

基於專案屬性資訊之知識探勘與再利用平台發展

A Platform for Supporting Knowledge Mining and Reuse Based on Context Information of a Project

吳怡瑾* 林書偉**
I-Chin Wu*, Shu-Wei Lin**

摘要

企業在面對產業環境快速變革之際，如何透過知識管理以創造組織價值活動，並善用智慧資產提升組織核心能力，已成為企業提昇商業智能的重要議題。專案計畫的執行為公司內重要的知識管理活動，專案包含了多項知識密集性高的工作，本研究透過歷史專案與工作集中萃取所需的資訊，設計專案情境屬性資訊模式（project-in-context, PIC），提出基於專案屬性資訊之知識發掘與再利用平台。因此本研究主要內容包括：(1)專案資訊擷取：分析專案執行過程各階段所需的相關知識物件，透過知識工程的程序，由歷史工作集中萃取、分析工作相關文件資訊；(2)專案知識發掘：結合文件探勘與資料探勘技術，萃取重要專案概念並根據PIC模型定義之知識物件以分析物件間的關聯；(3)專案屬性資訊利用：研究由主動搜尋與被動遞送的角度設計與探討由專案屬性建構情境資訊利用之實際運作模式。由專案計畫管理的角度出發，本研究提出之方法與系統可協助知識工作者快速的瞭解工作脈絡並有效解決問題。

關鍵字：資料探勘、知識發掘、知識再利用、專案情境資訊模式、文件探勘

Abstract

Organizations implement Knowledge Management Systems (KMS) to maximize the effectiveness and reuse of knowledge assets in order to increase productivity and profitability. Thus, effective project management can place great demands on knowledge management solutions designed to support and streamline the execution of project-related tasks. Accordingly, in this work we extract knowledge from historical projects, design a project-in-context (PIC) meta-model, and deploy a platform that facilitates the capture and reuse of project-specific information based on the context. The research areas addressed in the work are as follows. (1) Knowledge acquisition: analyzing the type of project and its associated attributes and defining general, but essential, project context information based on the PIC model. (2) Knowledge discovery: the use of text mining and data mining techniques to extract knowledge items needed by workers, and discover the relationships between various knowledge items. (3) Knowledge utilization based on the context: with the proposed model and methods, several applications related to the reuse of project knowledge by pull- and push-based knowledge management strategies are developed to achieve effective project management. From the perspective of project management, the proposed model and system can help knowledge workers understand information about a current research project and resolve problems effectively.

Keywords: Data Mining; Knowledge Discovery; Knowledge Reuse; Project Context Information; Text Mining

* 天主教輔仁大學資訊管理系助理教授（通訊作者）

Assistant Professor, Department of Information Management, Fu-Jen Catholic University. (To whom all correspondence should be addressed.)

E-mail: icwu.fju@gmail.com

** 天主教輔仁大學資訊管理系研究生

Graduate Student, Department of Information Management, Fu-Jen Catholic University

壹、前言

根據KPMG的研究報告，目前有80%的企業認為組織核心知識為企業之重要策略資產並將知識管理視為提昇企業競爭力的重要因素（Davenport & Prusak, 1998; KPMG Consulting, 2003; Nonaka & Takeuchi, 1995）。專案的執行為組織內重要的知識管理活動之一，而專案包含了多項知識密集性高的工作，如：專案工作所面對的問題、原因與解決方法…等，均包含大量的知識，因此，如何透過知識管理流程與策略以有效的支援專案計畫的執行為重要的研究議題。近幾年企業組織以專案方式組成團隊以建立開發新產品、技術或服務、縮短研發時間和減少研發成本，並保留研發過程的相關資源；在資訊科技研發方面，企業也經常透過專案以降低軟體製造風險和減少成本的支出。而專案開發時，若不充分瞭解企業本身或者外部的資源與企業工作流程，往往無法有效利用資源進而完成工作；此外，專案完成後若沒有一套完善的制度將執行過程的經驗或是專案的成果保留，如：專案的計畫書、問題解決、經驗學習…等等，將無法有效利用企業專案資源。

由於企業電子化，導致組織資料來源多樣且資訊量亦隨之劇增，外顯知識的儲存、管理與散佈亦為目前組織知識管理的重要工程。根據IDC的報告，內容管理市場於2005年已超過14億美元，並且為IT服務產業中成長率最為突出的產業。其中，文件管理系統被企業視為知識外顯不可或缺的要件，更為

實踐知識管理的第一步。著名的知識管理工具，如：Documentum 4i，OpenText Live Link，Autonomy KnowledgeServer，文件與內容管理皆為上述開發工具的核心功能。文件管理之目的為累積組織工作相關經驗與知識，並透過有效的蒐集、組織與散佈方式，充分發揮組織知識資產的價值。目前企業知識入口網站紛紛提出支援組織資訊搜尋服務的具體方法，如：全文檢索系統、知識目錄分類結構、關鍵詞查詢、問與答（Q&A）系統…等技術。相關技術與應用服務對組織部分資訊的搜尋有所助益但對協助知識工作者進行脈絡性知識的獲取與探索卻有其限制性；另一方面，雖有相關研究提出知識地圖的概念以協助組織知識的呈現與導覽，但是仍缺乏對特定專案或工作情境資訊的描繪，並缺乏探討如何系統化的保留與組織特定專案或工作情境資訊。

專案軟體開發程序中，每一個階段皆會產生知識，因此若能考量專案里程碑或階段性以進行資訊蒐集將有助於專案知識的組織與利用（賴聯福，1998；Ahn et al., 2004; Brezillion & Pomerol, 1999; Wu & Wang, 2007）。Ahn、Lee、Cho與Park（2004）指出專案過程中，組織、人員和活動為組成情境資訊的關鍵元素。研究由組織觀點（organizational perspective）和活動觀點（activity perspective）以保留專案完整生命週期之知識。其中，活動觀點（activity perspective）認為活動是組成情境知識的關鍵要素，活動與專案元素都有關係，如：活

動與成員關係，一位成員可以參加許多活動，一個活動也可以由許多位成員共同完成，每一位成員在活動扮演不同的角色，所以研究以活動為基礎而建立基本情境模式，試著將專案過程中產生的知識可以順利累積保留，並透過知識導覽（navigation）的方式順利傳遞知識。該方法的優點是利用情境的方式將知識和專案流程結合，以完整保留知識情境脈絡。相關研究提出適切的觀點保留與散佈專案知識，但是缺乏整合專案里程碑於專案知識擷取、發掘與利用程序之探討；此外亦缺乏結合情境資訊於主動資訊遞送方法的設計。本研究預計達成以下目的：

- (一) 萃取歷史專案資訊並挖掘其潛在有用知識：研究主要蒐集專案之案例與工作資訊並從中擷取重要資訊，以提供工作者在執行類似工作時有參照之藍圖。計畫首先由歷史專案執行經驗與參照軟體工程標準（SPMP, IEEE Std. 1058-1998），分析專案執行過程中各階段所需的相關知識物件，如：專案成員、專案所需工具、活動、參考標準、文件與專案目標…等。接著，透過知識工程的流程，由歷史工作集合中萃取、分析與呈現工作特定情境的資訊。
- (二) 利用文件探勘技術建立專案語料庫：由於專案文件中隱藏豐富資訊，研究透過文件探勘技術，建立專案語料庫以進行後續之資料分析。
- (三) 利用資料探勘技術萃取重要專案屬性並分析屬性關聯：運用資料探勘技術，

由歷史專案分析與擷取重要情境屬性資訊，如：專案參與成員、專案所需工具、活動、參考標準、…等，並探勘屬性間重要的語意關聯。研究預期能提供工作者完整的資訊脈絡，協助其有效率的搜尋專案相關內容並進而解決問題。

貳、文獻回顧

一、知識管理系統支援專案計畫

知識管理系統（Knowledge Management System, KMS）的佈署與建置被視為有效增進組織知識資產並提升企業競爭力之價值活動。Davenport與Prusak（1998）指出「組織知識之累積與管理必須透過資訊科技與人的結合，並且在知識共享的組織文化下，始能創造倍數的效果。」亦即資訊科技除落實知識管理活動亦加速了知識管理流程。專家系統（Expert System）、資料倉儲（Data Warehousing）、資料探勘（Data Mining）、決策支援系統（Decision Support System）與文件管理系統（Document Management System）為目前企業廣為採用之資訊技術（Nemati et al., 2002）。組織環境中，專案的執行為重要的知識管理活動之一，因此，如何透過知識管理流程與策略以有效的支援專案計畫的執行亦為重要的研究議題。此外，專案計畫包含了多項知識密集性高的工作，其執行過程中涉及較多型態的知識工作者，而工作者在執行工作期間亦須參閱大量的文件以深入瞭解工作的內涵，並獲取專業的知識與技能。

由於專案的執行為人員、方法、工具、系統…等的組合，因此，如何整合這些相關資源以達成有效的專案知識累積與再利用為重要的研究議題（Rubenstein-Montano, 2000; Tah & Carr, 2001）。目前相關的專案管理系統多半著重在專案的保存與累積、忽略由歷史專案集中探勘隱含資訊的重要性（Rubenstein-Montano, 2000）。此外，相關專案管理研究指出，專案知識的再利用與分享有賴於知識地圖的建構（Kim et al., 2003; Park & Hunting, 2003）。因此，本研究試圖提出完整的專案知識管理架構以支援後續專案地圖的建構與專案隱涵知識的探勘；即本研究目的之一在協助工作者透過對過去類似專案的瞭解，能快速解決其執行專案時所遭遇的問題。

二、結合工作情境資訊的知識管理

由於組織的活動主要以工作（task）為出發點，因此，若能考量工作特性與知識工作者資訊需求，勢必能提供有效的知識搜尋服務（Abecker et al., 2000; Fenstermacher, 2005; Staab & Schnurr, 2000）。目前從兩個面向探討以工作為基礎的知識支援服務，分別為知識的密集度（knowledge-intensive）與流程的複雜度（process complexity）（Eppler et al., 1999）；此外，為達成支援知識密集度高且流程複雜度高的工作，相關研究更提出結合工作情境的知識遞送（context-based knowledge delivery）（Abecker et al., 2000; Ahn et al., 2005; Liu et al., 2005）。

相關研究從工作環境之工作情境（context）觀點出發，分析組織工作特性，並根據工作者目前工作情境給予有效的工作導引與協助（Abecker et al., 2000; Ahn et al., 2005; Celentano et al., 1995; Staab & Schnurr, 2000）。Abecker、Bernardi、Maus、Sintek與Wenzel（2000）提出KnowMore Project，該雛形系統之特色為提出KIT（knowledge-intensive tasks）之概念，KIT為一般工作流程活動（workflow activity）中之知識密集性工作，系統以工作規格書（task specification）維護該工作之重要資訊。因此，系統可透過規格書啟動各式代理人（agent）搜尋工作相關知識給使用者。由於該系統是透過與流程的結合達到資訊蒐集之目的，因此，其主要貢獻為透過對資訊脈絡（context-aware）之認識，啟動價值流程相關資訊並主動提供工作相關（task-specific）資訊給使用者。此外，Kabiria系統（Celentano et al., 1995）強調工作環境中的文件搜尋系統需要考量辦公室之操作環境，並提出考量文件相關脈絡（context）知識的文件搜尋模式。該系統在提供工作者工作相關文件時，將同時考量文件靜態知識（如：文件內容、型態…等等）、操作性知識（如：工作相關活動、執行人員…等等）與領域知識（如：文件使用規則、法律…等等）。相關研究主要貢獻為透過流程瞭解工作資訊脈絡（task context），並啟動價值流程相關資訊，主動提供工作相關知識給使用者。該型態之知識遞送方法適合流程複雜度

較高的工作，其方式為透過流程或工作之資訊脈絡以建構靜態工作特徵檔（profile or specification），提供工作相關知識。該模式主要限制為若使用者的資訊需求沒有定義在靜態工作特徵檔中，系統將無法提供知識遞送服務；此外，該模式缺乏依據工作者動態的資訊需求，調整工作特徵檔以提供滿足工作現況的相關知識。

