國 立 交 通 大 學

資訊管理研究所

博 士 論 文

發掘情境導向之知識地圖以管理專案知識

Discovering Context-Oriented Knowledge Map for Managing Project Knowledge

研 究 生： 許 籌 尹

指導教授： 劉 敦 仁

中 華 民 國 九 十 五 年 一 月

國 立 交 通 大 學

資訊管理研究所

博 士 論 文

發掘情境導向之知識地圖以管理專案知識

Discovering Context-Oriented Knowledge Map for Managing Project Knowledge

研 究 生：許籌尹

研究指導委員會：許秉瑜　博士
呂瑞麟　博士
楊　千　博士
陳安斌　博士
指導教授：劉敦仁　博士

中 華 民 國 九 十 五 年 一 月

發掘情境導向之知識地圖以管理專案知識

Discovering Context-Oriented Knowledge Map for Managing Project Knowledge

研 究 生：許籌尹　　Student：Chouyin Hsu

指導教授：劉敦仁　　Advisor：Dr. Duen-Ren Liu

國 立 交 通 大 學

資 訊 管 理 研 究 所

博 士 論 文

A Dissertation

Submitted to Institute of Information Management

College of Management

National Chiao Tung University

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in

Information Management

January 2006

Taipei, Taiwan, Republic of China

中華民國九十五年一月

Discovering Context-Oriented Knowledge Map for Managing Project Knowledge

Student: Chouyin Hsu **Advisor: Dr. Duen-Ren Liu**

Institute of Information Management
National Chiao Tung University

Abstract

Forming projects to achieve different objectives and works is an essential work-type in most organizations. Moreover, many enterprises implement various business projects on the Internet for extending collaboration across different departments and organizations. Accomplishing projects essentially involves extensive resources and practical solutions which are valuable enterprise assets for supporting further project development. Therefore, systematically constructing project knowledge from historical projects is helpful for efficient knowledge support. Basically, a project worker is mostly engaged in various projects and seeks different references from project knowledge. Accordingly, interacting with users and responding the relevant part of project knowledge according to user information needs is an inevitable effort as integrating project knowledge.

Therefore, we propose the framework of project-based knowledge map for developing project knowledge and deliberately introduce project context for improving the communication and understanding in the framework. Based on Topic Maps, ISO/IEC 13250, the framework of project-based knowledge map is helpful for regulating project resources, interacting with users, extracting internal knowledge patterns, and constructing the project-based knowledge map for users. Particularly, project context improves the description of previous project experiences and the user interaction with current project developers for increasing the connections in the proposed framework. Multiple-phase data mining methods are therefore employed for knowledge discovery. Moreover, project context which provides the important operational information of project development is helpful for constraint-based data mining operation in the framework of project-based knowledge base. Consequently, the construction of project knowledge and the user-depend knowledge support are

fulfilled in the framework of project-based knowledge map.

A primitive system is developed for illustrating the significance of the framework of project-based knowledge map. Furthermore, RDF/XML technology is used for implementing the evolution of the project-based knowledge map in the system. The advantage facilitates the dissemination of project-based knowledge map across various applications over the Internet. Accordingly, the development and exploitation of project knowledge are accomplished in the framework of project-based knowledge map.

Keywords：Knowledge map, Data mining, Topic Maps, RDF/XML, Knowledge management

Acknowledgement

終於完成我的博士論文，心中激動萬分，難以言語。非常感謝劉敦仁教授最初的收留及持續的指導。劉老師嚴謹治學的態度，深深影響我對研究的堅持。從第一篇膽怯生疏的錯誤嘗試，一直到最後具有規模的論文撰寫，過程中的摸索，迷失，苦思，嘗試等階段，還好有劉老師適時的指引，才得以圓滿收斂完成。劉敦仁老師伉儷提攜的恩惠，永遠銘記在心。

在此，特別感謝口試委員許秉瑜教授，呂瑞麟教授，楊千教授及陳安斌教授熱心的指導，精采的討論過程加上延伸性的思考，頓時領悟更深一層。畢業前的這場灌頂大會，無疑是對博士身分最佳的洗禮，助益良多。我會更珍惜得來不易的學位，繼續加油。

為人師，為人妻，為人母，還有機會可以踏入交大校園，實現夢想，順利完成博士學位，是身邊許多人的功勞。一路走來，感激所有曾經幫助過我的貴人，所給予正面的能量；也謝謝眾多橫生的阻力，轉成逆向加持。感恩的心，祝福一路上緣份或深或淺的朋友，點點滴滴的交會，累積成就今日的我，有能力完成每個階段的任務。面對未來，期待與我的朋友與家人一起繼續努力前進！

當然，家中的小可愛瓅安及瓅帆是最貼心的小天使，總是玩累才肯去睡覺，留下我一個人面對電腦，看著你們酣睡的樣子，繼續工作。還好有老公永裕的支持勉勵，認真工作加上高度自我要求的建築師，是相互扶持二十年的夥伴。最後，特別將博士的榮耀獻給我的父母親，沒有你們，就沒有今日的我。感謝你們的祝福與體諒，我會更努力，扮演好自己的角色。

籌尹 (Kelly)
交大浩然圖書館 2006/1/16

Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Motivation

Knowledge management is crucial to the adaptation and survival of organizations in the face of continuous environmental changes [34]. The activities of knowledge acquisition, storage and distribution in a KM system enable the dynamic creation and maintenance of the enterprise intelligence [19][64]. According to KPMG research reports, about 80% companies look to KM to play an “extremely significant” or a “significant” role in improving competitive advantage, and consider knowledge as a strategic asset in business [27]. IBM took four years to reengineer their customer relationships by acquiring and disseminating knowledge to both customers and human experts [36]. Bolloju et al. proposed an integrative model for building enterprise decision support environments using model marts and model warehouses as knowledge repositories [7]. Schwartz et al. considers that memory-concept associations and e-mail systems should be used to manage organizational knowledge and deliver appropriate knowledge items in a timely and helpful manner [43]. In the last decade, knowledge management has covered a variety of disciplines and extends into many domains and applications.[53]

Surely, knowledge management also contributes many efforts to the field of project management for managing project knowledge. Rubenstein-Montano surveyed knowledge-based information systems for urban planning and suggested the importance of moving towards knowledge management [45]. Tah et al. applied knowledge management technology to identify project risk and further improve project management [52]. Barthès and Tacla developed an agent-supported portal to organize knowledge in complex R&D projects [4]. Deng et al. developed an integrated information system based on project-specific subjects [14]. Czuchry and Yasin offered a practical integrated informational approach to balance the strategic and operational concerns [12]. Most applications merely focus on the accumulation of projects, index or keyword search functions; however, they overlook the valuable knowledge patterns and working experiences hidden in historical projects. Therefore, extracting the internal knowledge patterns from the collection of previous projects is an

important contribution in this research.

Forming projects to achieve different objectives and works is an essential work-type in most organizations [40]. Internet technology has facilitated project processing across different departments, organizations and even countries. Many enterprises implement various business projects on the Internet. Network convenience significantly encourages processing projects with widespread project resources, such as project teams and systems. However, a project is a 'temporary' endeavor undertaken to create a particular product or service, and the project team is usually disbanded and reorganized for another new project [40]. Therefore, knowledge support is highly important in this kind of volatile relationship as developing projects in the distributed environment.

Many researches highlight the importance of projects in organizations [53]. Project accomplishment involves multiple project resources, including people, systems, methods and tools which contain valuable working solutions and experiences. The numerous historical projects are the important knowledge source. The advance of data mining techniques has inspired applications in different problem-solving domains [6][21][18] . Therefore, applying data mining methods to discover project knowledge from the collection of historical projects is essential in this research. The advantage is helpful to avoid committing the same errors and to reuse the practical experiences for efficiently facilitating further project development.

Project is an endeavor in which human, material and financial resources are organized in a novel way, to undertake a unique scope of work of given specification, within constraints of cost and time, so as to achieve unitary, beneficial change, through delivery of quantified and qualitative objectives [53]. Therefore, integrating project resources, processes and performances from the collection of historical projects are required to be organized in a flexible and consistent structure for developing project knowledge in this research [32]. For conceptual deployment of project knowledge and efficient knowledge support, knowledge maps which can visualize and explore complex abstractions is therefore applied for integrating historical projects, knowledge patterns and displaying project knowledge. Many works have proposed flexible structures of

tree or graph for developing knowledge map systems [11][26][29][31] . However, for considering the extension and dissemination of project knowledge on the Internet, Topic Maps, ISO/IEC 13250, is applied as the main structure of the knowledge map in this research. Topic Maps has been referred as the GPS (Global Positioning System) of the information universe. Therefore, Topic Maps which is a standard for representing interchangeable information is herein employed for governing the exchangeability and consistence of knowledge maps. Effective collaborations of project management over the Internet can thus be facilitated. The advantage not only facilitates the consistence and maintenance of the developed project knowledge, but also improves the meaningful representation and navigation of project knowledge.

Furthermore, as developing project knowledge, not only the attributes of project resources are important, but also the operational information is valuable context for conveying project experiences. Project attributes are useful for indicating what are involved in projects, such as tools and systems. Notably, project context is used for explaining the operational information of how project attributes are applied in projects, such as the location or duration of the system. Meanwhile, the project context is helpful for increasing the communication as developing the project knowledge. As interacting with users, the project context can provides the basic context conditions for user selection to determine the user information needs. Conversely, undiscriminating responses to all users in many keyword search systems easily impose many disadvantages on users.

- Users have to spend great amount of time to manually separate the relevant knowledge from others.
- Users probably miss the important knowledge or make mistakes by manually separation.
- The internal relationships among the result are not easily discovered by users if without further assistance.

A project is a temporary endeavor undertaken to create a unique product or service [40]. Temporary means that every project has a definite beginning and a definite ending. Inevitably, a project worker is engaged in different situation among different projects. Therefore, interacting with users is important for

increasing the understanding between the system and users. Since project context provides the operational information and circumstance of project development, such as the location of systems or the role of workers, users can select the similar context to locate knowledge, instead of entering keywords. Different from applying prior knowledge or experiences to enter keywords, the selection of context conditions is useful communication with users. From context perspective to manage project knowledge is practical and valuable challenge.

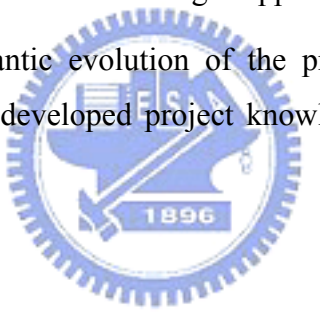
Therefore, we propose the framework of project-based knowledge map in Topic Maps for consistently describing context-rich project attributes, integrating project resources, discovering internal knowledge patterns and displaying meaningful guideline for user selection to locate the relevant part of knowledge map. Particularly, data mining methods are employed for extracting hidden knowledge patterns. The proposed project context emphasizes the import development annotation from previous project developers and also the useful interaction for further knowledge seekers. The implicit communication facilitates the intelligent service of the framework for responding the relevant project knowledge to support further project development.

Basically, three major stages are herein provided in this research for illustrating the development and evolution of the project-based knowledge map. The first stage explains the development and the advantage of the framework of the project-based knowledge map. The second stage implements a web-based system for demonstrating the performance, and the third stage applies RDF/XML technology for efficiently disseminating the project-based knowledge amp across various applications over the Internet. Therefore, the knowledge navigation and knowledge exploitation over the network are efficiently fulfilled in the framework of the project-based knowledge map with context information.

1.2. Goals

Major goals of this dissertation are listed below.

