S. Sudarshan, *Database System Concepts*, 5th Ed, McGraw-Hill, 2006.

[19] R. Srikant and R. Agrawal, "Mining generalized association rules," *Proc. of 21st Very Large Databases Conference*, pp. 407-419, 1995.

[20] K. Stoffel, J. Saltz, J. Hendler, J. Dick, W. Merz, and R. Miller, "Semantic indexing for complex patient grouping," *Proc. of Annual Conf. of the American Medical Informatics Association*, 1997.

[21] M. Taylor, K. Stoffel, J. Hendler, "Ontology-based induction of high level classification rules," *Proc. of SIGMOD Data Mining and Knowledge Discovery Workshop*, 1997.

[22] M.C. Tseng, W.Y. Lin, and R. Jeng, "Mining association rules with ontological information," *Proc. of 2nd Int. Conf. on Innovative Computing, Information and Control*, 2007.

[23] H. Zhu, *On-line Analytical Mining of Association Rules*, Master Thesis, SIMON FRASER University, 1998.

[24] *OWL Web Ontology Language Use Cases and Requirements*, http://www.w3.org/TR/webont-req/

[25] The Protégé Ontology Editor and Knowledge Acquisition System, http://protege.stanford.edu/