三、資料探勘與文件探勘相關技術

(一) 資料探勘技術

知識發掘（Knowledge Discovery）為由大量資料進行歸納與分析，以自動化的探勘流程，發掘尚未被發掘或是有潛在利用價值的規則、知識或資訊的一連串過程並進一步輔助企業做決策（Fayyad et al., 1996; Han & Kamber, 2006）。知識發掘根據所要挖掘資料特性的不同，可分為資料探勘（data mining）與文件探勘（text mining）。資料探勘（data mining）主要是從大量資料中發掘潛藏的有用資訊，以供決策人員參考；相關研究領域包含資料庫、知識學習、人工智慧、專家系統、特徵識別、統計學及資料視覺化…等（Chen et al., 1996; Han & Kamber, 2006）。資料探勘主要程序為資料清潔（data cleaning）與整合（integration）、資料轉換（transformation）與選擇（selection）…等；而資料探勘的技術包含關聯規則（association rules）、序列樣本（sequential patterns）、資料分類（classification）及資料分群（clustering）…等。

關聯法則是資料探勘主要的技術之一，關聯規則分析可以從大量資料中找尋資料項目（如：購買物品）之間的相關性，典型的應用如：超市購買行為分析…等。關聯式規則提供支持度（表示項目出現的頻率）及信賴度（表示項目間的強度），以表示關聯式規則的強度（Agrawal et al., 1993; Chen et al., 1996; Han & Kamber, 2006）。Agrawal、Imieliński與Swami（1993）所提出關聯法則定義如下，令 $I = \{i_1, i_2, \dots, i_m\}$ 為一群項目的集合（Items）， D 是一群交易（Transaction）所組成的集合，每筆交易 T 是由一群項目 I 中任意項目的子集合，若一個項目組包含有 k 個項目，稱之為 k -項目組（ k -itemsets），以 $itemset_k$ 表示之， $k \geq 1$ 。在項目組 X 與 Y 之間有一關聯規則被表示成 $X \rightarrow Y$ ， $X, Y \subseteq I$ ，且 $X \cap Y = \emptyset$ ，其中 X 與 Y 分別表示在資料庫中不同的資料項目組，即若購買項目集合 X 時，可能會在購買項目集合 Y ，每一條關聯法可透過支持度（support）及信賴度（confidence）以判斷所找出的關聯法則是否有意義，其公式如下：

$$\text{Support}(X \rightarrow Y) = P(X \cup Y) \quad (1)$$

$$\text{Confidence}(X \rightarrow Y) = P(Y | X) \quad (2)$$

關聯法則最具代表性的方法是 Apriori（Agrawal et al., 1993; Agrawal & Srikant, 1994），主要執行步驟為：(1)逐漸擴展找出所有高頻率項目集即找到支持度大於事先設定的最小支持度的項目集合；(2)透過分析高頻率項目集合以建立合適的關聯規則。透過

關聯法則所找出的規則可能非常多，但是並非所有關聯規則都是很有價值的，使用者可能只對其中一小部分項目集有興趣；因此為了提高資料探勘的效率及實用性，可以採用以限制式為基礎的關聯法則。限制式關聯法則能夠針對使用者所提需求進行資料探勘，即透過事先條件的設定或是約束，預期得到更有興趣或有用的法則（Han and Kamber, 2006; Srikant et al., 1997）；常見的限制如下：(1)知識型態的限制，即由association限制探勘的類型、(2)維度（dimension）與層級（level）的限制、(3)資料屬性的限制、(4)限制術語（predicate）的限制…等等。本研究在未來結合情境資訊的查詢語模式將進行分析與探討。

(二) 文件探勘技術

文件探勘（text mining）與資料探勘不同之處在於資料探勘主要處理結構化的資料，針對資料庫裡面存放的大量文件，找出其特殊或有趣的樣式；資料探勘主要運用於大型資料庫上，針對有特定需求的使用者，提供資料分析與統計功能、關聯分析、預測…等。文件探勘主要針對半結構或非結構的文件資料進行分析，可以有效的處理以文件模式存放的大量資料，發掘深藏在文件資訊中的知識。Sullivan（2001）定義文件探勘為一種編輯、組織及分析大量文件的過程，是為了要提供特定使用者特定的資訊（如：摘要、關鍵字），以及發現某些特徵及其間的關聯。文件探勘是用於在大量的文件庫或是知識庫，針對文件作資訊搜尋、訊息過

濾、知識萃取、知識管理及決策輔助等之用；因此，文件探勘主要藉由傳統資訊檢索、機器學習與人工智慧技術，企圖在大量的文件中歸納及推衍隱含的知識或知識脈絡。

分群技術（clustering）是根據樣本中各項目屬性之相似性進行分群，使群集內具有高度同質性，而不同群集間則有高度異質性。分群技術沒有事先明確定義類別以進行分類，資料是根據屬性間的相似性形成一群。分群法大致上分為階層式分群法（hierarchical clustering）、切割式分群法（partitioning methods）、密度導向分群法（density-based methods）、格狀導向分群法（grid-based methods）、模型導向分群法（model-based methods）（Han & Kamber, 2006）。*k*-Means於1967年由MacQueen提出，因為其簡單易於瞭解和使用的特性是最常用的分群方法。*k*-means是以距離作為相似性的依據，其群集之間是相鄰而不能重疊的，假若有兩個群集相似時，處於邊界的案例只能歸屬於一群集，為了解決此問題，改良的群集演算法紛紛被提出。若將絕對的距離觀念改為相對的機率觀念時，此時每一個案例就會同時屬於所有的群集，只不過視其中所屬各群集的機率高低而已，這類群集演算法稱之為柔性群集，其代表為*E-M*（Expectation-Maximization）演算法。*E-M*演算法即是期望值最大演算法，是一個迭代（iterative）程序，每次迭代分成 *E*（expectation step）與 *M* 步驟（maximization

step)。E 步驟推導給定觀察外顯變項下遺漏變項之條件機率密度函數，用以計算在目前參數估計值下對數概似函數之期望值。M 步驟則根據 E 步驟所得資料求參數之最大概似估計值，相當於完整資料之最大概似估計。E-M 演算法與 k-means 的計算流程類似，唯一不同的是 E-M 是使用高斯分配，如公式 (4) 描述該案例隸屬於某群集密度，以機率函數來取代距離函數。

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (4)$$

參、問題定義與系統架構

一、設計專案階段情境知識平台之動機

新進員工經常藉著諮詢專家以解決問題，若無法找到解決的方法將透過組織的知識庫尋找相關的文件。目前企業積極導入知識管理系統協助保留專案開發過程所有資源，其中文件管理系統主要保留開發專案過程中相關文件，如：討論會議記錄、備忘錄、工作日誌等。此外，目前商業資訊檢索系統與企業知識入口網站，雖提供組織知識庫支援組織各層級的工作者，但主要引進傳統搜尋引擎 (search engine) 技術，以查詢語模式 (query-based) 提供資訊檢索服務。亦有不少系統提供分類目錄，試圖解決傳統關鍵詞搜尋技術低精確度的問題。根據以上模式所進行的檢索，其限制為無法有效解決系統非透過化所造成的檢索障礙 (即工作者無法在短時間對組織知識結構有充分的瞭解)，另一方面無法於工作者檢索過程中，

適時辨識工作情境主動遞送相關知識。因此，當工作者需要相關資源時，透過系統提供的搜尋或目錄。可能搜尋出大量文件資料或資源，而這些資源未必符合工作者的需求，所以必須耗費更多時間重複搜尋、瀏覽與篩選。本研究著眼於組織中專案執行的複雜度，因此，本研究試圖透過專案情境資訊的輔助以協助知識工作者釐清專案的知識脈絡，進而解決工作中所遭遇的問題。此外，研究中亦提出對外顯知識，例如：計畫、專業文獻、know-how、專案、專利、與會議記錄...等等，透過知識工程的程序並參照相關標準，將非結構化的資訊轉換為以結構化方式加以保存，以協助組織的知識再利用與分享，茲將本研究重要定義分述如下：

- (1) 情境資訊定義與擷取：參照技術標準，設計與規範專案工作所需的相關情境資訊觀點。研究由歷史專案執行經驗並參照軟體工程標準 (參照 SPMP, IEEE Std. 1058-1998)，分析專案執行過程中所需的相關知識物件並定義專案的情境資訊模型 (project-in-context, PIC)。該模型主要包含兩個層面：專案資源領域觀點 (project resource domain) 與專案知識領域觀點 (information domain)，前者定義各項專案物件之間的關聯，而後者定義不同專案里程碑或工作階段所需參照之相關知識。模型中之專案或知識物件的屬性資料，將以詮釋資料方式定義與描述，如附錄一所描述。詮釋

資料是對專案相關物件屬性的一組描述，目的在促進資訊系統中對資料之檢索、管理與分析，進而促進知識的交換與分享。

- (2) 專案情境資訊利用：結合搜尋與探勘技術，萃取重要專案屬性並分析屬性關聯。研究提出滿足工作者特定情境資訊需求 (contextually project-specific information) 的概念，該想法主要源自專案執行期間其特定情境資訊將根據專案里程碑/工作階段而有所改變，亦即系統將根據工作者處於不同的工作階段而支援不同的情境資訊，以輔助其完成工作的階段性任務。本研究提出之系統由主動搜尋與被動遞送的角度設計與探討專案情境資訊利用之實際運作模式。因此，將由專案、工作與文件等不同層級 (granulites) 所擷取的情境資訊，透過關聯規則探勘 (association rule mining) 與限制式的關聯規則探勘 (constraint-based association rule mining) 方法，分析專案物件關聯以推理並過濾出使用者所需的資訊。基礎的情境資訊推理，如：工作者試圖找到「某項工作技能」與「具有此技能的工作者」的關聯、「某個工作主題」與「研究主題相關的文章作者」等關聯。此外，系統亦可根據使用者查詢語與工作情境資訊，如：所處的專案階段、遭遇問題…等，推論使用者的資訊需求，呈

現與情境資訊相關之專案核心知識或最佳案例…等知識。

二、系統架構

本研究之系統架構分為三個階層：知識蒐集與擷取、知識挖掘與呈現與專案發展之知識利用，如圖一所示，分別說明如下：

(一) 知識蒐集與擷取

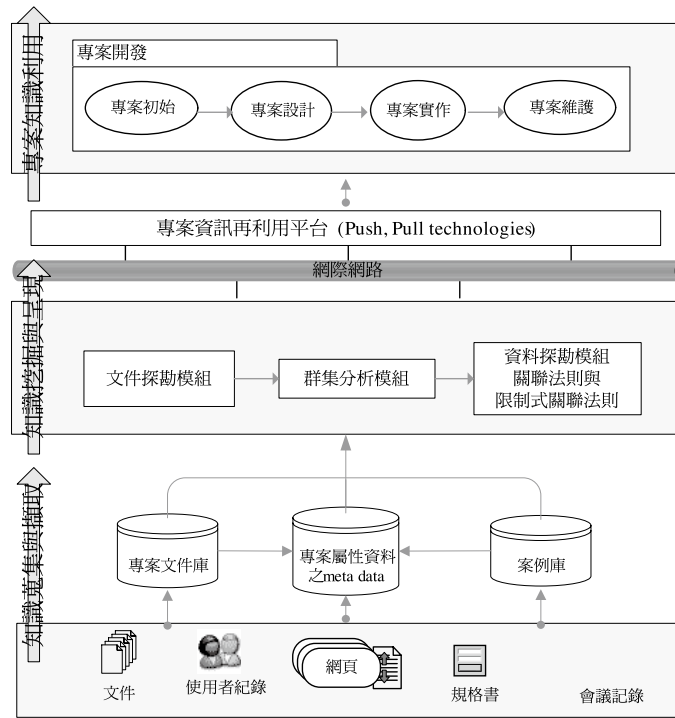
研究主要依不同專案階段蒐集資料與資訊，每一個專案階段會因為階段特性的差異而需要或產生不同的資訊、使用不同工具或諮詢不同的對象，若能將每個階段產生的資訊依當時的工作情境而保留下來，將可有利於後續相似專案或工作的執行。

(二) 知識挖掘與呈現

研究主要運用文件探勘、分群與限制式關聯法則技術與理論，對過去專案進行分析並擷取專案情境的語意關聯。如：成員在專案中扮演不同角色，而每個角色所遇到的問題不一樣，透過關聯法則可以找到關係，相關專案情境參數與規則定義將於文後敘述之。

(三) 專案發展之知識利用

設計以專案階段情境為基礎之專案資訊利用平台，透過所探勘出之專案情境規則以促進有效的知識存取。一方面系統可根據使用者之工作情境主動提供專案里程碑與工作階段相關知識；另一方面，當使用者主動搜尋專案相關資訊時，系統將結合情境資訊，有效過濾資訊。由於該資訊呈現方式考量各