- Developing the framework of project-based knowledge map in Topic Maps for persistently constructing project knowledge in the project-based knowledge map.
- Appling various data mining methods for extracting internal association patterns from historical projects to enrich the content of project-based knowledge map.
- Formally providing project context for annotating the essential operational information of project development to improve the performance of the project-based knowledge map.
- Providing context-oriented selection menu for interacting with users to facilitate user-dependent knowledge support in the proposed framework.
- Successively semantic evolution of the project-based knowledge map for disseminating the developed project knowledge across various applications on the Internet.



1.3. Contributions

Knowledge management involves a thorough, systematic approach to information repository of an organization by a sequence of collaborative processes. Many knowledge management applications for developing project knowledge mostly focus on the accumulation of projects, index or keyword search functions; however, the valuable knowledge patterns and working experiences hidden in historical projects are overlooked. Therefore, extracting the internal knowledge patterns from the collection of historical projects is an important contribution of the framework of the project-based knowledge map.

Moreover, employing Topic Maps as the foundation of knowledge maps for constructing project knowledge is a leading contribution of this research. Topic Maps, ISO/IEC 13250, a standard for representing interchangeable information, is herein employed for governing the exchangeability and consistence in the

proposed framework. Beyond the accessibility and extensibility of trees or graphs used in many traditional knowledge map applications, Topic Maps which has been referred as the Global Positioning System of the information universe is the basic model for constructing and representing the project-based knowledge map in a meaningful hierarchy. Various data mining methods work collaboratively for extracting association patterns to promote the fulfillment of Topic Maps in this work. The advantage improves the knowledge representation and user navigation on the project-based knowledge map for efficient knowledge support. Particularly, coupling Topic Maps and data mining methods in the proposed framework is a leading research and an important contribution.

Project context which is deliberatively provided for improving the communication and understanding as developing the project-based knowledge map is a novel contribution. Project knowledge is valuable intelligent asset in organizations. In order to support further project development efficiently, not only the outcome of projects is important for developing project knowledge, but also the operational information and circumstances of project development is an essential part of project knowledge. Therefore, project context is especially proposed in the framework for annotating operational experiences, such as the circumferential information of who, whom, how, when, and where is of project development. The information is helpful for providing a guideline for user interaction. As a result, project context offers various perspectives for constraint-based data mining to facilitate user-dependent knowledge support, instead of overwhelming users with indiscriminative outcome as using keyword search tools which are used in many traditional knowledge management applications. Therefore, with the benefit of project context, the previous project experiences are conveyed to current project workers in the proposed framework for improving the connections as developing the project-based knowledge map. Consequently, project context proposed is an important contribution in the framework of project-based knowledge map for determining the right part of project knowledge to right person for the right purpose.

Notably, the evolution of the project-based knowledge map developed in RDF/XML technology is a further contribution in this research. Although many

database systems and programming languages can provide the corresponding referential integrity and syntax to implement the project-based knowledge map, the flexibility of XML and the semantics of RDF indeed facilitate the exploitation and dissemination of the project-based knowledge map across different applications on the Internet. As the web service becomes the major highlight on the Internet in the future, the project-based knowledge map is even competent to extensively construct and disseminate project knowledge.

1.4. Organization of this dissertation

The rest of this dissertation is organized as follows. Section 2 reviews related work on technology background and related work. Chapter 3 then explains the advantage of context-oriented knowledge maps and emphasizes the definition and benefit of the project context. Chapter 4 presents the framework of project-based knowledge map and the meaningful structure of the project-based knowledge map. Plus, the important modules and essential operations are further specified. Next, Chapter 5 gives an illustrative system for developing the project-based knowledge map and explaining the performance. The evolution and progression of the project-based knowledge in RDF/XML are discussed in Chapter 6. Then, conclusions and future works are finally made in Chapter 7.

Chapter 2. Related Work and Technology Background

2.1. Related work

2.1.1.Knowledge maps for knowledge management

Knowledge management covers a variety of disciplines and extends into many general domains, such as general management, information systems, decision support systems, artificial intelligence, expert systems and more. On the technical viewpoint, knowledge management tools are the major considerations for implementing the effective knowledge management system. Beyond the traditional groupware and information systems, knowledge tools primarily can manipulate information, automate information search, retrieval agents, decision support, document management, and enable knowledge generation, codification transfer. Widely used keyword search engines and menu-driven interface tools have been employed for practical systems for years.

In the last decade, knowledge management has become the important issue for academic research and industrial practice. As shown in Table 1, the various definitions of knowledge management provide different viewpoints of managing knowledge.

Table 1. The various definitions of knowledge management

Author	Summary	Year
K. Wiig	Creation, Compilation and Transformation	1993
J. Han, Y. Fu, S. Tang	Dissemination, Application and Value Realization	1995
Nonaka, Takeuchi	To integrate machine learning methodologies with database technologies and discovers different kinds of knowledge from large database	1995
Petrash	KM is getting the right knowledge to the right people at the right time so they can make the best decision.	1996
M. Marquardt	Knowledge creation, conversion, Organizational knowledge-creation process and knowledge spiral	1996
T. Beckman	KM is the formalization of and access to experience, knowledge, and expertise that create new capabilities, enable superior performance, encourage innovation, and enhance customer value.	1997
R. Ruggles	Acquisition, creation, storage, transfer, and utilization of knowledge	1997

EC. Nevis, AJ. DiBella, JM. Gould	Building, Maintaining, and Sustaining the Learning Organization	1998
TH. Daevenport D. W. De Long, M. C. Beers	Knowledge management is the process of capturing, distributing, and effectively using knowledge.	1998
E. Zack	Embedding knowledge in processes, representing knowledge, facilitating knowledge, transferring , measuring the value of knowledge assets	1999
D. Sinclair, N. Hardie	Identify, Create knowledge maps, Find the knowledge gaps	2000
A. Rabarijaona, R. Dieng, O. Corby, R. Quaddari	Facilitate , Creation, Optimize knowledge use	2000
Fischer, Ostwald	Apply the XML meta language for corporate knowledge management for ontology query	2001
The Diffuse Org.	KM is collaboratively designed and constructed in the promise of social creativity, living organizational memory, and Attention economy to deliver the relevant knowledge for workers.	2001
Ernst & Young	knowledge generation, knowledge representation, knowledge codification, knowledge application	2001
M. Cannataro, A. Guzzo, A. Pugliese	The business standard of knowledge Capture, Classification, Dissemination	2002
M. Cannataro, D. Talia,	The reuse of semi-structured information and its integration in heterogeneous systems, and the creation of knowledge maps for the organization and sharing of information.	2003
E. Woods	Designing, building, and implementing an architecture for distributed knowledge discovery as Grid application	2003

However, upwards to the conceptual deployment for efficient knowledge support, most tools do not the meet the challenges [59]. Therefore, knowledge maps which can visualize and explore complex abstractions become the increasingly important and popular knowledge management tool. The definition proposed by Vail illustrates the advantage of visual display for efficient communication with users [54].

"*A knowledge map is a visual display of captured information and relationships, which enables the efficient communication and learning of knowledge by observers with differing backgrounds at multiple levels of detail*."

Successive applications take advantage of the nature of visualization and navigation of knowledge map for locating and publishing knowledge [13][26]. Vail highlighted the visual display of captured information [54]; Browne et al. combined knowledge maps with reasoning-based methodology to elicit probability assessment for decision making [8]; Leathrum et al. employed knowledge maps to design an intelligent questioning system for learning [29]; Chung et al. proposed a knowledge map framework for discovering business intelligence to alleviate information overload on the Web [11]; Lin and Hsueh applied information retrieval algorithms to generate and maintain a knowledge map system for virtual communities of practice [31]; Kim et al. proposed a road map to develop knowledge maps in the industrial community [26]. The performance of knowledge maps has been receiving considerable attention in the many applications of knowledge management practice, excluding project management. Notably, we first apply knowledge maps for developing project knowledge extensively for support project development [32].

2.1.2.Context information

Context can be a list of situational factors or any information that characterizes the situation of an object, where an object is a person, place, or object considered relevant to the interaction between a user and an application, including the user and application themselves [17]. Therefore, with the help of context, numerous applications become user-friendly, flexible, and adaptable. For example, context information in web search is considered for improving the personalized knowledge search [28]. In mobile spontaneous network, context is useful for managing knowledge exchange [49]. Context information is applied to integrate independent tools into a pervasive computing system with that monitors user interactions with the computer and data sources [55]. Document standardization is proposed by merging the documents and work context to improve the effectiveness in document management [47].

In this work, we deliberatively employ context information for improving the communication and understanding in the proposed framework. Accordingly, context information promotes knowledge discovery and representation of the

project-based knowledge map for efficient knowledge support.

2.2. Technology background

2.2.1.ISO/IEC 13250 Topic Map

Topic Maps, the ISO standard ISO/IEC 13250, defines a model for the semantic structuring of link networks in 2000 [20]. With growing information repository on Internet, Topic Maps, the need for enhanced access mechanisms for linking, navigating and exploring information from site to site is getting more and more evident. Basically, topics, associations, occurrences are major components in Topic Maps.

- A topic is concrete description to represent a real world subject, such as a person, a theme, or a concept.
- An occurrence is one of the real information objects link to related topics, such as a file, a video, or a report. A topic is an abstract label, and occurrences are substantial references.
- Associations are used to describe the relationships between topics, and an association name is a meaningful concept for defining a relationship.

2.2.2. XML/RDF specifications

Extensible Markup Language (XML), derived from SGML, is a core technology that defines a universal standard for structuring data. XML version 1.0 was defined in 1998 by the World Wide Web Consortium [64] and the Second Edition was published in 2000. XML is a global standard for storing structured data in an editable file that is useful for data storage, data exchange and document publishing on the Internet. XML exclusively focuses on manipulating data, and Extensible Stylesheet Language (XSL) is useful for interchangeably displaying data in different templates for HTML, PDF, WAP view types and many Internet-based applications.

A custom DTD is used to constrain the syntactical rules that govern the

proposed knowledge maps, and describe their major elements, including topic names, occurrences and association relationships, in XML documents. This work emphasizes simple definitions of XML/DTD specifications to increase the practicability of the project-based knowledge map.

RDF (Resource Description Framework) provides the foundation for metadata interoperability across different applications via a syntax specification and schema specification. Moreover, concrete RDF syntax uses the XML by which RDF can specify semantics for data in a standardized and interoperable manner [10]. RDF is based on the idea of identifying things using web identifiers (called Uniform Resource Identifiers, or URIs), and describing resources in terms of simple properties and property values.

RDF model intrinsically supports binary relations (a statement specifies a relation between two resources). Higher arity relations have to be represented using binary relations [42]. Basically, RDF model consists of three major components, and the explanation is given below.

- Resources: All things being described by RDF expressions are called resources. A resource may be an entire Web page, a part of a Web page, an entire web site, or an object that is not directly accessible via the web page (e.g., a printed book).
- Properties: A property is a specific aspect, characteristic, attribute, or relation used to describe a resource. Each property has a specific meaning, defines its permitted values, the types of resources it can describe, and its relationship with other properties.

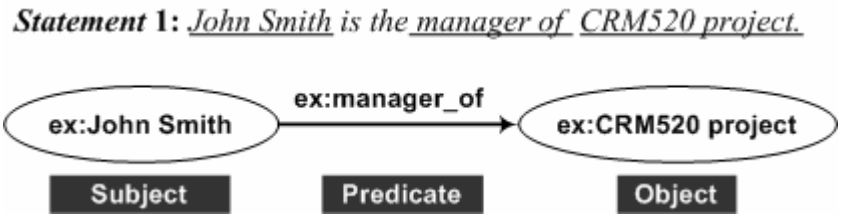


Figure 1. A simple RDF statement with RDF graph

- Statements: A specific resource together with a named property plus the value of that property for that resource is an RDF statement. The three

individual parts of a statement are called the subject, predicate and object, of the statement respectively. For example, a statement example and RDF graph are given below, as well as the corresponding RDF graph.

Also, RDF requires the XML namespace facility to uniquely and precisely identifying the governing authority of the vocabulary. XML namespaces provide a method for unambiguously identifying the semantics and conventions governing the particular use of property-types.

2.2.3. Data mining methods

Data mining is powerful for extracting patterns of business interests, including associations, changes, anomalies and significant structures from information repositories, in efficient and productive ways. Data mining involves several tasks for different mining purposes, including association rule mining, clustering, classification, prediction, and time-series analysis [18][6][21]. We mainly employs clustering mining, generalization mining, and constraint-based association rule mining to extract different patterns based on context perspectives.

Clustering mining aims at identifying groups of similar objects without known or specified classes and therefore helps to discover distribution of patterns and interesting correlations in large data set. Clustering has different names in other research fields, such as unsupervised learning (in pattern recognition), numerical taxonomy (in biology, ecology), typology (in social sciences) and partition (in graph theory). Clustering can be considered the most important unsupervised learning problem. A loose definition of clustering could be “the process of organizing objects into groups whose member s are similar in some way”. A cluster is therefore a collection of objects which are “similar” between them and are “dissimilar” to the objects belonging to other clusters [24][3].

Many clustering methods, including hierarchical clustering, partitional clustering, density-based clustering, grid-based clustering and model-based clustering, are applied in various applications [46][44][21]. Basically, Agglomerative Hierarchical Clustering (AHC) that is more commonly used for

document clustering carries the following advantages [56][61]

- One can either prescribe the number of clusters or let the number of clusters determining by demanding a certain minimum similarity within a cluster.
- The number of clusters can be rapidly adjusted by given minimum similarity, instead of rerunning clustering algorithm again.

Bottom-up hierarchical clustering continues until some stopping condition is reached. Typical stopping conditions include a threshold for the number of remaining clusters (e.g., stop if 4 clusters remain), or a threshold for the required similarity between clusters (e.g., stop if the two most similar clusters are not very similar). The procedure is given as follows [3][24].

1. Agglomerative clustering starts with the set of objects as individual clusters.
2. At each step merges the most two similar clusters. The process is repeated until a minimal number of clusters have been reached, or, if a complete hierarchy is required then the process continues until only one cluster is left.

An important component of a clustering algorithm is the similarity measure between pairs of clusters. Many distance measures are herein employed for different applications, including Euclidean, Dice, Jaccard, Cosine, Manhattan [3]. If the components of the data instance vectors are all in the same physical units then it is possible that the simple Euclidean distance metric is sufficient to successfully group similar data instances. Note that it weights all features/dimensions "equally". Pair-wise vectors between *i*th and *k*th vectors, denoted by *d(i, k)* as follows, where *t* is the number of attributes.

$$d(i,k) = \left[\sum_{j=1}^{t} (w_{ij} - w_{kj})^2 \right]^{1/2} \quad \textbf{(2.1)}$$

Agglomerative clustering algorithms are mainly classified as single-link, complete-link and average-link depending on the method they define inter-cluster similarity. Single-link and average-link methods typically take $O(n^2)$ time, while

complete-link methods typically take $O(n^3)$ time. Moreover, many research works hold various experiments to explain that average-link performs better than others in document/text clustering experiments [56][61]. The details of three methods are explained as follows.

Single Linkage Method: The similarity between two clusters S and T is calculated based on the minimal distance between the elements belonging to the corresponding clusters. This method is also called "nearest neighbor" clustering.

$$\|T - S\| = \min_{x \in T, y \in S} \|x - y\| \quad \textbf{(2.2)}$$

Complete Linkage Method: The similarity between two clusters S and T is calculated based on the maximal distance between the elements belonging to the corresponding clusters. This method is also called "furthest neighbor" clustering.

$$\|T - S\| = \max_{x \in T, y \in S} \|x - y\| \quad \textbf{(2.3)}$$

Average Linkage Method: The similarity between two clusters S and T is calculated based on the average distance between the elements belonging to the corresponding clusters. This method takes into account all possible pairs of distances between the objects in the clusters, and is considered more reliable and robust to outliers. This method is also known as UPGMA (Unweighted Pair-Group Method using Arithmetic averages).

$$\|T - S\| = \frac{\sum_{x \in T, y \in S} \|x - y\|}{|S| \cdot |T|} \quad \textbf{(2.4)}$$

Association rule mining finds interesting associations or correlations among large sets of data, and creates practical rules that describe how frequently events or objects have occurred together. It has been applied in numerous applications and successfully applied in market basket analysis, recommender systems, user behavior analysis, and other areas [3]. Apriori algorithm, one of useful association rule mining methods, counts the support of individual topic name and an itemset is considered frequent if the support of that itemset exceeds a user-specified mimum support [1].

Confidence and support, two major parameters in the formal statement of association rule, are shown below.

(1) I is a set of items. D is the transaction database. An association rule is an implication of the form $X \Rightarrow Y$, here $X \subset I$, $Y \subset I$ and $X \cap Y = \varnothing$.

(2) Rule $X \Rightarrow Y$ has a confidence c in D if c% of transactions in D that contains X also contains Y. Also, rule $X \Rightarrow Y$ has a support s in D if s% of transactions in D contains $X \cup Y$.

Generalization mining is a process that abstracts a large set of task-relevant data from a relatively low conceptual level to higher conceptual levels. Generalization mining, one of descriptive data mining methods, is helpful to describe the data set in a concise manner and present in interesting properties or attributes. Attribute removal and attribute generalization are two efficient methods for generalization [21].

- Attribute removal is used to eliminate a constraint and thus generalize rules in a large set of distinct values for an attribute.
- Attribute generalization is to apply a generalization operator on the attribute, in which there is a large set of distinct values.

Constraint-based association mining is performed under the guidance of various kinds of constraints, including knowledge type, data, dimension, interestingness and rules [21]. The advantage is helpful for filter out the uninteresting data and result.

- Data constraints specify the set of data to be mined. Data constraints and knowledge type constraints are applied before mining. The advantage is helpful for reducing the complexity of data mining operation and simply focusing on the interesting data.
- Rule constraints specify the form or property of rules to be mined, including rule form constraint and rule content constraint [37]. Rule constraints, dimension constraints, interestingness constraints are applied after mining operation. The advantage is helpful for focusing on the interesting result for further interpretation and manipulation.

Chapter 3. Context-oriented Knowledge Maps

A project is performed by a project team, which performs a group of processes within a particular period, including project initiation, planning, executing, controlling and closing processes [40]. Project resources are therefore generated during these processes. As shown in Figure 2, various project resources, including people, systems, methods and tools which contain valuable working solutions and experiences, are valuable references for developing project knowledge to support further projects.

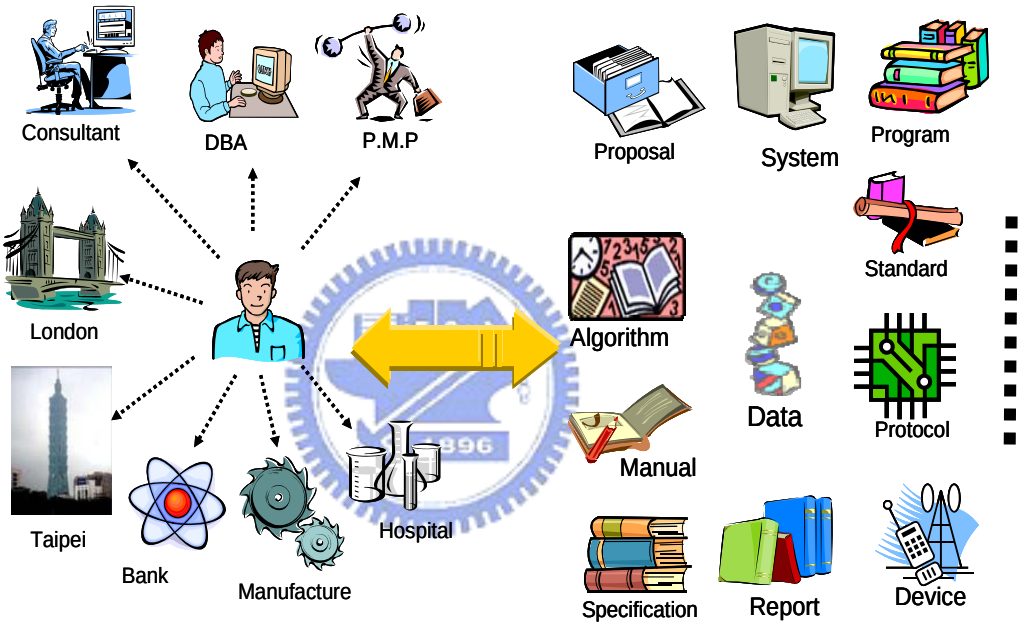


Figure 2. Context information in the development of project knowledge

Moreover, a project involves team members from various departments, organizations or even virtual organizations to collaborate on the Internet. A global network diminishes the concern of geographic boundaries, but increases the diversity of project development environment. A project worker who is usually engaged in different situations among different projects since a project is a temporary endeavor undertaken to create a unique product or service [40]. Therefore, as managing project knowledge from diverse project resources, annotating the circumstances and background of project operations is valuable guideline for interacting with users. Not only improving the representation of

project knowledge, but also the useful guideline of the user interaction in the framework of project-based knowledge map.

3.1. The project-based knowledge map in Topic Maps

With the inherent advantage of knowledge map, the proposed framework of project-based knowledge map integrates and displays project knowledge systemically and meaningfully. Upon the popularity of Internet, the modern idea behind a knowledge map is required to construct a global architecture to show the content of knowledge. Therefore, we apply Topic Maps, ISO/IEC 13250, as the major model in the proposed framework for systemizing project knowledge, including project resources, important concepts and up way to association patterns. Consequently, many advantages improve the performance of project-based knowledge map as follows.

- The instinctive hierarchy of Topic Maps is the main structure of project-based knowledge map. Therefore, the expressive structure is helpful for conducting the consistent project knowledge.
- The strong linking capability in Topic Maps supports the internal connection of project-based knowledge map, including physical and abstract links. Physical links are useful for connecting to distributed project resources, and abstract links are helpful for indicating associations.
- The notation in Topic Maps is applied for declaring the important definitions and reducing the conflicts and ambiguity in the proposed framework.
- The classification system in Topic Maps for managing topics and associations is helpful for improving the expression of the rule statements in the project-based knowledge map.

3.2. Context information in the project-based knowledge map

As shown in Figure 3, when we humans interact with other person, the previous contact experience and current situation information, such as location or job title, are helpful context that intuitively affects human reactions [16]. Conversely, computers can not spontaneously exploit such context information when interacting with users. Thus, employing context in the project-based knowledge map is practical for improving intelligent communication as discovering project knowledge.