圖一 專案資訊再利用系統架構圖

個不同專案的特性與不同專案於各工作階段所需的資訊，因此，可以滿足工作者於不同工作情境的資訊需求，有效利用專案情境資訊。

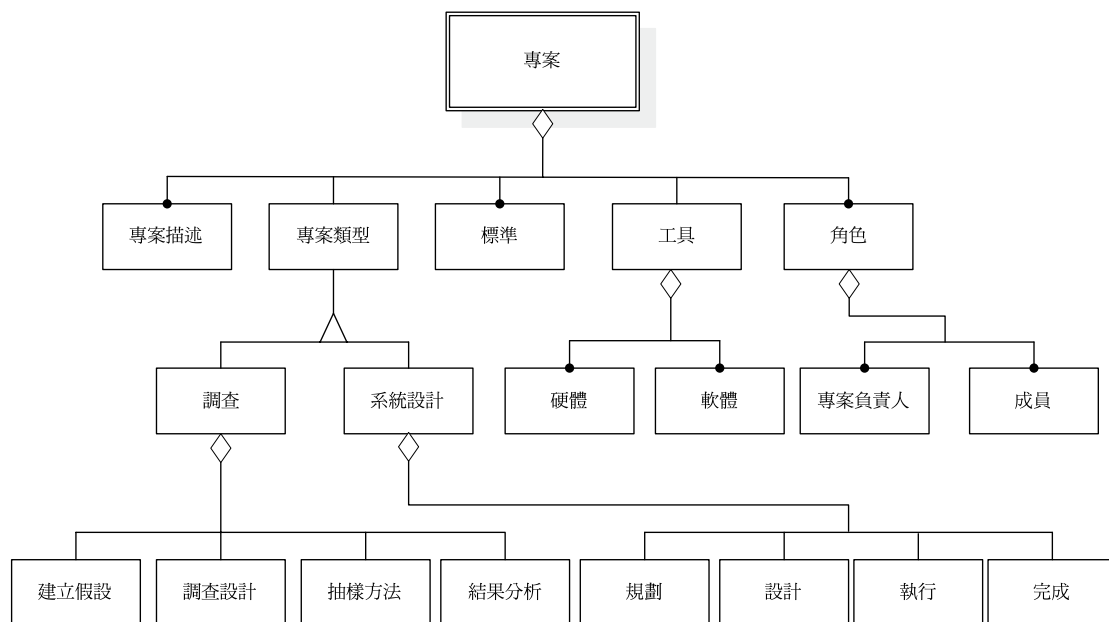
三、情境模型

專案設計是企業的核心工作，為了達成某些目標或目的，成立專案小組進行研發。若每當成立一個專案時，都必須重頭開始規畫專案將浪費許多時間和資源，因此，有效的再利用專案資訊已成為組織知識管理的重要議題。我們使用物件模型化技術（Object Modeling Technique, OMT）（Rumbaugh

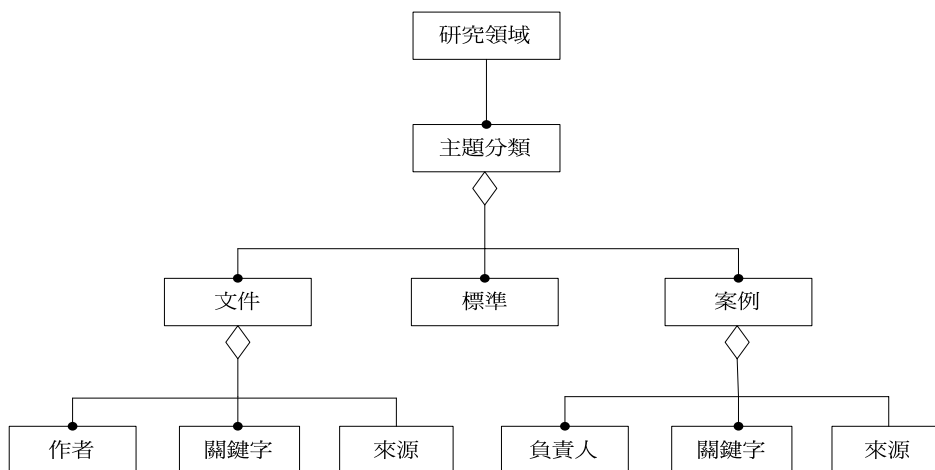
et al., 1991）描述專案情境模型（PIC model），該模型主要包含兩個層面：專案資源領域觀點（project resource domain）與專案知識領域觀點（information domain），如圖二與三所示。本研究根據所建立的專案的情境資訊模型（PIC model），標註重要詞彙的詮釋資料型態（meta-data type）。

（一）專案資源觀點

主要描述領域中專案屬性定義與關聯，即構成一個專案所包含的物件，例如：專案描述、專案主題語料庫、專家、標準、工具、專案類型、...等。研究所設計之專案情



圖二 專案的情境資訊模型 (PIC model) 之專案資源觀點



圖三 專案的情境資訊模型 (PIC model) 之知識領域觀點

境資訊模型 (PIC Model) 之專案資源觀點特點在我們定義專案階段里程碑的概念，根據本研究目的所述，預計有效組織不同專案階段里程碑之重要專案資源。此外，研究後續亦將結合知識領域觀點之專案分類樹與專案里程碑，發掘與分析特定專案物件間之語法與語意關聯，例如：專案成員常用特定工具完成特定類型專案工作的開發，將於第四章中詳述專案情境知識發掘程序。

(二) 知識領域觀點

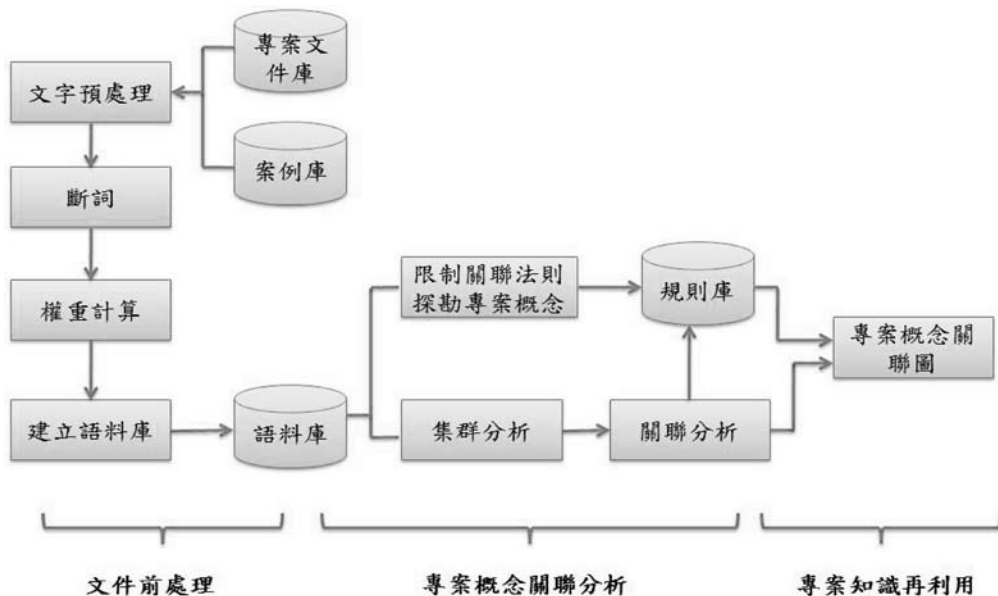
知識領域觀點主要描述研究對象的知識領域範疇，本研究中所建構之專案知識分類樹的第一層為重要的知識領域分類 (field)，第二層為由過去專案所得之重

要知識主題 (topic)，而最下層為分別儲存於專案文件庫與案例庫的專案相關文件 (documents) 與案例 (cases)。該知識領域觀點之最上層為專家或有經驗豐富的工作者所定義，表示專案主要的研究與應用領域，第二層主題主要根據第一階層之專案知識領域並根據過去相似專案分析而得知重要主題。

肆、專案情境知識發掘程序

一、專案知識語料庫建立

本研究的專案概念挖掘程序主要分成，專案文件前處理、專案概念分析關聯和專案知識再利用，如圖四所示。



圖四 專案概念挖掘程序

(一) 專案文件前處理

本研究蒐集資料來源為輔仁大學資訊管理學系，九三至九七學年度期間大學畢業資訊系統專案，研究自專案文件中擷取摘要、知識背景以及問題集…等，分別依以下程序處理以建構專案文件庫和案例庫供後續分析使用。

1. CKIP斷詞處理

中文斷詞是將中文語句依語意與文法結構切割至以詞為基本單位的動作，由於從單一中文字瞭解其代表意義較為有限，相較之下由兩個以上單字所組成的詞能提供之解釋性較單字來得高，且能降低逐字比對所可能產生的誤判情形，因此絕大部分的斷詞研究都是以詞為單位。系統採用中研院CKIP (Chinese Knowledge and Information Processing) 小組提供之中文斷詞服務進行文章斷詞 (Chen & Liu, 1992)。使用CKIP線上斷字系統前須先向中研院登記申請使用API Key，本系統自行製作介面連結到CKIP線上斷字系統，系統利用TCP Socket送出文本、資料的交換方式採用XML，連線傳送驗證資訊及文本至伺服器，伺服器經過處理後經由原連線傳回結果。研究藉由CKIP的斷詞服務將不必耗費多餘人力與時間建置詞庫，只需將斷詞的結果與相關的關鍵詞做比對，更能提高搜尋與分析的速度，主要撰寫如圖五以下重要方法。

2. 關鍵詞萃取與語料庫建立

本研究根據不同的詞性挑選字詞，所

挑選的字詞須能代表類別主題的意義，如普通名詞 (Na)、專有名稱 (Nb)、地方詞 (Nc)、時間詞 (Nd)，又專案文件包含一些英文單字，這些英文單字對於某些領域來說也具有代表意義，因此萃取外語標記 (FW)，接下來便進行關鍵詞權重值之計算。研究根據Salton與Buckley (1988) 所定義的權重架構 (TFIDF scheme)，分別計算字彙頻率 (term frequency, TF) 和文件反轉頻率 (inverse document frequency, IDF)，計算每個字詞在文件中的權重。TF指字詞頻率，如果字詞出現在文章之頻率越高則表示越重要；IDF是文件集合頻率的倒數，以字詞出現在其它文件的多寡來衡量詞彙的重要程度，當字詞出現在許多文件則表其代表性較低，反之如果只出現在少數文件，則字詞將具有代表性，公式如下。

- (1) 詞頻 (term frequency, TF)：表示計算關鍵詞在某文件的出現頻率，關鍵詞 T_j 代表在文件 D_i 出現的頻率， TF_{ij}
- (2) 文件反轉頻率 (inverse document frequency, IDF)：主要衡量關鍵詞在所有文件集中是否具有鑑別力。IDF_j 表示關鍵詞 T_j 在所有文件裡內的權重值。 N 代表所有文件的總數， df_j 代表有出現關鍵詞 j 的文章總數。

$$IDF_j = \log_2 \left(\frac{N}{df_j} \right) \quad (5)$$

- (3) 研究採用TFIDF權重架構以代表關鍵詞 T_j 在文件 D_i 的權重 (weight)，TFIDF同時考慮到詞彙出現的頻率及詞彙出

-
- (1) `setRawText(String)`：接收欲斷字之專案文件，準備將原始字串傳入中研院 CKIP 斷詞服務（尚未轉成 XML 格式）。
- (2) `getRawText`：為 `setRawText(String)` 之對應方法，檢查傳出字串是否正確。
- (3) `send()`：將傳出字串封裝為 XML 格式並送出至 CKIP 線上斷字系統進行斷字服務，以下為線上斷字系統中研院所規定之格式。
- ```
<?xml version="1.0" ?>
<wordsegmentation version="0.1">
<option showcategory="1" />
<authentication username="iis" password="iis" />
<text>台新金控 12 月 3 日將召開股東臨時會進行董監改選。</text>
</wordsegmentation>
```
- (4) `getReturnText()`：取得回傳的 XML 格式的字串，該字串為斷詞結果並標註斷詞後的詞性，如下所示。傳回後的字串會比原先傳出字串更長，因此需設定較大的 buffer 接收斷詞後的字串。
- ```
<?xml version="1.0" ?>
<wordsegmentation version="0.1">
<processstatus code="0">Success</processstatus>
<result>
<sentence> 台新(N) 金控(N) 1 2 月(N) 3 日(N) 將(ADV) 召開(Vt) 股東
(N) 臨時會(N) 進行(Vt) 董監(N) 改選(Vt) ，(COMMACATEGORY)
</sentence>
</result>
</wordsegmentation>
```
- (5) `String[] getSentence()`取得回傳的句子陣列：將透過 API 所提供之斷詞服務處理過後得到 XML 文件資料，經由此物件轉換為處理過後的句子陣列。此外，中研院 CKIP 以句號“。”來區別一個句子。
- ```
<sentence> 台新(N) 金控(N) 1 2 月(N) 3 日(N) 將(ADV) 召開(Vt) 股東
(N) 臨時會(N) 進行(Vt) 董監(N) 改選(Vt) ，(COMMACATEGORY)
</sentence>
```
- (6) `Term[] getTerm()`取得斷詞後的詞陣列：此物件可透過中研院的 API 所提供之斷詞服務，取得先前傳送文件內容字串中重要的中文關鍵詞，並透過此物件接收斷詞後的字詞陣列。
- (7) `CountWordFreq`統計詞出現的數量：此物件提供了統計字詞結果數量的方法，將 CKIP 斷詞後的字詞陣列透過計算求出文獻資料中字詞所出現的次數頻率。
- 

圖五 CKIP重要函式

現於文件集合的多寡以衡量詞彙的重要性。

$$W_{i,j} = TF_{i,j} \times IDF_j \quad (6)$$

研究選擇每篇文件中權重高於門檻值之字詞建立語料庫 (corpus) 並根據對應之專案儲存。

## (二) 專案概念關聯分析

### 1. 群集分析與命名

由關鍵詞權重值可以得知文章中哪些字詞較為重要，再透過EM群集演算法，找出主題相似之群組。根據先前研究，用來表示文件最適當的關鍵詞個數為10~25個為較佳

(Koller & Sahami, 1997)，過多的關鍵詞反而會降低重要關鍵詞的顯著性。因此本研究從每篇文章挑選出15個關鍵詞來代表文章的特性。研究將文件資訊轉換成向量空間模型，將文件視為由一組字詞所構成，表示為  $D_i = (T_1, T_2, \dots, T_n)$ 。本研究利用SQL Server 2005進行群集分析，將文件對應文字之空間向量矩陣為資料來源，如表一所示。EM群集分析結果如圖六所示，圖六中群集之間的連線表示群集彼此的關聯。表二所示為五個群組之群內相似度計算結果，群集顏色越深表示群內相似度較高，因此群集1和群集2其群內相似度比較高，分別為0.7105與0.5787，文章所談論的主題也為類似。研究由專家根據每群的群組屬性特徵分別為群

組命名，試圖找出主題特色，如群集1之群體中心關鍵詞為：適應值、IGA、評比、演化、染色體、適應力、Model、交配、GA、適應、突變、持股、最佳解、遺傳、交配率、進化論、基因、基金、乖離率、RSI、平均線、RFM、投資、績效、機制、股價、股市、股票、報酬率、演算法、投資人…等，可以得知此群體特徵大多為關於股市或是投資方面之相關概念，因此本研究將此群集重新命名為『投資理財』。

## 2. 以關聯法則進行概念關聯分析與概念關聯圖建立

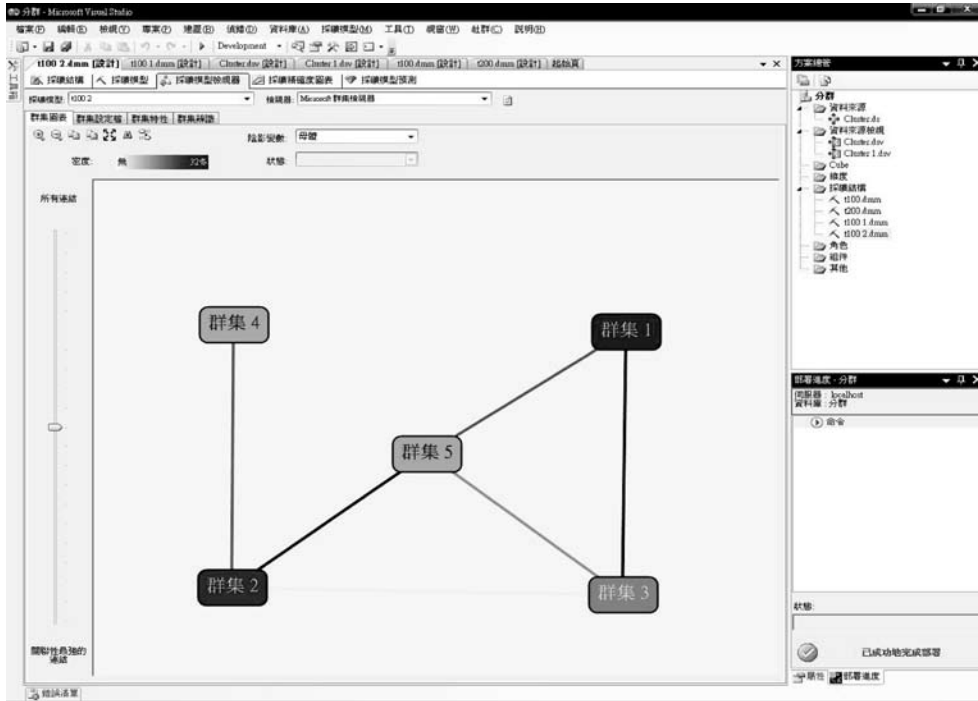
本研究分別針對分群後五個領域進行關聯分析，試圖找到專案與專案間概念關聯。

表一 空間向量矩陣

	基金	知識	標籤	金鑰	RFID	DES	投資	...	Tn
D1	0	1	1	0	1	0	0	...	1
D2	0	0	1	1	0	1	0	...	0
D3	0	1	0	1	1	1	1	...	1
...	...	...	...	...	...	...	...	...	0
D60	1	1	1	1	0	0	0	...	1

表二 群集相似度與命名

群集	群集名稱	群內相似度
群集 1	投資理財	0.7105
群集 2	文件與資料探勘	0.5787
群集 3	電子商務	0.5094
群集 4	醫療	0.4851
群集 5	資訊技術與應用	0.4063



圖六 文件群集分析結果圖

研究根據權重計算公式計算關鍵字詞權重植，設定門檻值是根據每個領域裡文章全部字詞的權重加總之後取平均數當作門檻值，並選出具有意義之字詞，作為後續關聯法則的語料庫資料來源。本研究採用Apriori演算法計算概念詞之間的關聯並進而建立初步的概念關聯網路。以投資理財領域為例，首先計算全部文章字詞之權重值，接著取平均數作為門檻值，因此門檻值為0.08；系統設定支援度為0.4，找出所有支持度值高於最低支持度門檻值之高頻物件集合。針對每個高頻項目找其規則，因此設定信賴度為0.5，結果如表三所示。

系統若是將經由Apriori演算法所得到的全部規則進行概念關聯圖建立將非常雜亂且不易使用。由於信賴度（confidence）越高表示規則具有高度準確但並非信賴度高就等於是有意義的規則，因此研究利用公式(7)，針對所有規則分別計算其重要性，設定重要性必須大於零，其規則才被視為具有意義的將反向規則剔除，來簡化概念關聯圖，提高關聯的正確性。

$$\text{Importance}(A \Rightarrow B) = \log\left(\frac{p(A|B)}{p(A|B)}\right) \quad (7)$$

概念關聯圖是具有方向性之圖型，箭頭方向表示關聯方向，單向箭頭表示關聯是由另一方向指向箭頭的另一端，如若有規則

表三 投資理財前30名之規則表

信賴度 (Confidence)	規 則	信賴度 (Confidence)	規 則
0.74818	達爾文→進化論	0.69897	資本→指數
0.74818	GA→達爾文	0.69897	報酬，指數→資本
0.74818	達爾文→GA	0.69897	資本，指數→報酬
0.7482	進化論，GA→達爾文	0.6990	資本，報酬→指數
0.7482	達爾文，GA→進化論	0.6990	報酬，報酬率→資本
0.7482	達爾文，進化論→GA	0.6990	資本，報酬率→報酬
0.7482	達爾文，進化論→Evolutionary	0.6990	報酬，投資→資本
0.7482	達爾文，生物→進化論	0.6990	資本，投資→報酬
0.7482	進化論，子代→達爾文	0.6990	報酬，股票→資本
0.7482	達爾文，子代→進化論	0.6990	資本，股票→報酬
0.7482	達爾文，演算法→進化論	0.6990	資本，報酬率→指數
0.7482	進化論，基因→達爾文	0.6990	指數，投資→資本
0.7482	進化論，演算法→達爾文	0.6990	指數，報酬率→資本
0.7482	達爾文，基因→進化論	0.6990	資本，投資→指數
0.6990	風險，股票→資產	0.6990	指數，股票→資本

為“達爾文”→“進化論”信賴度為0.7482，其意義為文章出現“達爾文”並會出現“進化論”的機率。若雙向箭頭表示兩者關聯密切，舉例來說，若有兩個規則分別為“達爾文”→“進化論”，和進化論→“達爾文”，其關聯強度都為0.7482，表示達爾文與進化論經常共同出現於文章中，兩個概念詞具有高度關聯。研究未來亦將考慮該節點為權威（authority）節點或聚集中心（hub）節點以分析該節點為一般化或特定的概念。

### 3. 以限制關聯法則探勘專案情境資訊關聯

本步驟主要是透過事先條件的設定或是

約束，以定義不同關聯型態並建立限制式關聯規則樣版。而該樣版主要參照本研究第一部分之專案情境模型定義的專案的情境資訊模型（PIC model）與附錄一所示的專案物件屬性資訊描述（Metadata Structure）而設計。本研究定義之限制關聯樣版包括：人↔領域、人↔主題、人↔工具、領域↔主題、領域↔工具和主題↔工具。其項目意義為人表示屬於特定專案的人員、領域表示描述專案的應用領域、主題表示描述專案使用的方法或是其專業領域相關主題、工具表示描述專案使用哪些工具開發專案，定義如下：



- 人（角色）：屬於特定專案的人員，此為專案的情境資訊模型（PIC model）之專案資源觀點所定義之人（member）或角色（role）。
- 領域：描述專案的應用領域，此為專案的情境資訊模型（PIC model）之知識領域觀點所定義之所屬知識領域（domain knowledge）。
- 主題：描述專案使用的方法或是其專業領域相關主題，此為專案的情境資訊模型（PIC model）之知識領域觀點所定義之主題知識（topic）。
- 工具或標準：描述專案使用哪些工具開發專案，此為專案的情境資訊模型（PIC model）之專案資源觀點所定義之工具（tool）或標準（standard）。

本研究根據將擷取以上專案屬性資訊，附錄二所列為本研究之相關屬性的實體資料（instance）。研究主要藉由限制關聯條件之方式找到對應專案屬性之規則，如表四所示；接著建立概念圖讓使用者可以透過單一主進行相關項目的擴展以更清楚瞭解該主題之相關資訊，即研究所謂之情境資訊關聯。

## 二、專案情境知識呈現與利用

### (一) 知識推力：專案里程碑核心知識遞送

知識推力（knowledge push）策略為藉由歷史專案所建立的專案情模型分析過去專案的核心基礎知識，將其核心基礎知識呈現給工作者。系統可協助瞭解工作者相關知識，對於有特定問題的工作者，亦可以利用

工作情境，找到過去相似問題的解決方法與知識。本研究結合主題地圖的概念，於概念圖的每個重要節點都提供相關資訊，使用者可以點選節點以瞭解主題並可透過節點間的關聯知道其它相關概念。而系統將依照關聯強度排序，列出五個關聯較強的詞彙，讓使用者可以針對這些詞彙作相關的延伸閱讀，如圖七當使用者點選「達爾文」時，就可以知道達爾文在理財投資領域出現之意義；因為達爾文提出「進化論」概念，因此

有人將此概念運用在理財投資方面；此外，研究又提及「基因」、「演算法」…等應用。故研究可以透過相關詞彙知道「達爾文」常與某些概念一起出現，此外，系統可以主動提供使用者過去哪些專案已經運用相同概念的資訊或知識，讓資訊有效再利用。圖七之知識地圖（Knowledge Map，簡稱K-Map）為本研究以關聯法則進行概念關聯分析並進而建立的概念關聯圖。