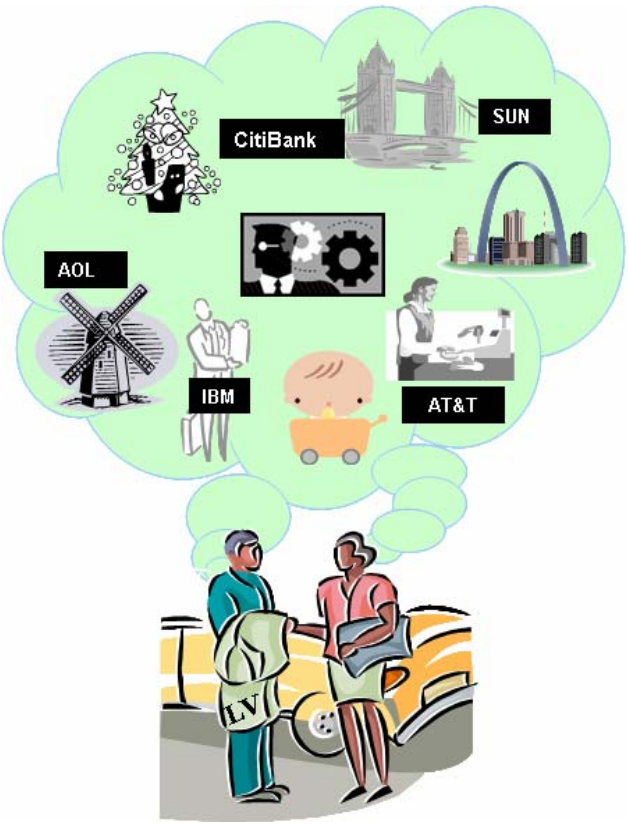


Figure 3. Context information in human communication

Therefore, we employ context information in the framework of project-based knowledge map for improving the representation of project knowledge and the communication with users. With the helpful of context, both of the expression of project-based knowledge map and the guideline of user interaction become meaningful and practical to users in this work. Notably, the novel synergy of context and knowledge maps increasingly improves the knowledge exploitation and user navigation for efficient knowledge support in the framework of

project-based knowledge map.

A project team is usually disbanded and reorganized before another new project, meaning that such cooperation is temporary. Plus, the convenient network increases the diversity of cooperation across geographical boundaries. This kind of volatile and complicated relationship increases the difficulty of communication as developing project knowledge. Therefore, providing sufficient project description from project workers to indicate the essential operational information and circumstance of project development is helpful for improving the communication and understanding as managing project knowledge.

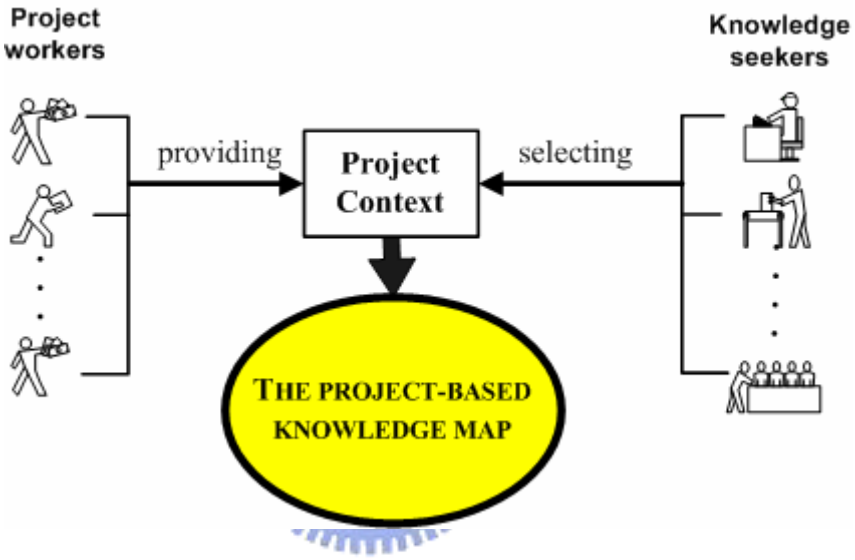


Figure 4. Context-oriented development for the knowledge map

As shown in Figure 4, along with the system itself, project participants contribute supplementary operational information of project development to build project context. On the other hand, various situations in project context provide fundamental guideline for user selection menu. The advantage is helpful for perceiving certain criteria from the result of user selection. Accordingly, project context which annotates the circumstance of project operations and the background of project workers is useful criteria as determining the proper part of project-based knowledge which is relevant to user information needs.

Basically, the term of 'project context' is not formally defined in the domain of project management or knowledge management. A non-official definition provided by Wideman is that project context is the background or environment in

which the project is conducted, also the background that justifies the project in the first place [62]. For example, the project context of the Cadastre Reengineering Project provided the general project context as follow [9].

The program was launched with the passage of the Law for Quebec Cadastre Reform in 1985. In 1989, the program bases were changed and the involvement of DMR increased with the allocation of a second contract to the company, which would cover organizational aspects of the project. In January 1993, the Ministry of Natural Resources (MNR) restarted the cadastre reengineering program now reevaluated at C$ 500 million and spanning a period of 13 years, from 1993 to 2006.

Different form above definition of project context for illustrating the inceptive background and environment of a project abstractly, we propose a novel and modern definition for project context from the consideration of inevitable network and complicated project types. In the framework of project-based knowledge map, work, project context is used for annotating project attributes, instead of a whole project. Since the convenient network facilitates projects processing in a distributed and heterogeneous environment, various situations and circumstances are essential annotation along with project attributes for increasing understanding and communication across geographic boundaries. Moreover, the description about project developers is also an essential part in project context. The situation, background, and viewpoint of project developers are useful for comprehending how the project attributes were applied and activated in previous projects. Therefore, the previous project experiences are naturally conveyed in project context for supporting current project workers in this work.

3.2.1. The definition of project context

Project context is persistently defined as follows for effectively discovering context-oriented knowledge maps.

Project context annotates the operational information and the viewpoint of project developers for each project attribute as describing the important features of project objects.

Definition 1(project context and project attribute): Given a project object *PO*, and project attribute *PA* is one of its project attributes which describe the important features of project object *PO*. Then, project context *PC* is a set of annotations which explain the details of how the project attribute *PA* is dedicated in the project object *PO*. In order to completely supplement the operational annotations, the existential dependency ⟨*PA*, *PC*⟩ is a one-to-many relation, as well as the dependency ⟨*PO*, *PA*⟩.

Definition 2 (context type): Given a set of project context *PC* for annotating a project attribute *PA* which describes a project object. Then context type is a basic set of classification for grouping *PC* into the sense of know-who, know-where, know-for, know-when, know-how and know-for to indicate the rational relationship between *PA* and *PC*.

Then, an example is given below along with different situations for explaining the advantage and importance of the project context for elucidating project attributes.

Example 1: For the tool of OLAP which is one of the project attributes used for describing the specification file spec_dw.pdf., project context is a set of annotations which provide more operational information to elucidating how the attribute OLAP is dedicated in spec_dw.df. As shown in Table 2, a set of project context (DBA, Taipei, Bank, 1996-1998, Cube, Sales department) clearly provides users the operational information of project attribute OLAP in the specification file spec_dw.pdf, including the information of role, location, duration, operation, client, and organization type.

The advantage enriches the sense of know-who, know-where, know-for, know-when, know-how, and know-whom about the project attribute OLAP used in the specification file. As a result, not only the project attribute provides the important feature of the project object, but also the project context annotates the essential operational information of the project attribute.

Table 2. An example of project context

Project object	Project attribute	Project context	Context type	Rational sense
spec_dw.pdf A specification document of the project DW007	**OLAP**	***DBA***	role	know-who
		Taipei	location	know-where
		Bank	organization	know-for
		1996-1998	duration	know-when
		Cube	operation	know-how
		Sales department	client	know-whom

Basically, there are four types of project workers who seek the support of project knowledge, including novice, junior, senior and expert workers. Four various user experiences are described below, and the corresponding situations are explained right after.

- A novice worker usually has little or none practical working experience. Therefore, a novice worker has to learn everything about project development from beginning level. As involving a project, a novice worker who know less and easily makes a mistake will need complete knowledge support to make progress.
- A junior worker has some or partial project experiences. Thus, a junior workers need to review some previous experiences and learn new knowledge. As involving a project, a junior worker who probably causes some delay will need sufficient knowledge support to improve the efficiency.
- A senior worker has much project experiences. Then, a senior user can apply the previous knowledge and understand new knowledge quickly. However, a senior worker who is usually burdened with heavy loading will need relevant project knowledge to promote the performance.
- An expert worker has abundant project experiences in development and

management. Therefore, an expert worker who usually is a project manager or a project leader will need pertinent knowledge support to facilitate project development entirely, such as key internal and external stakeholders and important contracts.

Situation 1-1: From the Table 2, a novice worker can easily learn the practical operational information about OLAP from the set of project context. Before referring the project object, the user can learn some operational experiences about the project attribute OLAP, including that a DBA may utilize OLAP in some bank-type projects, and the operation of cube is useful for analyzing the data for sales department. Also, the time period is helpful for reminding users the trend and popularity of OLAP. Accordingly, a novice worker can locate these important parts as referring the specification file.

Situation 1-2: From the Table 2, a junior worker can reinforce or supplement operational instruction about OLAP from the set of project context. For example, the worker applied some operations of OLAP, rather than cube operation. Then, the worker can right focus on the description on cube operation to shorten the total learning cost as referring the specification file.

Situation 1-3: From the Table 2, a senior worker can decide to refer the specification file or not since the project context reveals important information. For example, the worker can overlook the specification file if the worker is familiar with the cube operation in OLAP and bank-type projects. Then, the worker can look for other support without delay.

Situation 1-4: From the Table 2, an expert worker can learn some managerial information from the project context, such as a DBA may apply OLAP in some bank-type projects and the cube operation is useful in sales data. For example, the expert worker who is a project manager will select a person who knows OLAP to be a DBA. Also, the reasonable project schedule is possibly arranged according to the useful duration in project context.

The sensible project context provides meaningful information for different types of project workers. Conversely, another opposite example is given below. In contrast with above situations, four types of project workers may encounter more difficulties and problems whereas the case in example 2 is given.

Example 2: OLAP is a project attribute for describing the important feature of the specification file spec_dw.pdf. None of annotation or comment is offered for the project attribute OLAP.

Situation 2-1: A novice worker has no choice but referring the file. Lack of further annotation to indicate how OLAP is applied in the specification file, a novice worker has to go through the file or ask for help to figure out the application of OLAP. Many inconveniences are possible resulted from missing or misunderstanding the important parts. As a result, making progress is not easy under this situation.

Situation 2-2: A junior worker still has to manually open the file to learn the relevant parts for supporting current projects. However, the learning effect is totally depends on the worker. If the user missed or misunderstood something important, no one could help. Lack of further annotation, a junior worker may slowly look for the application of OLAP. The disadvantage imposes the worker laborious learning cost to understand the operational information of OLAP in the specification file. As a result, the working efficiency is not easy improved under this situation.

Situation 2-3: An expert user may use OLAP before, but he or she still has to refer the file again to see if something is useful. Simply providing the project attribute of OLAP to an expert user is insufficient to make the right decision. As a result, project development is not easy promoted under this situation.

Situation 2-4: An expert user or a project manager who may use OLAP before still has to manually open the file to find the operational information about OLAP to confirm or refresh his or her experiences, such as the user role or the period as applying OLAP. As a result, the project assignment or reasonable schedule is not easily controlled under this situation.

3.2.2.Project context in the user interaction

Interacting with users is important in modern knowledge management system. The advantage facilitates user-dependent knowledge support to meet with user situations. In this research, project context can provide a useful guideline for user interaction. As involving in a project, a project worker is basically assigned a particular position and working places for carrying out certain assignments in a period of time. Therefore, the situations of previous project developers which are mainly described in project context are useful guideline for current users to find previous project experiences.

Itemizing some important context types used for expressing the situations of project developers in project context for users to select the interesting context conditions is the useful user interaction in this research. As a user selecting the context conditions from the user selection, user interests are explicitly expressed to the system. According the selection result, the proposed framework can determine the relevant parts of project-based knowledge map for users, rather than passively displaying the overall project-based knowledge map. Instead of entering a series of keywords or laboriously tracing user behaviors, the user selection is convenient and adaptable to change user criteria for supporting different projects. Accordingly, the efficient user-dependent knowledge support is fulfilled in the framework of the project-based knowledge map.