### (二) 知識拉力：瀏覽與檢索相關知識

在知識的拉力（knowledge pull）部分，系統除提供使用者基本的站內與站外搜尋功能，並考量專案情境特徵資訊，提供滿足使用者工作情境的專案相關知識。研究將根據之前定義的限制關聯樣版實作關聯查詢，使用者若要查詢特定屬性資訊關聯，可以先選擇查詢項目，如：作者、領域、主題或工具，系統將根據探勘結果以概念圖方式呈現相關資訊，讓使用者可以有效瀏覽現在資訊或根據相關屬性資訊進行檢索。圖八之限制式搜尋（constraint-based search，簡稱

表四 限制關聯前20名之規則表

信賴度 (Confidence)	規 則	信賴度 (Confidence)	規 則
0.6092	露天拍賣, 奇摩拍賣→ MySQL	0.6092	醫療, Tomcat→ MicrosoftSQLServer2005
0.6092	露天拍賣, MVC模式→ MySQL	0.6092	醫療, Tomcat→ MicrosoftSQLServer2005
0.6092	露天拍賣, eBay→MySQL	0.6092	醫療, Tomcat→ MicrosoftSQLServer2005
0.6092	露天拍賣→MySQL	0.6092	醫療, RFID→ MicrosoftSQLServer2005
0.6092	醫療, 邱瑞科→PHP	0.6092	醫療, RFID→ MicrosoftSQLServer2005
0.6092	醫療, 翁頌舜→MySQL	0.6092	醫療, Apache→PHP
0.6092	醫療, 邱瑞科→MySQL	0.6092	醫療, Apache→MySQL
0.6092	醫療, 電子標籤→ MicrosoftSQLServer2005	0.6092	醫療, Apache→MySQL
0.6092	醫療, 電子標籤→ MicrosoftSQLServer2005	0.6092	醫療→PHP
0.6092	醫療, 電子商務→MySQL	0.6092	醫療→MySQL

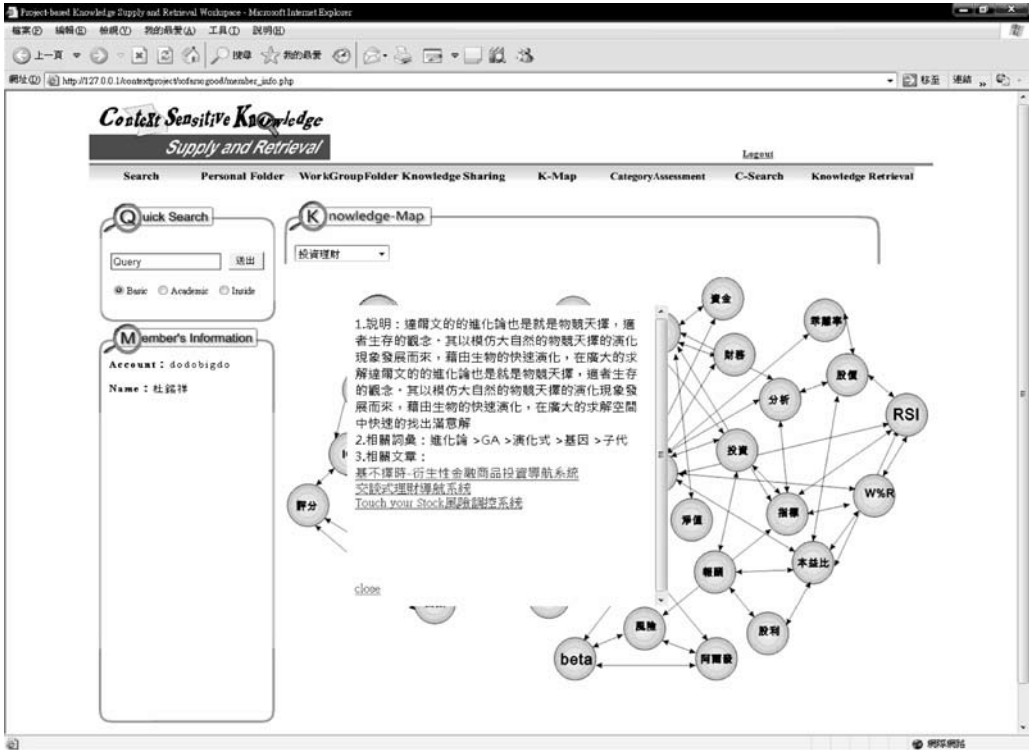
C-Search) 為本研究以限制關聯法則探勘專案情境資訊關聯, 舉例來說, 若使用者想知道某位專家常參與的專案領域、工具、有興趣的主題等, 使用者可以根據系統列出的作者直接選取, 如圖八所示可以知道邱瑞科教授經常參與醫療領域之專案, 常與地理資訊系統和模糊專家系統有關。使用者可以點選網路圖上的結點, 系統將提供過去相似專案與及與該節點有關的資訊。

## 伍、實驗設計與結果

### 一、實驗設計

#### (一) 實驗設計流程

實驗流程主要分成兩個部分, 第一部分先給予受測者任務並要求使用者依序使用傳統介面與研究提出的系統分別完成任務。傳統介面為使用輔仁大學資訊管理學系的歷屆畢業專題系統, 要求受試者先閱讀專案文件或是利用搜尋引擎找尋該領域相關知識, 受試者必須填寫出十個關鍵詞以代表特定領



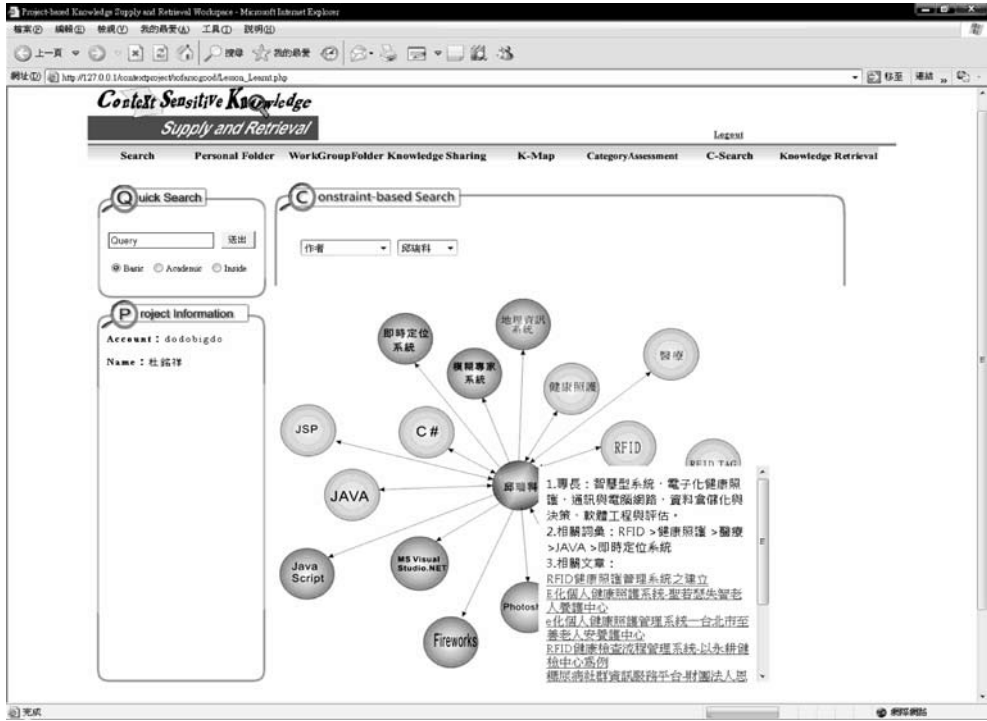
圖七 投資理財概念關聯圖 (K-Map介面)

域。輔仁大學資管系歷屆畢業專題系統平台如附錄三所示，使用者必須選擇年度與感興趣的專題再逐一點選瀏覽。接著受測者再使用本研究系統的K-Map與C-Search介面，以瞭解各領域的專業知識，受測者透過本研究提出的介面可以瞭解各關鍵詞的相關資訊，並針對其前一步驟所列舉的關鍵詞進行修改，若是受測者認為K-Map與C-Search介面所提供的概念更可以表達該領域知識就將其概念詞與之前所挑選的關鍵詞做替換。第二部分主要是受測者使用整體系統後給予的回饋，本研究採用李克特量表 (likert scale) 將問題程度分為五個等級，「非常不喜

歡」、「不喜歡」、「普通」、「喜歡」及「非常喜歡」，給予對應之分數為1至5分，讓受測者使用過系統之後，回答六個問題。

## (二) 實驗規模、參與者與限制

本研究資料來源主要為輔仁大學資訊管理學系九三至九七學年度之專案文件，本研究先經過初步過濾，主要移除專案文件品質不佳且相關研究主題較少的專案，故挑選出六十篇專題文件，因此所研究所建立概念圖可能並未包含所有資訊管理領域，未來將增加文件擴大其概念。本研究主要的實驗對象分為兩類，首先針對輔仁大學資訊管



圖八 邱瑞科教授之概念關聯圖 (C-Search介面)

理學系大學部一、二年級學生，因為這些學生尚未執行專題，不會對關鍵詞有先入為主的觀念，因此可以驗證傳統介面與本研究所提供的K-map與C-search介面，是否可以幫助其找到相關資訊。第二類則為從事「投資理財」與「文件探勘」領域的研究員，主要為研究該領域之碩二學生與教師。研究分別針對集群分析結果中兩個重要領域「理財投資」和「文件與資料探勘」領域進行評估，而每一個領域各找尋五位學生和三位專家分別進行測試。值得注意的是本研究利用關聯法則所找到專案屬性關聯，再依據關聯規則建立專案概念圖，對於其相關

研究領域必須有瞭解，才能對其進行分析與解釋，若是本身不瞭解此領域，所建立出來關聯可能會不符合該領域，因此還是需要專家知識來輔佐，修正專案概念圖。此外，本研究採用來自中研院CKIP中文斷詞系統，且本研究只有針對名詞與英文專有名詞進行分析，其它詞性並沒有納入分析。研究另一個主要限制為以大學部專題計畫為研究對象，我們主要著眼於大學資訊系統開發仍然具有企業專案系統開發之特質，在系統發展過程都需要經歷各個系統開發生命週期，主要包括：需求蒐集與分析、系統分析、系統設計、系統實作與系統測試階段。此外，本

研究亦根據SPMP標準中定義了五種主要的類別資源，包括：成員（member）、目標（goal）、活動（activity）、工具（tool）與標準（standard），並分析專案屬性之間的關聯。

### (三) 衡量指標

本研究利用正確率（Accuracy）與新穎度（Novelty）來衡量系統所提供的關鍵詞是否可以代表該領域知識，如公式(8)與(9)。正確率的定義為所取得的項目中，符合系統之關鍵詞比率，而新穎度定義使用者後來新增檢索的關鍵詞所佔的比率。對於本研究而言，正確率越高代表使用所挑選的關鍵詞跟系統所推薦的領域關鍵詞越接近，而新穎度越高則表示系統所提供的概念越能提供給使用者更多資訊。

$$\text{正確度(Accuracy)} = \frac{\text{使用者所挑選與系統相同的關鍵詞}}{\text{系統該領域的關鍵詞}} \quad (8)$$

$$\text{新穎度(Novelty)} = \frac{\text{使用者選擇之系統回饋關鍵詞}}{\text{使用者所挑選與系統相同的關鍵詞} + \text{使用者選擇之系統回饋關鍵詞}} \quad (9)$$

## 二、實驗分析與結果

以下分別針對「理財投資」和「文件與資料探勘」領域進行評估，分別驗證系統正確度與新穎度。

### (一) 系統正確度與新穎度評估

#### 1. 投資理財領域

根據公式計算關鍵詞的正確率和新穎度，學生與專家結果分別為表五與表六所

示。由學生受試者結果觀之（見表五），其正確率之平均值為0.38，可以知道學生對於「投資理財」領域的概念詞選擇與系統所選擇之概念詞，認知不太相同。然而受測者若進一步藉由本系統提供的K-Map介面對投資理財領域進行瞭解，並再重新挑選重要的相關概念詞，研究透過新穎度計算，發現新穎度平均值為0.68，表示該K-map介面有助於學生受測者獲得更多資訊。由專家受試者結果觀之（見表六），其正確率之平均值為0.60，表示專家之理財投資概念與系統所呈現之專案概念較為相近。在新穎度部分其平均值為0.48，這表示專家較無法由K-Map介面獲得較多新的概念。此外，由標準差觀之，無論學生或專家在正確率或新穎度評估結果皆有很大的共識度。

#### 2. 文件與資料探勘領域

由學生受試者結果觀之（見表五），其正確率之平均值為0.20，可以知道使用者對於「文件與資料探勘」領域的概念詞與系統所呈現之專案概念詞非常不相同。學生受試者的新穎度平均值為0.79，這表示使用者經由操作K-Map介面可以更瞭解文件與資料探勘專案概念領域並獲得更多資訊。兩者的標準差評估，正確率和新穎度的標準差值小且差異不大，表示學生對於文件與資料探勘之概念差異性不大。由專家受試者結果觀之（見表六），其正確率之平均值為0.50，表示專家對「文件與資料探勘」領域的概念詞與系統所呈現之專案概念詞較為符合。在新穎度部分其平均值為0.53，這表示專家經由

表五 關鍵詞評估 (學生部分)

	投資理財領域		文件與資料探勘領域	
	Accuracy	Novelty	Accuracy	Novelty
User 1	0.30	0.73	0.20	0.70
User 2	0.40	0.69	0.10	0.90
User 3	0.60	0.54	0.20	0.83
User 4	0.20	0.8	0.30	0.70
User 5	0.40	0.63	0.20	0.81
平均數	0.38	0.68	0.20	0.79
標準差	0.15	0.10	0.07	0.09

表六 關鍵詞評估 (專家部分)

	投資理財領域		文件與資料探勘領域	
	Accuracy	Novelty	Accuracy	Novelty
Expert 1	0.60	0.45	0.40	0.66
Expert 2	0.70	0.36	0.50	0.52
Expert 3	0.50	0.63	0.60	0.42
平均數	0.60	0.48	0.50	0.53
標準差	0.10	0.14	0.10	0.12

K-Map介面，並無法像學生受試者一樣獲得較多新穎的概念詞。此外，由標準差觀之，無論學生或專家在正確率或新穎度評估結果皆有很大的共識度。

## (二)受試者滿意度評估

### 1. 投資理財領域

學生受試者滿意度評估整理如表七，學生受試者對於「投資理財」領域之六個問題平均分數皆大於3。各問項中，學生對於「K-Map有助於關鍵詞選取」問項，其平均

分數偏低且標準差偏高，該結果與關鍵詞選取正確率 (Accuracy) 略低之結果相仿，可能為學生受試者對於該專案領域認識不深且對相關詞彙所知不多。若對應問項「K-Map呈現專案概念對於專案領域之認知符合」偏低之平均分數，可以知道大部分學生受試者對該領域的認知與系統所呈現的結果有差異。由專家受試者結果觀之 (見表八)，其平均數皆高於4分，即專家普遍認同K-Map有助於使用者瞭解專案。其中專家一致認為K-Map最能助於使用者在關鍵詞選取且

表七 K-Map滿意度問卷結果（學生部分）

投資理財領域		
問 題	平均數	標準差
K-Map（概念地圖）呈現專案概念有助於瞭解專案領域知識	4.00	1.00
K-Map有助於瞭解知識間之關係	4.20	0.84
K-Map可以幫助獲取較多的資訊	3.80	0.45
K-Map有助於關鍵詞選取	3.60	1.14
K-Map呈現專案概念對於專案領域之認知符合	3.40	0.55
K-Map可以有效率瀏覽專案資源	3.80	0.84
文件與資料探勘領域		
問 題	平均數	標準差
K-Map（概念地圖）呈現專案概念有助於瞭解專案領域知識	4.20	0.45
K-Map有助於瞭解知識間之關係	4.20	1.10
K-Map可以幫助獲取較多的資訊	4.20	0.83
K-Map有助於關鍵詞選取	4.00	0.71
K-Map呈現專案概念對於專案領域之認知符合	3.40	0.55
K-Map可以有效率瀏覽專案資源	4.40	0.55

K-Map可以幫助獲取較多的資訊。但專家指出若能改善該地圖之設計與使用性將可以提昇使用者瀏覽專案資源之效率。

## 2. 文件與資料探勘領域

學生受試者對於「文件與資料探勘」領域之各問項滿意度均略高於「投資理財」領域。由表七可以得知使用者認為以K-Map呈現專案概念有助於瞭解專案領域知識且K-Map有助於瞭解知識間之關係，但是學生受試者於「K-Map呈現專案概念對於專案領域之認知不相符合」問項分數較其問項為低。由專家受試者結果觀之（見表八），其結果與學生受試者相仿且與「投資理財」領

域有一樣的問題，即專家建議介面仍有改善空間，故在「K-Map可以有效率瀏覽專案資源」問項分數偏低。

## (三) 討論

綜合以上可以得知學生評估之正確率（Accuracy）平均值低於專家，但學生之新穎度（Novelty）平均值卻高於專家。其主要原因是學生缺少領域之相關知識，因此，所挑選的文章關鍵詞與系統所呈現之關鍵詞差異較大；而專家已經具備該領域之知識背景，以致於所挑選之關鍵詞與系統所呈現概念較為一致。在合理的情況下，學生在新穎度的表現值會比較高，主因為學生在瀏覽過

表八 K-Map滿意度問卷結果（專家部分）

投資理財領域		
問 題	平均數	標準差
K-Map(概念地圖)呈現專案概念有助於瞭解專案領域知識	4.33	0.57
K-Map有助於瞭解知識間之關係	5.00	0.00
K-Map可以幫助獲取較多的資訊	4.67	0.58
K-Map有助於關鍵詞選取	4.67	0.58
K-Map呈現專案概念對於專案領域之認知符合	4.33	0.58
K-Map可以有效率瀏覽專案資源	4.00	0.00
文件與資料探勘領域		
問 題	平均數	標準差
K-Map（概念地圖）呈現專案概念有助於瞭解專案領域知識	4.33	0.58
K-Map有助於瞭解知識間之關係	4.67	0.58
K-Map可以幫助獲取較多的資訊	4.33	0.58
K-Map有助於關鍵詞選取	4.00	0.00
K-Map呈現專案概念對於專案領域之認知符合	4.33	0.58
K-Map可以有效率瀏覽專案資源	3.67	0.58

相關詞彙後可以更肯定新詞彙與領域的相關性。滿意度評估部分，學生認為藉由K-Map呈現專案概念地圖，主要功能為有助於瞭解知識間之關係，其次是助於瞭解專案領域知識；但學生認為K-Map呈現專案概念與其對專案領域的認知不盡相同，可能為學生受試者較缺乏該專案領域的相關知識。專家受試者皆肯定K-Map對認識專案知識的幫助，但建議在介面上做更好的設計。

## 陸、結論與未來展望

本研究提出專案情境知識發掘程序並結合專案階段與知識物件間的關聯，由主動

搜尋與被動遞送的角度設計與探討專案情境資訊利用之實際運作情形。本研究另一貢獻為由專案資源領域觀點（project resource domain）與專案資訊領域觀點（information domain），提出專案情境資訊模型（project-in-context model, PIC），並根據該模型建構結合專案情境知識的知識發掘與再利用平台。研究主要目的在達成有效的專案資訊再利用，即收集專案之案例與工作資訊並從中擷取重要資訊，以提供工作者在執行類似工作時有參照之藍圖。此外，研究協助工作者在不同階段對其執行的專案與工作有較為明確與完整的視野（view）。由組織角度觀



之，該設計透過技術標準，可協助知識的擷取與利用，促進組織內與組織間各專案知識的交換與分享。由個人角度觀之，該系統可協助新進工作者快速的探索所需的專業知識。

後續研究將進一步由使用者層面探討系統使用行為觀點，由程序、知識與行為角度探討專案情境知識搜尋與遞送機制與模組之設計，並進而完成結合專案里程碑與情境概念之專案知識支援系統設計。另一方面，研究預計以語彙關聯網路建立專案主題知識語意地圖，透過主題地圖（Topic Maps）呈現專案相關資訊與知識（Sigel, 2000; Park & Hunting, 2003），以標準化之方式組織專案資源，有助於專案相關資訊的索引、搜尋與呈現，並且亦有助於未來專案資源的交換。

## 致謝

本研究經費承國科會專題研究計畫補助（編號：NSC96-2416-H-030-005）與輔仁大學研究發展處計畫補助（編號：9991A01），僅以此致謝。

## 參考書目

賴聯福Lai, Lian Fu (1998)。應用工作基礎概念圖發展自動化軟體工程*Task-based conceptual graphs as a basis for automated software engineering*。未出版之碩士論文Unpublished master's thesis，國立中央大學資訊工程研究所 Department of Computer Science & Information Engineering, National Central

University，桃園縣Taoyuan County。

Abecker, A., Bernardi, A., Maus, H., Sintek, M., & Wenzel, C. (2000). Information supply for business processes: Coupling workflow with document analysis and information retrieval. *Journal of Knowledge Based Systems, 13*(1), 271-284.

Ahn, H. J., Lee, H. J., Cho, K., & Park, S. J. (2005). Utilizing knowledge context in virtual collaborative work. *Journal of Decision Support Systems, 39*(4), 563-583.

Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 207-216). New York: ACM Press.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. *Proceedings of 20th International Conference on Very Large Data Bases* (pp. 487-499). Morgan Kaufmann: ACM Press.

Brezillon, P., & Pomerol, J. Ch. (1999). Contextual knowledge sharing and cooperation in intelligent assistant systems. *Journal of Le Travail Humain, 62*(3), 223-246.

Celentano, A., Fugini, M. G., & Pozzi, S. (1995). Knowledge-based document retrieval in office environment: The Kabiria system. *Journal of ACM Transactions on Information Systems, 13*(3), 237-268.

- Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: An overview from a database perspective. *Journal of IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866-883.
- Chen, K. J., & Liu, S.H. (1992). Word identification for Mandarin Chinese sentences. *COLING '92: Proceedings of the Fifteenth International Conference on Computational Linguistics* (pp. 101-107). Nantes: ACM Press.
- Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manages what they know*. Boston MA: Harvard Business School Press.
- Eppler, M. J., Seifried, P. M., & Ropnack, A. (1999). Improving knowledge intensive process through an enterprise knowledge medium. *SIGCPR '99: Proceedings of 1999 Computer Personnel Research* (pp. 222-230). New Orleans, LA: ACM Press.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthuramy, R. (1996). *Advances in knowledge discovery and data mining*. California, USA: AAAI Press.
- Fenstermacher, K. D. (2005). A process for delivering information just in time. *Journal of Lecture Notes in Computer Science*, 3782, 679-687.
- Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Hansen, P., & Järvelin, K. (2005). Collaborative information retrieval in an information-intensive domain. *Information Processing and Management*, 41(5), 1101-1119.
- Institute of Electrical and Electronics Engineers. (1998). IEEE standard for software project management plans (SPMP). (Technical report: IEEE Std 1058-1998).
- Kim, S., Suh, E., & Hwang, H. (2003). Building the knowledge map: An industry case study. *Journal of Knowledge Management*, 7(2), 34-45.
- Koller, D., & Sahami, M. (1997). *Hierarchically classifying documents using very few words*. *Proceedings of the Fourteenth International Conference on Machine Learning* (pp.70-178). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- KPMG Consulting (n.d.), *Insights from KPMG European knowledge management survey 2002/2003*. Amstelveen, The Netherlands: KPMG.
- Liu, D.-R., Wu, I.-C., & Yang, K.-S. (2005). Task-based K-support system: Disseminating and sharing task-relevant knowledge. *Expert Systems with Applications*, 29(2), 408-423.
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability* (vol. 1, pp. 281-297). Berke-

- ley: University of California Press.
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company*. New York: Oxford University Press.
- Nemati, H. R., Steiger, D. M., & Iyer, L. S. (2002). Knowledge warehouse: An architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decision Support Systems*, 33(2), 143-161.
- Park, J. & Hunting, S. (2003). *XML topic maps: Creating and using topic maps for the web*. Boston MA: Addison-Wesley.
- PMI Standard (1996). *A guide to the project management body of knowledge*. Pennsylvania, U.S.: Project Management Institute.
- Rubenstein-Montano, B. (2000). A survey of knowledge-based information for urban planning: Moving towards knowledge management. *Computers, Environment and Urban Systems*, 24(2), 155-172.
- Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., & Lorensen, W. (1991). *Object-oriented modeling and design*. Englewood Cliffs, N.J.: Prentice-Hall.
- Salton, G., & Buckley, C. (1988). Term weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513-523.
- Siegel, A. (2000). Towards knowledge organization with topic maps. *Proceedings of the XML Europe 2000*, Paris, France.
- Sowa, J. F. & Zachman, J. A. (1992). Extending and formalizing the framework for information systems architecture, *IBM Systems Journal*, 31(3), 590-616.
- Staab, S., & Schnurr, H.-P. (2000). Smart task support through proactive access to organizational memory. *Knowledge-Based Systems*, 13(5), 251-260.
- Srikant, R., Vu, Q., & Agrawal, R. (1997). Mining association rules with item constraints. *KDD '97: Proceedings of the International Conference on Knowledge Discovery and Data Mining* (pp: 67-73). California, USA: AAAI Press.
- Sullivan, D. (2001). *Document warehousing and text mining* (pp. 326). NJ, USA: Wiley Computer Publishing.
- Tah, J. H. M., & Carr, V. (2001). Towards a framework for project risk management in the construction supply chain. *Advances in Engineering Software*, 32(10), 835-846.
- Wu, I.-C., & Wang, Y.-C. (2007). An exploratory study of capturing and reusing project knowledge from a stage-based perspective. *AMCIS 2007: Proceedings of the Americas Conference on Information Systems*, Keystone, Colorado, USA.