A usually case is given in the following example. A project worker involved in different projects looks for relevant project knowledge to support different projects. Various situations in different knowledge management systems are discussed for explaining the advantage of the project context in the user interaction.

Example 3: Mary is a project worker, and she is assigned in A1 and A10 projects as DBA and consultant, separately. Mary needs different parts of project knowledge to carry out A1 and A10 projects properly

Situation 3-1 (keyword search in KM system): As Mary uses a keyword search KM system for locating relevant project knowledge, she has to learn some

prior knowledge to know what keywords are meaningful for project A1 and A10. However, a project may need different set of keywords to find out useful outcome, since different keywords will result in different searching outcome. Therefore, entering numerous keywords for different projects will easily cause heavy burden on users. As a result, Mary may enter more than 10 keywords for one project to search the relevant knowledge, and then manually separate the useful information from the overwhelming outcome. At the worst, if Mary is not familiar with the project, she could not get any relevant outcome without entering proper keywords.

Situation 3-2 (user profile in KM system): As Mary uses a knowledge management system which offers user profiles for locating relevant project knowledge, she gets the support dependent on the operations of previous system users instead of the original previous project workers. User profiles usually collect user browsing paths in long-term monitoring archives for further analysis. However, dynamic situations in the distributed project environment easily result in unstable and capricious user profiles. The outcome of the further analysis is not always promising to locate the relevant parts of knowledge for users. At the worst, Mary will get none or irrelevant parts because the insufficient archives or unsuitable analysis.

Table 3. Context conditions in the user selection

Project	Name	Context type	User condition	Rational Sense
A1	Mary	*role*	*DBA*	show-who
		location	*Taipei*	show-where
		organization	*Government*	show-for
A10	Mary	*role*	*Consultant*	show-who
		location	*NY*	show-where
		organization	*Bank*	show-for

Situation 3-3 (context in the project-based knowledge map): As Mary uses the project-based knowledge map system with context information, she can select the interesting context conditions from the user selection to locate the relevant parts of project knowledge. For example, the system provides context types of role, location and organization in user selection. As shown in Table 3, Mary expresses her conditions in project A1 and project A10 separately, including Mary works as a DBA for the government-type project A1 in Taipei, and she works as a consultant for the bank-type project A10 in NY (New York). The sense of show-who, show-where and show-for is used to compare with the sense of know-who, know-where and know-for in project context for further knowledge discovery operation to determine the relevant parts of project-based knowledge for Mary. Accordingly, Mary can receive the relevant knowledge support to improve the performance in project A1 and A10.

Chapter 4. The Framework of Project-Based Knowledge Map

We propose the framework of project-based knowledge map for developing project knowledge from historical projects. The proposed framework provides project attribute builder, context information services, knowledge discovery module, map generator module and for linking project resources, interacting with users, extracting association patterns, and displaying the proper part of knowledge map for users. Particularly, the project knowledge base collects and maintains the consistent definitions and important association patterns in the framework.

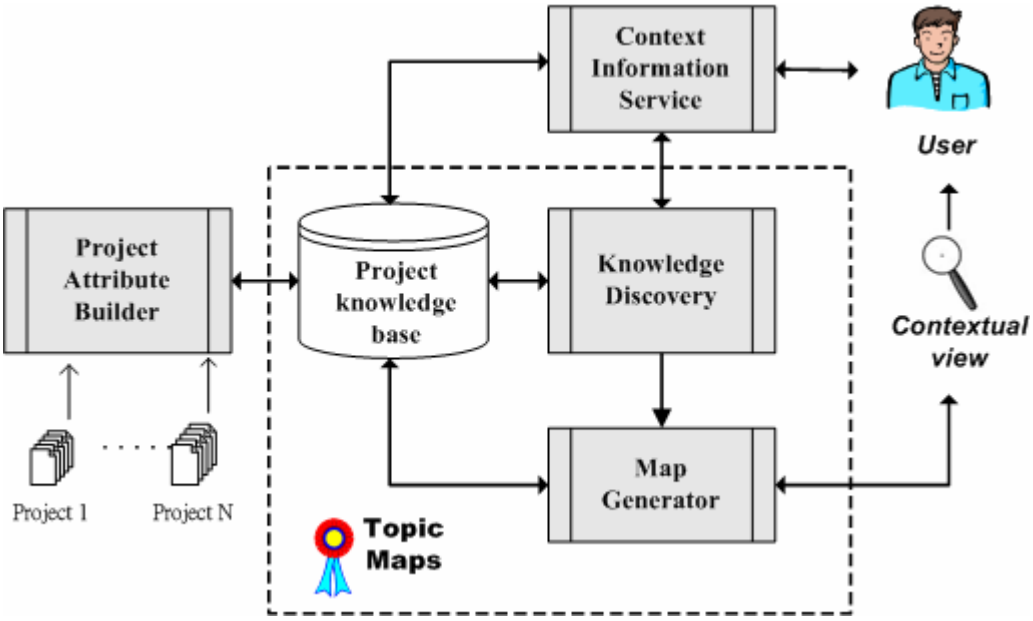


Figure 5. The framework of project-based knowledge map

Notably, Topic Maps is the major standard for controlling the consistence and agreement for the construction and representation of knowledge in this framework. As shown in Figure 5, project knowledge base, knowledge discovery and map generator in the dot-line area persistently conform to the notation and regulation in Topic Maps for archiving, extracting and representing project knowledge. Besides, project attribute builder aims to generate applicable knowledge resources for the system, and context information service offers user selection to interact with users according to the perspectives of project context.

As a result, the relevant part of project-based knowledge is determined for the user from the contextual view in the proposed framework. The structure of the project-based knowledge map is illustrated in the following subsection. Afterward, the essential components of the framework are also explained in separate subsections.

4.1. The structure of project-based knowledge map

The structure of project-based knowledge map is required to be flexible, expressive and navigable for representing the integrated project knowledge. Therefore, based on Topic Maps, context is specially proposed for enriching the structure of the project-based knowledge map. The first layer is the root entrance; the second layer contains the category names which are meaningful classification for topic names; the third layer contains the important topic names and associations; the bottom layer collects occurrences. Particularly, we define the context layer for conveying important operational experiences for explaining how topic names and occurrences are connected. The hierarchy outlook of a knowledge map is shown in Figure 6.

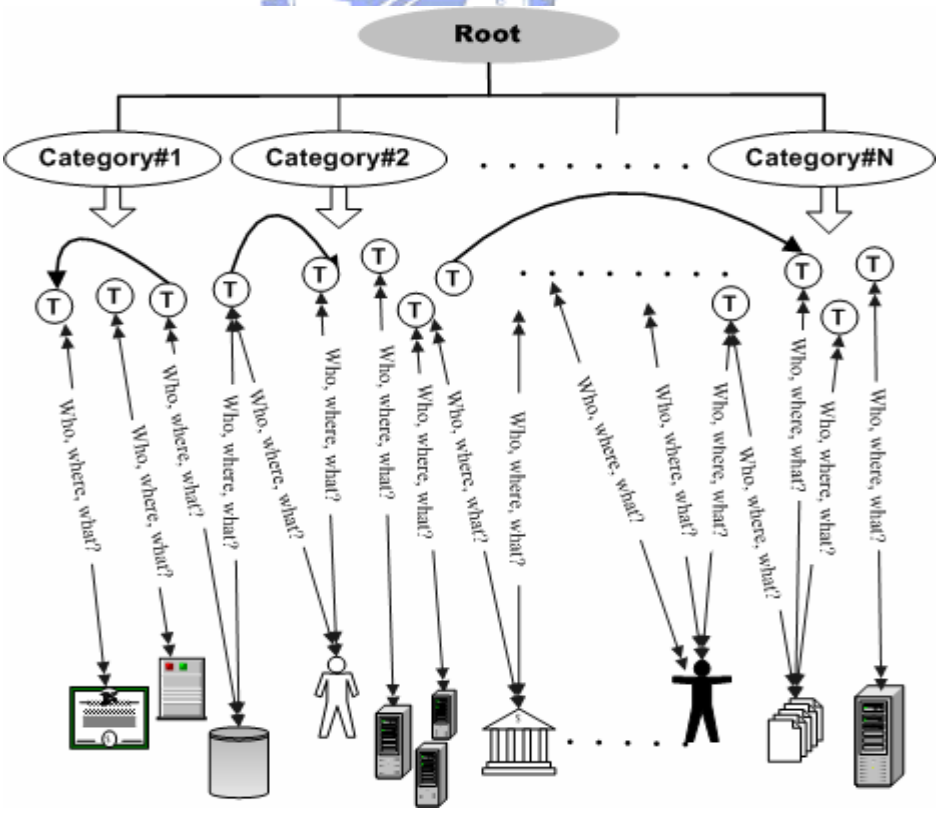


Figure 6. The structure of the project-based knowledge map

Category names are helpful to consistently classify topic names and to meaningfully transform association rules. Hence, a set of agreeing category names is rather important for users to introduce a project-based knowledge map. Generally, top-down and button-up are two basic processes to set the category names. Top-down process is suitable for well-grown domain, which has widely-accepted domain ontology, such as the species of animals. It is very useful to refer the ontology to conclude the category names. Button-up process is to generate the category names from the mergence or generation operation of topic names. Data mining can provide many algorithms to set the criterion to form the domain ontology gradually. The human expert, international standard, domain experience and ontology mostly are valuable to make the conclusion. However, a group of category names is such important to classify various topic names in a domain. Correctness, completeness, comprehensibility and consistence are important principles to define the category names.

Topic names are a set of important concepts in the knowledge map. Project attributes are described as formal topic names. The standard notation of Topic Maps for topic names is helpful to unify the concepts, including base name, display name and sort name. The base name is a name by which a topic name may be mostly used; the display name specifies the name to be formally explained to users, and the sort name specifies a name that is used in sorting procedures. For example, Taiwan Academic Network/TANet/TANET, is the base name, display name and sort name, respectively. This three-deck formation is helpful for reducing conflicts and misunderstanding in different operations.

Occurrences practically correspond to referable project objects. A group of multiple project objects may contain reports, files, videos, or program codes in which various important methods, models or theories are described with well-defined topic names for further user references.

Context offers the annotation of operational circumstances for topic names and related occurrences. A topic name is an important project attribute, and context is the useful information for identifying how the topic is applied in the occurrence. Particularly, with the help of context, the connection between topic names and occurrences becomes meaningful and expressive. Users are capable of

understanding and applying the integrated project knowledge efficiently.

Associations are used to indicate the certain relationships between topic names in the project-based knowledge map. An association is an abstract relationship between two topic names, and association name is helpful for explaining the relationship to describe how these topic names work together. For example, an association between the topics of 'call center' and 'CRM' is found, defining the association name of '*assist_in*' is helpful for making the rule statement that call center assists in CRM projects. Notably, context-relevant associations and context-specific associations are purposefully extracted in context perspectives for enhancing practical rule supports based on user situation.

Particularly, binary associations are the major extracted associations in the proposed framework. The advantage is helpful for the transformation of rule statements. The outcome increases the semantic support of the project-based knowledge map. The major advantages of binary associations are explained as follows.

- The semantics of higher-order associations which have high-arity of an association is often unclear.
- Higher-order associations can often be decomposed in binary associations.
- Most useful associations are binary.
- The further inference is consistently adopted in binary associations.