(投稿日期：2010年2月28日 接受日期：2010年3月18日)

附錄一 專案物件屬性資訊描述 ( Metadata Structure )

Entity	Metadata	Description
階層一：專案<projects>		
專案描述	<description>	專案規格書，主要描述專案目標、範圍、需求分析、預估時程、專案工作與任務，...等。
專案主題語料庫	<topiccorpus>*	透過文件前處理技術，發掘不同專案主題之重要詞彙。
角色，標準，工具	<role>*，<standard>*，<tool>*	定義專案之重要知識物件，主要包含專案重要角色、標準與所需工具...等。 SynthesizedfromIEEEStd.1058-1988 (1988)
專案類型	<type>	定義不同專案類型之專案樣本，其中主要包含「系統開發」、「研究調查」與「理論發展」三個類別並各自定義不同的專案里程。後續專案可繼承不同專案樣本而有效再利用過去相關知識與經驗。 SynthesizedfromAhn，etal. (2005) andWu&Wang. (2007) (2008)
所屬領域知識	<domainknowledge>	參考知識領域觀點，領域知識為專案分類樹第一階（頂）層之主題。 AdaptedfromLiuetal. (2005)
重要概念	<keyconcept>*	透過文件探勘技術萃取不同類型專案之重要概念。
階層二：專案里程碑<Milestonesoftheproject>		
專案類型	<type>	參考知識領域觀點，專案類型為專案分類樹第一階（頂）層之主題。 SynthesizedfromAhn，etal. (2005) anLiuetal. (2005)
專案里程碑	<milestone>	專案包含了不同的里程碑，主要為專案的起始、計畫、執行、控制與完成階段。本研究之重點在支援專案計畫起始的資訊蒐集與專案執行所需之相關知識物件。 AdoptedfromPMIStandard (1996) andCIRframeworkforthepatentdomainofHansenandJärvelin (2005)
專案文件	<document>*	存放於專案文件庫之報告、文獻、簡報與手冊等文件資訊。

案例	<case>*	存放於專案案例庫之最佳案例（bestpractices）與經驗（lessonlearnt）。
主題知識	<topics>*	參考知識領域觀點，主題知識為專案分類樹第二階層之主題。 AdaptedfromLiuetal.（2005）
專案主題語料庫	<topiccorpus>*	透過文件前處理技術，發掘不同專案主題之常見詞彙。
階層三：專案活動<Activities>		
任務描述	<activities>*	主要定義不同專案里程碑之執行任務與活動。
角色成員	<role-members>*	定義專案不同活動之角色成員。
側寫檔	<keyword>* <topic>*	為專案知識工作者之資訊需求側寫檔，主要以關鍵詞 與主題分類類描述知識工作者的專案相關資訊需求。

附錄二

(1) 專案屬性之主題知識 (topic)

主題 (topic)			
3D	概 構圖	乖離率	Web 2.0
關聯法則	搜尋引擎	私密金鑰加密法	VR指標
線上購物	傳統搜尋法	即時定位系統	VOD隨選視訊
線上分析處理	部落格	自組織映射圖	Treynor指標
模糊專家系統	統計分析	成交量指標	TFIDF
蟻元系統演算法	移動平均線理論	安全電子交易標準	TF
數位簽章	晨星基金評	地圖	SSL加密認證機制
數位憑證	推論引擎	地理資訊系統	SET安全協定
數位影音串流	專家系統	合作式過濾推薦系統	RSI
顧客關係管理	健康照護	共變異數法	RFM理論
數位內容	高階主管資訊系統	共同基金	RFID
露天拍賣	針灸	仿生物演化分析	PSY指標
類神經網路	秘密金鑰加密法	交談式基因演算法	MVC模式
價格指標	夏普指標	本體論	K線理論
遺傳基因演算法	相關性回饋	平衡計分卡	Kmeans
蒙地卡羅模擬法	威廉指標	文字探勘	KD指標
電子標籤	阿爾發係數	支援向量機	IDF
電子商務	知識管理	分類	Google Maps API
電子市場	知識傳遞	資料探勘	eLearning
道氏理論	知識庫	公開金鑰加密法	eBay
資訊檢索	知識分享	內容導向方法	DES 密法演算法
資訊過濾	波浪理論	內容為基礎的推薦系統	CKIP
資訊安全	拍賣競標	互動式多媒體服務系統	blowfish 加密演算法
資料探勘	奇摩拍賣	三重 DES	Apriori
資料倉儲	協同過濾法	$\beta$ 係數	文字探勘
經絡	協同行銷	WordNet	分群

(2) 專案屬性之人或角色、領域、工具或標準

人 (member or role)	領域 (domain knowledge)	工具或標準 (tool or standard)
吳怡瑾	投資理財	Apache
吳濟聰	文件與資料探勘	ASP
李俊民	電子商務	ASP.Net
林文修	醫療	C#
邱瑞科	資訊技術與應用	IIS
胡俊之		Java
翁頌舜		Java Script
張銀益		JSP
莊雅茹		Microsoft SQL Server 2000
蔡明志		Microsoft Visual Studio.NET
		MySQL
		Oracle 9i
		Tomcat
		XML
		PHP

附錄三 輔仁大學資管系歷屆畢業專題系統畫面



# A Platform for Supporting Knowledge Mining and Reuse Based on Context Information of a Project

I-Chin Wu\*, Shu-Wei Lin\*\*

## Extended Abstract

### 1. Introduction

In the modern business environment, knowledge management activities (process) are regarded as the keys to support and streamline the execution of project-related tasks. Obviously, knowledge that has not been captured cannot be reused; thus, knowledge capture is one of the most fundamental tasks of KM related activities. Accordingly, capturing knowledge from completed projects will govern how knowledge can be reused for the task at hand. Knowledge accumulated in the process of project execution is an important organizational asset. The existing literature has indicated the usefulness of project milestones in organizing

project knowledge (Ahn, et al., 2004; Brezillion & Pomerol, 1999; Lai, 1998; Wu & Wang, 2007). Ahn, Lee, Cho, & Park (2004) pointed out the importance of project context and what constituted essential context information, i.e., information related to organization, members, and activities. However, few studies have incorporated project context information in the design of knowledge management systems. In this study, we developed a project-in-context (PIC) meta-model based on the knowledge extracted from completed projects using text mining and data mining techniques, and we designed a platform that facilitated the capture and reuse of project-specific information based

---

\* Assistant Professor, Department of Information Management, Fu-Jen Catholic University (To whom all correspondence should be addressed.)

E-mail: icwu.fju@gmail.com

\*\* Graduate Student, Department of Information Management, Fu-Jen Catholic University

Note. This extended English abstract is supplied by the JLIS editors and approved by the authors.

To cite this article in APA format: Wu, I. C., & Lin, S. W. (2010). A Platform for Supporting Knowledge Mining and Reuse Based on Context Information of a Project. *Journal of Library and Information Studies*, 8(1), 97-136. [Text in Chinese].

To cite this article in Chicago format: Wu, I-Chin, and Shu-Wei Lin. "A Platform for Supporting Knowledge Mining and Reuse Based on Context Information of a Project." *Journal of Library and Information Studies* 8, no. 1 (2010): 97-136. [Text in Chinese].



on project contexts. This paper describes the model and reported the evaluation results of the proposed platform.

The following aspects were addressed: (1) knowledge acquisition: we analyzed the types of projects and their associated attributes and defined the general yet essential project context information based on the PIC model; (2) knowledge discovery: we used text mining and data mining techniques to extract knowledge items needed by workers and discovered the relationships between various knowledge items; (3) knowledge utilization based on the context: with the proposed model and methods, we developed several applications related to the reuse of project knowledge using pull- and push-based knowledge management strategies to achieve effective project management. The model and system described in this paper can help knowledge workers understand a current research project quickly and resolve problems effectively.

## 2. System Architecture

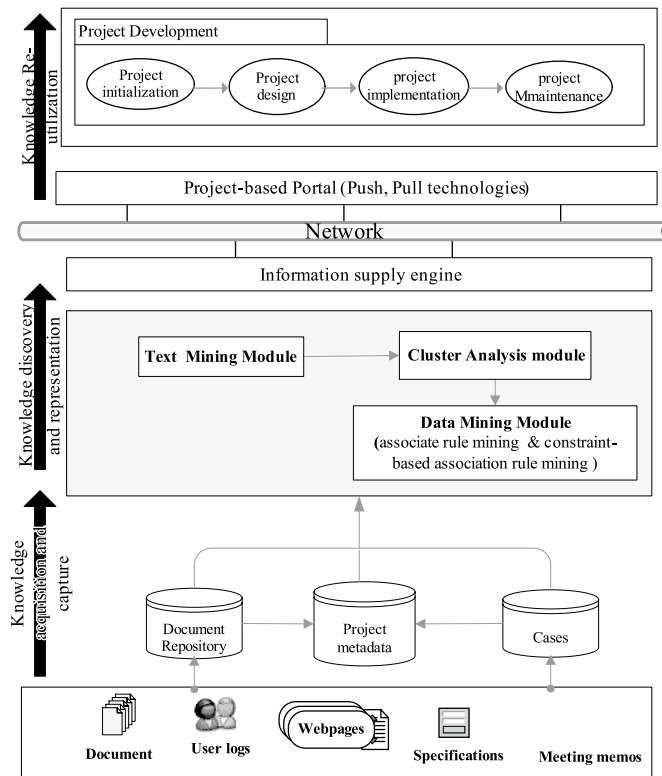
The proposed system was composed of three layers, i.e., knowledge acquisition and capture; knowledge discovery and representation; and knowledge utilization (see Figure 1). Based on completed projects, knowledge was acquired and captured from related project documents and by project stage. Text mining, clustering, and constraint-based

association rule mining techniques were used to analyze the project information and identify associations between context information knowledge items. Finally, the platform promoted the effective organization and reuse of knowledge by project stage.

This study used completed projects and referenced the SPMP standard (IEEE std., 1058-1998) to analyze the knowledge items required in project execution. Based on which, the proposed project-in-context (PIC) model describes a project's relevant contexts from two domains: the project resource domain and the information domain. First, the project resource domain defined the associations between project entities, e.g., project descriptions, project corpora, experts, standards, tools, project types, etc. (see Figure 2). Second, the information domain defined various project milestones and information referenced in different work stages (see Figure 3). We design a three-level meta-data structure based on the XML/XML Schema to fully describe the attributes of the projects as well as the knowledge items in the proposed PIC Model.

## 3. The Mining Procedures of the Project Context Information

3.1 Building the project corpora: this study used the course projects completed by the students of the Department of Information Management in Fu-Jen Catholic University in



**Figure 1 Framework of the Context-sensitive Knowledge Reuse Process**

the course of information systems project during 2004-2008 as the foundation for building the corpora. The following procedures were used to extract summaries, background knowledge, and FAQs in order to construct project document repository and case repository for later analyses.