4.2. Project attribute builder

A project is composed of various project objects, each of which is associated with some certain models, tools, or methods, objectives, particular manipulations, and even special solutions. Therefore, project attributes are used for describing the important information for each project object. Moreover, this builder applies generalization data mining for diminishing the conflicts and faults in the definition of project attributes. Besides, the annotation of a project attribute is considered as the important project context which indicates how a project object

is applied, such as its conditions of activation and duration of validity. Each project attribute is annotated with a small set of project contexts. The fundamental undertaking of providing project operational information is useful for developing practical project knowledge in this framework.

4.3. Knowledge discovery module

A collection of consistent project attributes delivered from the project attribute builder is prepared for discovering internal knowledge patterns in this module. Since a project object is typically represented by a set of attributes and a project basically contains a group of project objects. Collecting all projects together will usually have a huge collection of project attributes. The high-dimensional data representation is easily caused. Moreover, a great diversity of projects also raises the dimensionality and burden the knowledge discovery task. Therefore, multiple-phase data mining approach is provided for discovering project knowledge in this module.

Table 4. The summary of clustering operations in the first phase

1st phase	Advantage	Disadvantage
Clustering operation	A certain degree of similarity in a cluster is helpful for facilitates further data mining operations.	The clustering result which is dependent on the clustering algorithm and threshold is difficult to be confirmed.
Without clustering	The cost of computation is saved in the first phase.	The heterogeneity easily increases the difficulty of further data mining operations.

In the first phase, clustering mining is proposed to group data as collecting numerous and manifold projects. The advantage is helpful for gathering similar project attributes together which are the essential foundation of a project-based knowledge map. With the benefit of similar project attributes in a cluster, the further association mining method is rather promising and productive. Lack of clustering operation in advance, the further data mining methods are tiresome and

laborious. The comparison is summarized in Table 4.

In the second phase, association mining is used for extracting the association relationship in a cluster of project attributes. An association relationship which is determined by the coexistences of some topics reveals the operational experiences applied in previous projects, such as some tools are used together in most cases. Furthermore, with the help of context information service, constraint-based association mining is used for extracting context-relevant associations and context-specific associations to support users with general-purpose rules and special-purpose rules distinctively. The advantage helps users to focus on the practical association rules for facilitating project development purposefully.

4.4. Context information service

Context information service mainly provides user selection to interact with users for simply perceiving user information needs. According to the result of user selection, the essential criteria for further knowledge discovery is therefore established for determining the relevant part of project-based knowledge map for users. Therefore, context information service simply helps users to focus on the appropriate project knowledge based on the indicated context conditions in the user selection.

Interacting with users for perceiving user interests and information needs are important in modern knowledge management systems for improving the efficiency of knowledge support. According to the five primary human-computer interaction styles identified by Shneiderman, menu selection is emerging as an important model of human-computer interaction, besides command language, form fill in, direct manipulation, and natural language [50]. The advantage of menu selection reduces keystrokes, improves recognition process, and conducts proper sequence of user interaction in the systems [33] .

Therefore, context information service takes the advantage of menu selection to list major context conditions in the simple user selection for interacting with users and recognizing user information needs. Moreover, the user selection is also adaptable to user dynamic changes without entering a series of keywords or

laboriously tracing user behaviors. Consequently, context information service is helpful for interacting with users and perceiving user interests to improve the user-dependent knowledge support in the proposed framework.

4.5. Project knowledge base

Project knowledge base mainly archives and organizes project knowledge in the structure of project-based knowledge map, including the important definitions of category and association names, formal definition template of topic names with project context and occurrence links, and the discovered association patterns. Furthermore, Topic Maps is the major model for maintaining the consistence and uniqueness of all definitions. The advantage is helpful to sustain the integrity and validity of the project-based knowledge map, as it benefiting the integration of extensible and complex project knowledge.

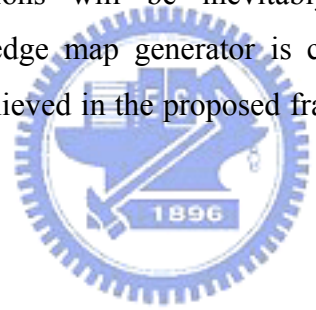
As developing the system of the project-based knowledge map, the database type of project knowledge base is dependent on the information system used in an organization. Based on the adaptability of the proposed framework, database system and file system are eligible to for project knowledge base. Basically, relational database systems established in most organizations provide mature technology to support the performance of project knowledge base. The advantage helpfully develops the system of project-based knowledge map based on the existing infrastructure of relational database systems in organizations without too much extra cost.

Moreover, the newly launched web services technology which can overall support the dissemination of information and knowledge over the network attract many applications and communities. Web services provide interoperability with open standards and protocols between various software applications running on disparate platforms, making it easy for developers and systems to comprehend. For considering the extension of the project-based knowledge map, the advantage of web services is beneficial to implement project knowledge base for further sharing and integrating project knowledge across different organizations and systems over the network easily and extensively.

4.6. Knowledge map generator

Knowledge map generator mainly converts association patterns into expressive rule statements and represents the selected parts of project-based knowledge map for users. Instead of primitive symbolic notations, the valuable associations are further transformed to natural language-like rule statements. The advantage helps users to understand the associations easily and to apply them efficiently in new projects. Currently, web-based pages are mainly employed for displaying the project-based knowledge map after the collaborative operations in the proposed framework. The advantage is helpful for user navigation on the project-based knowledge map over the network easily and rapidly.

Currently, knowledge map generator mainly provides the friendly interface with human users for efficient knowledge navigation. As web services technology becomes the major platform on the Internet, the interface with various machines in multiple applications will be inevitably conducted in this generator. Consequently, knowledge map generator is crucial for properly exporting the project knowledge achieved in the proposed framework to all kinds of users over the network.



Chapter 5. Discovering the Project-Based Knowledge Map

A primitive system is herein developed for illustrating the performance of the framework of project-based knowledge map. It was thanks to the support from an international R&D company in Taipei that we collected a set of project documents used for describing historical computer-based projects. The advantage of constructing the project-based knowledge map is helpful for integrating the valuable project knowledge to facilitate further project development in the organization.

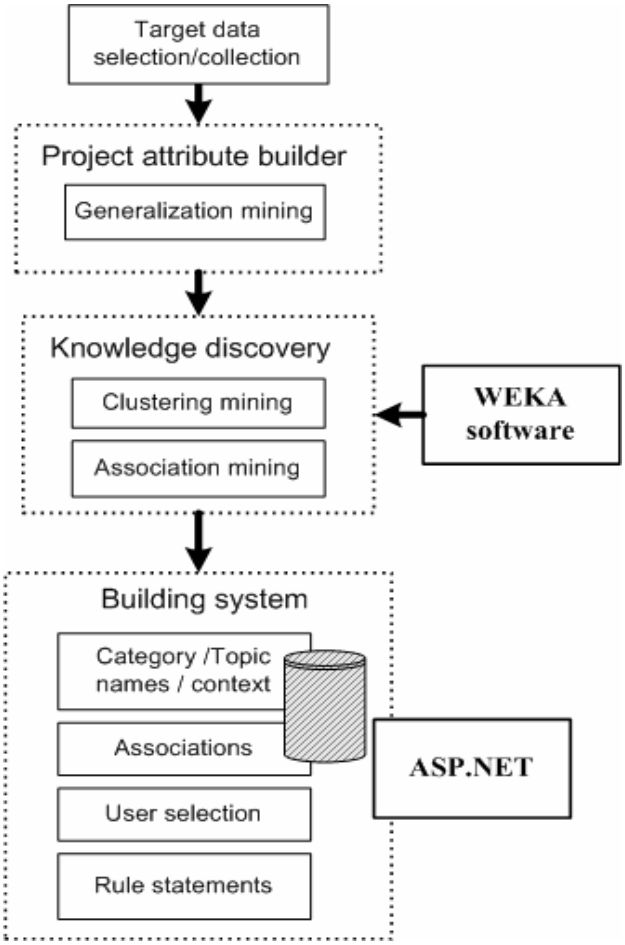


Figure 7. The experimental system flow

Based on the proposed framework, the experimental system flow for developing the project-based knowledge map is shown in Figure 7. Generating consistent project attributes, discovering and storing knowledge patterns, and

building the web-based system for displaying the project-based knowledge map are the major operations. Weka software is helpful for processing knowledge discovery task with various data mining methods [60], and ASP.Net is useful for building the database and system for illustrating the project-based knowledge map for users.

Historical projects in an organization gathering valuable organizational memories and intelligent assets usually are not open for research study or investigation. Also, for the consideration of confidentiality, some personal or commercial attributes are replaced with synthetic names, tiles or labels in this work. Basically, about thirty-six historical projects are collected and the involved project objects are described in two hundred and twenty-five project documents. According to the keywords originally provided in the documents, there are more than three hundred and twenty five keywords used to represent different attributes of project objects. Therefore, generating the consistent project attributes from the primitive keywords is the first essential work.

5.1. Generating consistent project attributes with project context

Generating consistent project attributes is the underlying endeavor to construct the reliable project-based knowledge map. Initially, project workers provided primitive project attributes for describing the project objects they have used in projects. Inevitably, many conflicts and faults are always occurred since they had various aspects and domain knowledge. Therefore, attributes removal and attribute generalization, two basic generalization mining operations, are applied in the project attribute builders. As shown below, many conflicts and errors are easily occurred in these primitive definitions.

(1) Synonyms: two different definitions of project attributes have the same or similar meaning. For example, Electronic Medical Record, Electronic Medicinal History and Electronic Patient Record (EPR), given by different authors, have the same meaning. Therefore, EPR is selected to represent all because it is mostly used.

(2) Homonym in acronym: the abbreviation widely used in information

system projects easily causes acronym misunderstanding. A term means differently in different domains. For example, ATM means auto seller machine in business applications; it means Asynchronous Transfer Mode in network area; it means Adobe Type Manager in commercial packages. This kind of homonym problem is prevented by the standard notation in Topic Maps. Also, the category name supplements the classification concept for user to realize the proper definition easily.

(3) User-defined terms: authors may propose their self-defined terms as keywords in meta information, such as KDWZ (knowledge discovery wizard system). Mostly these terms are removed for diminishing the noise data in the system.

(4) Deficiency of description: the original information is very important for further operations in this work. If not attainable, the expert's opinion will help a lot.

(5) Concept generalization: Since different terms may share the same concept, it is adaptive to use the general vocabulary for concept integration. For example, both B2C and B2B are generated to the topic name of *Electronic Commerce*. It is a very important procedure to integrate the concepts in this work. Also, concept generalization is helpful for reducing the amount of keywords. Moreover, the generation rules are required to apply continuously and consistently for ever unifying the concepts in the system.

As the description shown in Table 5, the project title, project object, project attribute and project context are arranged for completely expressing the project resources. The project DW007 contains two project objects of spec_dw.pdf and service_sys.doc. The important features of spec_dw.pdf are John Smith, OLAP and DBMS, and the important features of service_sys.doc are Mary Brown, call center, UML and CRM. Besides, the corresponding project context timely annotates the useful operational information and development conditions. For instance, as developing spec_dw.pdf of project DW007, John Smith was a SA and worked in the department of R&D in Taipei from 1998-1999; OLAP was operated by a DBA in spec_dw.pdf of project DW007 for using the operation of

cube to analyze sales data from 1995 to 2000 for the bank organization; DBMS was used by DBA and programmers in Taipei since 1998 until now for the bank and government organizations.

Table 5. The collection of project objects attributes with project context

Project title: DW007		
Project objects	**Project attributes**	**Project context (context type)**
spec_dw.pdf	John Smith	SA (role) Taipei (location), 1998-1999 (duration) R&D (department)
	OLAP	DBA (role) Cube (operation) Sales (client) 1995-2000 (duration) Bank (organization)
	DBMS	DBA (role) Programmer (role) Taipei (location) 1998- (duration) Government (organization) Bank (organization)
service_sys.doc	Mary Brown	DBA (role) Taipei (location) 1999-2002 (duration) R&D (department) Government (organization)
	Call center	Sales Manager (role) Taipei (location), NY (location) 2000- (duration) Government (organization)
	UML	SA(role) DBA(role) User case (operation) v1.4 (version no.) Bank (organization)
	CRM	Consultant (role) Sales Manager (role) Programmer (role) 2000-2005 (duration) Bank (organization) Hospital(organization)

Consequently, many ambiguities, errors and conflicts are herein removed to regulate the consistent project attributes for describing the important features of project objects. Therefore, not only the projects which consist of various project objects are described with representative project attributes, but also the corresponding project context provides the operational information for indicating how the project objects are applied. As a result, total one hundred and fifty-seven project attributes with proper project context are summarized for describing two hundred and twenty project objects in thirty-six projects. Substantially, the complete description as shown in Table 5 for each project objects is critical for preserving consistent project attributes with project context to construct the meaningful project-based knowledge map.

5.2. Predefining the important agreements

5.2.1.The definition of category names

The consistent project attributes with formal definitions are deliberatively introduced for the essential topic names of the project-based knowledge map. Furthermore, a well-defined set of category names is helpful to classify topic names for enhancing the concrete recognition of topic names. Currently, we summarize the opinions from domain experts and the IEEE Standard for Software Project Management Plans (SPMP IEEE Std. 1058-1998) for proposing five category names, including *Member, Tool, Standard, Goal,* and *Activity* [22].

- ***Member:*** key workers or project leaders
- ***Tool:*** the applied packages, methods, languages
- ***Goal:*** key purposes, objectives of projects
- ***Activity:*** held actions, events, contests in projects
- ***Standard:*** the most applied standards in projects

As a result, the integral declaration for a topic name in the system includes the 3-deck standard notion of Topic Maps and the proper category name. The formal declaration of topic names is as shown in Table 6. The advantage improves the meaningful expression of topic names and reduces the conflicts.

Table 6. The formal declaration of topic names

base name: John display name: John Smith sort name: Smith, J Category name : ***Member***	base name: OLAP display name: Online Analytical Processing sort name: OLAP Category name: ***Tool***
base name: CC display name: call center sort name: call center Category name: ***Activity***	base name: KM display name: Knowledge Management sort name: KM Category name: ***Goal***
base name: ATM display name: Adobe Type Manager sort name: ATM Category name: ***Tool***	base name: XML display name: Extensible Markup Language sort name: XML Category name: ***Standard***
base name: ATM display name: Auto Teller Machine sort name: ATM Category name: ***Activity***	base name: ATM display name: Asynchronous Transfer Mode sort name: ATM Category name: ***Standard***

5.2.2.The definition of association names

Next, the association names are helpful for enriching the conceptual classification of topic names and improving the meaningful transformation of rule statements. The template of binary association is $R(X, Y)$, where X is the antecedent topic name of the association, Y is the subsequent topic name of the association, and R is an abstract relationship for further definition of association name. According to the opinions from experts and senior project managers, four pre-defined association names are proposed in this work. The explanation is given below and the summarization is given in Table 7.

(1) ***use*:** It is meaningful when X is a person and Y is a tool.

(2) ***engage_in:*** It is meaningful when X is a person and Y is a standard, activity or goal.

(3) ***work_with:*** It is meaningful when X and Y are in the same category.

(4) ***assist_in:*** It is meaningful when X and Y are in the category set of Standard, Tool, Activity and Goal. X and Y are in the different categories, where X could be a standard, tool or activity.

Table 7. The pre-defined association names

The category of antecedent topic	Association name	The category of subsequent topic
Member	***use***	*Tool*
Member	***engage_in***	*Standard, Activity, Goal*
Member *Tool* *Activity* *Standard* *Goal*	***work_with***	*Member* *Tool* *Activity* *Standard* *Goal*
Standard	***assist_in***	*Goal, Tool, Activity*
Tool		*Goal, Standard, Activity*
Activity		*Goal, Standard, Tool*

5.3. Discovering knowledge with data mining approach

Two-phase data mining method is herein employed for knowledge discovery. The first phase employs clustering mining method to cluster project attributes into small groups. The second phase aims to discover internal association relationships for each cluster. Mainly, vector model is the basic data representation formation for the knowledge discovering operations. Furthermore, with the rapidly growth of project knowledge, we employ constraint-based data mining for determining relevant part of project-based knowledge map users from various context perspectives. The advantage facilitates the knowledge support and exploitation of the project-based knowledge map.

5.3.1. Vector model

A project attributes O_j is associated with a multi-dimensional vector, that is, $(w_{1j}, w_{2j}, \cdots, w_{kj})$, a weight value $w_{ij} \geq 0$ denotes the importance of attribute i on project j. For simplicity, this work uses the values 1 and 0 for the weight value to indicate whether the attribute is important (presence) or not important (absence)

to the project, respectively. Each dimension in the vector stands for a distinct term in the defined-term space of project attributes.

Based on the consistent project attributes, one hundred and fifty seven attribute dimensions are used for representing each project object. Therefore, vector models are intrinsically suitable to stand for multi-dimensional data in a computational format [24]. As shown in Table 8, an excerpt of the vector model describes the important attributes for each project object. Consequently, the further clustering and association mining operations are promisingly dependent on the vector model.

Table 8. An excerpt of the vector model for representing project attributes

Project	Project object	John	OLAP	DBMS	ISO	Mary	UML	CRM	KM
DW007	spec_dw.pdf	1	1	1	0	0	0	0	0
	service_sys.doc	0	0	0	1	1	1	1	0
	server_m1.sql	0	1	0	0	1	1	1	0
	manual_sys.doc	0	0	0	0	1	0	1	0
	long_sa.dpf	1	0	1	1	0	0	1	0
KK008	spec_user.pdf	0	1	1	0	1	0	0	1
	spec_sys.doc	1	0	0	1	0	1	0	1
	transfer.sql	1	0	1	0	0	0	0	1

5.3.2. Clustering process

Clustering mining is helpful for grouping ‘close’ data together based on similarity measurement. Forming clusters is to form clusters based on the dissimilarity matrix. A threshold value is an important parameter to decide how ‘close’ the project attributes will form a cluster. Project attributes with distance values less than a threshold value are grouped into the same cluster. For saving the computation cost in the high dimensional vector model, Euclidean distance metric and single-link agglomerative clustering method are used for grouping the collection of project attributes.

Table 9. A dissimilarity matrix using the measure of Euclidean distance

	KA001	KA002	KA011	KA100	KA101	KA109
KA001	0	4.9	5.6	3.4	***2.2***	5.7
KA002		0	5.5	4.5	4.6	5.8
KA011			0	5.5	5.6	***2.3***
KA100				0	3.2	5.9
KA101					0	5.6
KA109						0

According to the example of dissimilarity matrix given in Table 9, project attributes KA001, KA101, KA100 and KA002 could form a cluster, and project attributes KA011 and KA109 could form a cluster if the given threshold value was 5.0. The algorithm places each object in its own cluster and gradually merges these atomic clusters into larger and larger clusters. If the threshold values is greater than 6.0, all objects are grouped are in a single cluster. The clustering dendrogram/tree of different threshold values is shown as in Figure 8.

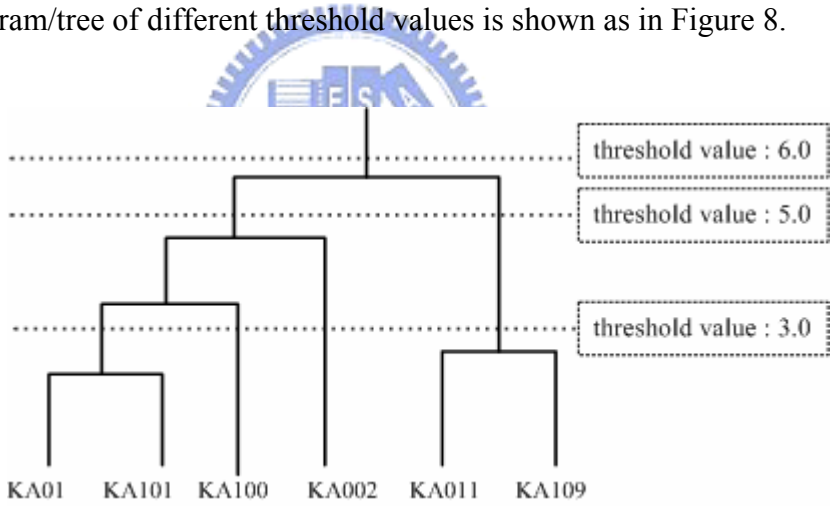


Figure 8. Agglomerative clustering forms clusters in various thresholds

As a result, one hundred and fifty-seven topic names are grouped into three clusters. The clustering outcome is shown in Table 10. Each cluster of topic names is prepared for discovering association patterns to construct the foundation of project-based knowledge map.

Table 10. The result of the clustering mining operation

Cluster No.	Content
#1 (83)	John Smith, Mary Brown, Joyce English, Robert King, Stephen Adams, David Campbell, DBMS, OLAP, Java, Clustering, Neural Net, Fuzzy set, Active Service Page , Visual Basic, Flash, Genetic Analysis, ANOVA, Strength Weakness Opportunities Threats, Call center, Certificate Authorities, Conference, Marketing, Contest, Questionnaire, Sampling, Knowledge Management, Tax, Customer Relationship Management, Enterprise Resource Planning, Electronic Commerce, e-Learning, Point of sale, Logistics, Enterprise Information Portal, ISO, Unified Modeling Language, XML, Synchronized Multimedia Integration Language, Simple Object Access Protocol, Health Level 7, MP3, Document Object Model, Secure Sockets Layer, Scalable Vector Graphics, ebXML, Standard Generalized Markup Language, XML Access Control Language, Extensible User Interface Language, Geography Markup Language, Concurrent Versions System, Extensible HyperText Markup Language, Microsoft Extensible Application Markup Language, Agent System, Text Categorization, Text Mining, Business Process Reengineering, Web clipping, Web-based system, Frank Hale, Sharon Regan, Lisa Taylor, Bobby Kao, Steven Thomas, Karen Lee, Nancy Rice, Virtual Reality Modeling Language (VRML), Electronic Business, ActiveX, Distributed Component Object Mode (DCOM), Enterprise Application Integration (EAI), Universal Description Discovery and Integration (UDDI), Java Server Page (JSP), Quality management, Executive information system (EIS) , Materials requirements planning (MRP), Object Linking and Embedding (OLE), Supply chain management (SCM), Electronic mail, Computer-aided software engineering (CASE), index server, OSP (online service provider), Very large database (VLDB), Robert Green, Peter Martin
# 2 (55)	John Smith, Mary Brown, Frank Hale, Warren Chen, Jennifer Liu, Authentication, Consumer Behavior, Decision Support Systems, Electronic Document Delivery, Fuzzy reasoning, Information Acquisition, Instructional Strategy, Learning Outcome, Recency Frequency Monetary (RFM) model , Learning Process, Learning Context, Longest Detour Problem, Material Requirement Planning, Proxy server, Political Behavior, Mobile Communication Networks, Spanning Sub-graph, Real-Time Model Selection, Regular Graphs, Encryption scheme, Conflict Theory, Social Exchange, Problem-solving Technique, Template Recommendation, Decision Model, Digital Library, Mobil Commerce, Graph Sandwich, Home Location Register, Overflow Control, Homogeneous Set,

	Idea-Generation Support System, Peer-to-Peer, Just-in-time, Message Recovery, Shop Floor Control Information System, Mobility Management, Perfect Replacement, Credit Card Payment, Organizational Politics, Simulation Analysis, Strategy Contingency Theory, Structuration Theory, Task Technology Fit, Petri Net, Stephen Adams, David Campbell, Nancy Hopkins, June Matthews, Robert Liu
# 3 (19)	Congestion Control, Technology Valuation, Active Delay Control, Automated Negotiation, Game Theory, Groupware Development, Trust Dynamics, Virtual Group, TCP flows, Multi-issue Negotiation, Technology Transfer, Information Systems Strategic Planning (ISSP), Interpretive Scheme, Structural Equation Modeling (SEM), Bobby Kao, Steven Thomas, David Campbell, Joyce English, Karen Lee

5.3.3.Context information service

As interacting with users, context information service summarized the important context conditions from project context, including the context types of role, location and organization. The description is shown in Table 11.