### 3.1.1 Project documents preprocessing

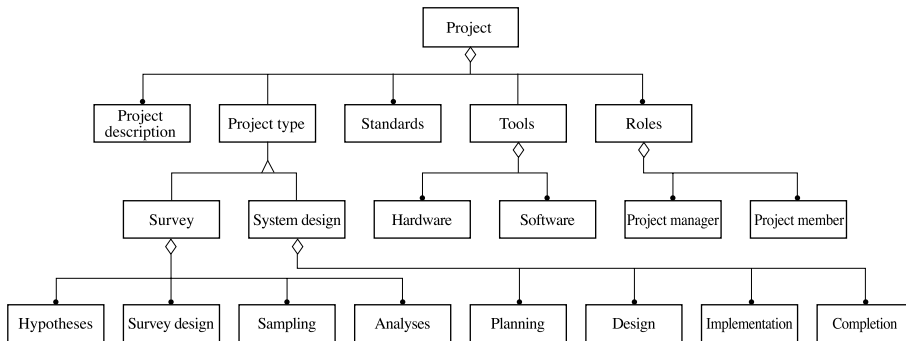
a. Word truncation: this study adopts the word segmentation technology developed by the CKIP (Chinese Knowledge and Information Processing) group of the Institute of Informa-

tion Science and the Institute of Linguistics at Academia Sinica to do word truncation.

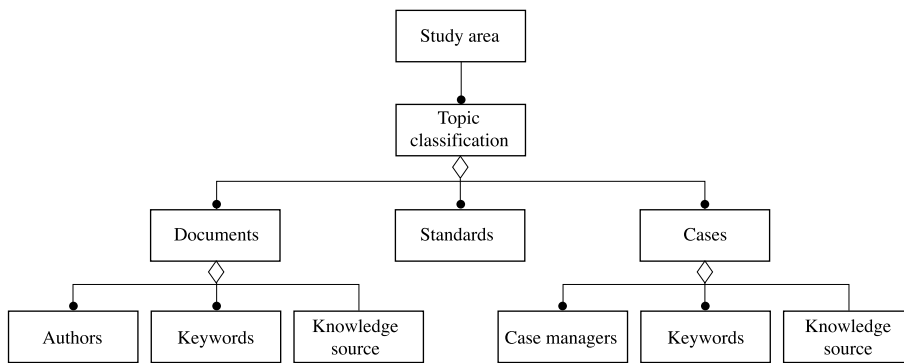
b. Extraction of keywords: the TFIDF (term frequency, inverse document frequency) scheme for *term (keyword) weighting* by Salton and Buckley (1988) was used to extract keywords in order to build the project corpora.

### 3.1.2 Cluster analysis of project concepts

a. Cluster analysis and naming: this study uses SQL Server 2005 to conduct cluster analysis, this study obtained five clusters of concepts,



**Figure 2 The Meta-model of the PIC Model – the Project Resource Domain**



**Figure 3 The Meta-model of the PIC Model – the Information Domain**

among which two contained conceptually highly homogeneous documents. These two clusters were named as “finance and investment” and “text and data mining.”

- b. Analysis based on the association rule: this study employs the *Apriori algorithm* used in associate rule mining to find relationships between key concepts and generate concept networks for the five clusters.
- c. Analyses via the constraint-based association rule mining: with predetermined criteria, this

study specify the association type and adopt constraint-based association rule mining to find frequent patterns among project attributes. This study develops concept maps by which users may understand a certain topic’s contextual association.

### **3.2 The representation and utilization of the project context information.**

3.2.1 *Push-based Knowledge Reuse*: based on the analyses of the completed projects,

the platform was able to deliver the core knowledge of a project to its users. That is, the platform offered essential information based on the project’s milestones.

3.2.2 *Pull-based Knowledge Reuse*: the platform provided users with in-site and global search capabilities and present the mined information in concept maps. It offers useful browsing and effective retrieval capabilities based on specified project attributes.

## 4. Experiment Design and Results

### 4.1 Experiment Design

This study designed a two-phased experiment to test the effectiveness of the interface in proposed system (referred to as K-Map). In the first phase, study subjects were assigned project tasks and were asked to use the traditional project management system interface to complete the tasks. They were asked to write down ten keywords about the project. In the second phase, the subjects were asked to use K-Map to carry out the same tasks. If the subjects considered the keywords provided by K-Map as better than those of the first phase, they would replace the keywords with those identified in the second phase. After the experiment, subjects were asked to rate their use satisfaction in a 5 -scale Likert scale. Furthermore, the study subjects were divided into two groups. The first group (“the

student group”) was composed of freshmen and sophomore undergraduate students, while the second group (“the expert group”) was faculty members and graduate students who represented expert users in the areas of “finance and investment” and “text and data mining.” The evaluation of the keywords provided by K-Map was based on their accuracy and novelty, calculated as follows,

$$Accuracy = \frac{|\text{system-supplied keywords selected by the users}|}{|\text{system-supplied keywords of that area}|}$$

$$Novelty =$$

$$\frac{|\text{user selected keywords via system feedback}|}{|\text{system-supplied keywords selected by the users} + \text{user selected keywords via system feedback}|}$$

### 4.2 Results and Discussion

In both “finance and investment” and “text and data mining” projects, the average accuracy rates of the student group (accuracy in the “finance” project: 0.38; the “text” project: 0.20) were less than those of the expert group (accuracy in the “finance” project: 0.60; the “text” project: 0.50). But the novelty rates of keywords were higher in the student group – novelty in the “finance” project: 0.68; the “text” project: 0.79; versus novelty in the expert group in the “finance” project: 0.48; the “text” project: 0.53. The findings were reasonable. Students lacked the topic knowledge of the projects and thus their selection of keywords varied more

from the system supplied keywords. However, the system apparently offered greater keyword novelty for the students after they browsed the associated keywords in the proposed K-Map. The standard deviations in both groups also showed that the study subjects were rather consensual in the evaluations of keyword accuracy and novelty.

In terms of user satisfaction, the student group considered K-Map as useful firstly in facilitating an understanding of the knowledge association relationships via concept maps, and secondly in providing the topical information of the project. But there were existing gaps between the students' perception of the projects and the subject topics of the projects. This might have been due to students' lack of subject knowledge in those two areas. In contrast, the experts considered K-Map as useful in gaining project knowledge, although they recommended that improvements should be made to the user interface.

## 5. Conclusion

In conclusion, this study found that K-Map, a system offering project context information via knowledge mining techniques as useful in supporting the knowledge management of project execution. For future studies, the investigators will continue to explore the design of the context information discovery and delivery mechanism from the

perspectives of process, content, and the user's behaviors. The investigators also plan to incorporate the concepts project milestones and project contexts to design projects' knowledge support systems. Furthermore, the investigators will continue to construct the semantic maps of topical knowledge and standardize the organization of project knowledge resources. These efforts will enhance the indexing, discovery, and representation of project knowledge and can facilitate the exchange of projects' knowledge.

## References

- Abecker, A., Bernardi, A., Maus, H., Sintek, M., & Wenzel, C. (2000). Information supply for business processes: Coupling workflow with document analysis and information retrieval. *Journal of Knowledge Based Systems, 13*(1), 271-284.
- Ahn, H. J., Lee, H. J., Cho, K., & Park, S. J. (2005). Utilizing knowledge context in virtual collaborative work. *Journal of Decision Support Systems, 39*(4), 563-583.
- Agrawal, R., Imieliński, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 207-216). New York: ACM Press.
- Agrawal, R., & Srikant, R. (1994). Fast

- algorithms for mining association rules. *Proceedings of 20th International Conference on Very Large Data Bases* (pp. 487-499). Morgan Kaufmann: ACM Press.
- Brezillon, P., & Pomerol, J. Ch. (1999). Contextual knowledge sharing and cooperation in intelligent assistant systems. *Journal of Le Travail Humain*, 62(3), 223-246.
- Celentano, A., Fugini, M. G., & Pozzi, S. (1995). Knowledge-based document retrieval in office environment: The kabiria system. *Journal of ACM Transactions on Information Systems*, 13(3), 237-268.
- Chen, M. S., Han, J., & Yu, P. S. (1996). Data mining: An overview from a database perspective. *Journal of IEEE Transactions on Knowledge and Data Engineering*, 8(6), 866-883.
- Chen, K. J., & Liu, S. H. (1992). Word identification for Mandarin Chinese sentences. *COLING '92: Proceedings of the Fifteenth International Conference on Computational Linguistics* (pp. 101-107). Nantes: ACM Press.
- Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manages what they know*. Boston MA: Harvard Business School Press.
- Eppler, M. J., Seifried, P. M., & Ropnack, A. (1999). Improving knowledge intensive process through an enterprise knowledge medium. *SIGCPR '99: Proceedings of 1999 Computer Personnel Research* (pp. 222-230). New Orleans, LA: ACM Press.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurasamy, R. (1996). *Advances in knowledge discovery and data mining*. California, USA: AAAI Press.
- Fenstermacher, K. D. (2005). A process for delivering information just in time. *Journal of Lecture Notes in Computer Science*, 3782, 679-687.
- Han, J., & Kamber, M. (2006). *Data mining: Concepts and techniques*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- Hansen, P., & Järvelin, K. (2005). Collaborative information retrieval in an information-intensive domain. *Information Processing and Management*, 41(5), 1101-1119.
- Institute of Electrical and Electronics Engineers. (1998). IEEE standard for software project management plans (SPMP). (Technical report: IEEE Std 1058-1998).
- Kim, S., Suh, E., & Hwang, H. (2003). Building the knowledge map: An industry case study. *Journal of Knowledge Management*, 7(2), 34-45.
- Koller, D., & Sahami, M. (1997). Hierarchically classifying documents using very few words. *Proceedings of the Fourteenth International Conference on Machine*

- Learning* (pp.70-178). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
- KPMG Consulting (n.d.), *Insights from KPMG European knowledge management survey 2002/2003*. Amstelveen, The Netherlands: KPMG.
- Lai, L. F. (1998). *Task-based conceptual graphs as a basis for automated software engineering*. Unpublished master's thesis, National Central University, Taoyuan County. [Text in Chinese].
- Liu, D.-R., Wu, I.-C., & Yang, K.-S. (2005). Task-based K-support system: Disseminating and sharing task-relevant knowledge. *Expert Systems with Applications*, 29(2), 408-423.
- MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability* (vol. 1, pp. 281-297). Berkeley: University of California Press.
- Nonaka, I., & Takeuchi, H. (1995). *The knowledge creating company*. New York: Oxford University Press.
- Nemati, H. R., Steiger, D. M., & Iyer, L. S. (2002). Knowledge warehouse: An architectural integration of knowledge management, decision support, artificial intelligence and data warehousing. *Decision Support Systems*, 33(2), 143-161.
- Park, J. & Hunting, S. (2003). *XML topic maps: Creating and using topic maps for the web*. Boston MA: Addison-Wesley.
- PMI Standard (1996). *A guide to the project management body of knowledge*. Pennsylvania, U.S.: Project Management Institute.
- Rubenstein-Montano, B. (2000). A survey of knowledge-based information for urban planning: Moving towards knowledge management. *Computers, Environment and Urban Systems*, 24(2), 155-172.
- Rumbaugh, J., Blaha, M., Premerlani, W., Eddy, F., & Lorenzen, W. (1991). *Object-oriented modeling and design*. Englewood Cliffs, N.J.: Prentice-Hall.
- Salton, G., & Buckley, C. (1988). Term weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513-523.
- Siegel, A. (2000). Towards knowledge organization with topic maps. *Proceedings of the XML Europe 2000*, Paris, France.
- Sowa, J. F. & Zachman, J. A. (1992). Extending and formalizing the framework for information systems architecture. *IBM Systems Journal*, 31(3), 590-616.
- Staab, S., & Schnurr, H.-P. (2000). Smart task support through proactive access to organizational memory. *Knowledge-Based Systems*, 13(5), 251-260.
- Srikant, R., Vu, Q., & Agrawal, R. (1997). Min-

ing association rules with item constraints. *KDD '97: Proceedings of the International Conference on Knowledge Discovery and Data Mining* (pp. 67-73). California, USA: AAAI Press.

Sullivan, D. (2001). *Document warehousing and text mining* (pp. 326). NJ, USA: Wiley Computer Publishing.

Tah, J. H. M., & Carr, V. (2001). Towards a

framework for project risk management in the construction supply chain. *Advances in Engineering Software*, 32(10), 835-846.

Wu, I.-C., & Wang, Y.-C. (2007). An exploratory study of capturing and reusing project knowledge from a stage-based perspective. *AMCIS 2007: Proceedings of the Americas Conference on Information Systems*, Keystone, Colorado, USA.

(Received: 2010/2/28; Accepted: 2010/3/18)