Table 11. The context conditions in the user selection

Context type	Description	Conditions
Role	normal or customary title of a person in the project team.	database administrator systems analyst project manager, sales manager commercial designer senior programmer junior programmer consultant software engineer auditor
Location	the site or place of a certain object in the project, such as a person, or an activity.	Taipei London Los Angeles New York Tokyo
Organization	the company or organization who supports the projects or activities.	Bank Account Hospital Government Manufacture Academy Trade

The context types of role, location and organization are important for describing the background of the project developers. Therefore, the user can chose the relevant context conditions for explicitly indicating the user interests. Instead of intentionally entering keywords, context information service offers an easy yet useful selection menu for users to perceive user information needs. Notably, the result of user selection is therefore used as the essential criteria for determining the relevant part of project-based knowledge map in the following association mining operations.

5.3.4. Association rule mining

Based on a cluster of similar project attributes, the further internal associations are rather valuable for developing project-based knowledge map. In order to extract the useful associations to meet user current information needs, user context is therefore applied as the important criteria in the association mining operation. According to the context conditions selected by users, various constraints are therefore established for extracting the useful associations to support different situations.

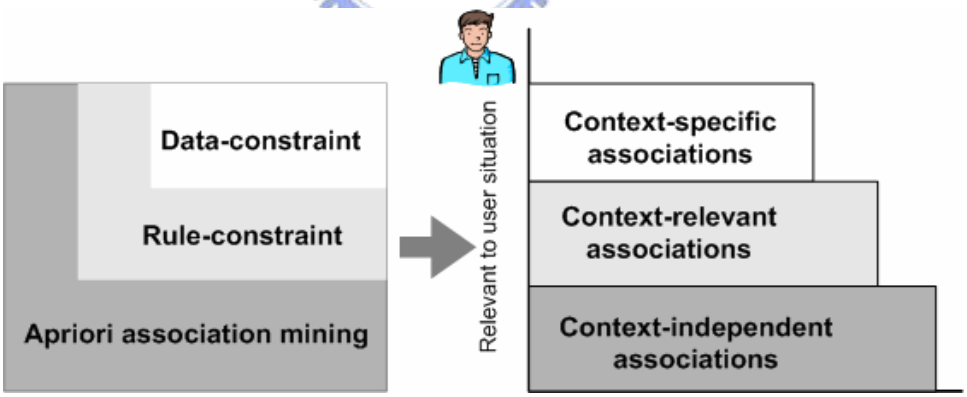


Figure 9. Context in constraint-based association mining

Since the user selection is proposed for improving the understanding with users, this work provides three-level association rule mining for extracting useful associations patterns from various context perspectives. As shown in Figure 9, various constraint-based association mining methods are used for extracting

context-independent associations, context-relevant associations and context-specific associations to support various information needs from different context perspectives. The details are explained as follows.

- The first phase applies Apriori algorithm for mining context-independent associations. The outcome offers overall associations for users without differentiating user situation [32].
- The second phase employs rule-constraint mining operation for selecting context-relevant associations. The consequence provides generally workable rules based on the outcome of user selection. The advantage saves the cost of manual separation for users.
- The third phase employs data-constraint mining operation for discovering context-specific associations for users. The outcome extracts various specific associations based on different context perspectives. The advantage supports users with pertinent associations to learn the pertinent rules directly and efficiently for facilitating further project development.

5.3.5.Context-independent association mining

In the first phase, we aim to discover general association patterns without considering user selection. Each project object in the occurrence layer of project-based knowledge map is linked to some proper topic names and is regarded as a transaction of topic names. Collecting the transactions of topic names for mining the topic names mostly appeared together in project objects is helpful for discovering useful associations. Therefore, Apriori algorithm, the well-known association mining method, is herein applied for extracting context-independent associations [1].

Procedure Apriori association mining
Input: (*D*: the collection of transactions of topic names, minimum support threshold)
Ouput: (*CIA*: the collection of context-independent associations in *D*)

C_k: Candidate itemset of size k;
L_k : frequent itemset of size k;
for (k = 1; L_k !=; k++) do
  C_{k+1} = candidates generated from L_k;
  for each transaction $t \in D$ do increment the
    count of all candidates in C_{k+1} that are contained in t
  L_{k+1} = candidates in C_{k+1} with min_support
end for;
Return $CIA = \cup_k L_k$;

Figure 10. Apriori association rule mining

Let *D* be the collection of the transactions of topic names in the cluster. The operation of Apriori association mining generates *CIA*, the fundamental set of context-independent associations. The procedure is shown in Figure 10, and the effect generates a set of fundamental associations for further constraint-based association mining operations. Accordingly, more than twenty association rules are extracted from the cluster 1 as the support rate is 20%, as shown in Table 12.

Table 12. The set of extracted context-independent associations

OLAP ⇒ DBMS	KM ⇒ XML	Sharon ⇒ Lisa
ISO ⇒ EPR	Karen Lee ⇒ JSP	SOAP ⇒ DOM
DBMS ⇒ Call center	EIS ⇒ KM	XACL ⇒ XUL
Call center ⇒CRM	GML ⇒ GIS	ActiveX ⇒ OLE
David ⇒ Flash	ebXML ⇒ EB	SCM ⇒ CASE
John ⇒ Java	Web clipping ⇒ Agent	MP3 ⇒ SMIL
John ⇒ Mary	MRP ⇒ SCM	SGML ⇒ HTML
GA⇒ Frank	DCOM ⇒ OLE	Marketing ⇒ SWOT

The benefit covers the overall associations to the utmost, but the outcome easily confuses users without further separation and classification. When the project schedule is tight, the use has no idea which association rule is most productive or practicable. Some research works ordered the association rules by the support or confidence rate for assisting users to make the proper choice. However, it is not really efficient to support users. Therefore, the context-oriented solutions, context-relevant association mining and context-specific association mining, are proposed for further selection and differentiation to support users based on various context perspectives.

5.3.6.Context-relevant association mining

Then, the result of the user selection which indicates user information needs is applied as filtering constraints. For each association in the set of context-independent associations, if the context of the involved topics satisfies the context conditions of user selection, then the association will be added to the set of context-relevant associations. The procedure is shown in Figure 11.

Procedure Context_relevant association mining
Input: (*US*: the set of context conditions of user selection,
CIA: the set of context-independent associations)
Output: (*CRA*: the set of context-relevant associations)

CRA = {}
for each context-independent association R_k of *CIA* do
if R_k satisfies a context condition contained in *US*
then R_k is added to *CRA*
return *CRA*;

Figure 11. The procedure of context-relevant association mining

Let *CIA* be the set of context-independent associations which has been extracted in previous phase and *US* be the set of context conditions specified by

the user. For example, the result of user selection is DBA, Taipei and Government to indicate the context conditions of role, location and organization respectively. The operation of context-relevant association mining generates *CRA*, the set of context-relevant associations. The advantage separates the relevant associations for users according to the user selection. The conceptual outcome of the context-relevant association mining operation is given in Figure 12. From the collection of context-independent associations extracted in previous phase, the selected context conditions of DBA, Taipei and Government are applied for determining the context-relevant associations which satisfy the context conditions.

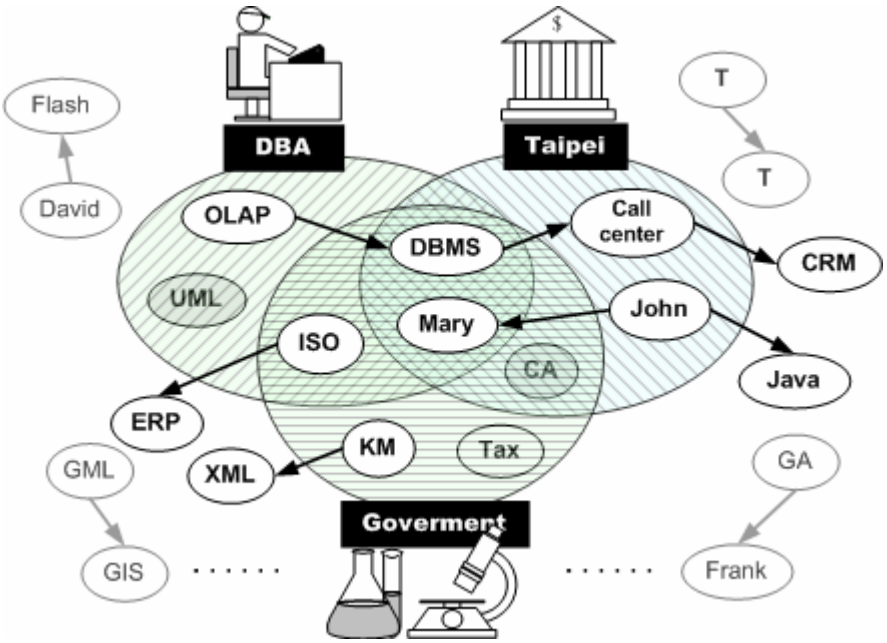


Figure 12. The conceptual outcome of context-relevant associations

The result of the context-relevant mining procedure is shown in Table 13. Based on the outcome in Table 12, the selected context conditions is the important criterion for separating the relevant associations. As a result, the context-relevant associations contains (OLAP, DMBS) (ISO, ERP), (DBMS, Call center), (Call center, CRM),(John, Java),(John, Mary) and (KM,XML).

Table 13. The result of the context-relevant association mining procedure

Context-independent Associations	Project Context (context type)	
(OLAP, DMBS)	**DBA** (role) Cube (operation) Sales (client) 1985-1995 (duration) Bank (organization)	**DBA** (role) Programmer (role) **Taipei**(location) 1998- (duration) v6.0 (version no.) **Government**(organization)
(ISO, ERP)	**DBA**(role) Consultant (role) 9002 (version no.) **Government** (organization)	Consultant (role) Manager (role) 2000-(duration) Manufacture (organization)
(DBMS, Call center)	**DBA** (role) Programmer (role) **Taipei**(location) 1998- (duration) v6.0 (version no.) Bank (organization) **Government**(organization)	Sales Manager (role) **Taipei** (location), NY (location) 2000- (duration) Bank(organization)
(Call center, CRM)	Sales Manager (role) **Taipe**i (location), NY (location) 2000- (duration) Bank(organization)	Consultant (role) Sales Manager (role) Programmer (role) 2000-2005 (duration) Bank (organization)
(David, Flash)	Programmer(role) London (location) 1999-(duration)	v7.0 (version no.) 2000-2002(duration) Bank (organization)
(John, Java)	SA (role) **Taipei** (location), 1998-1999 (duration) R&D (department)	v1.3 (version no.) 1998-2000 (duration)
(John, Mary)	SA (role) **Taipei** (location), 1998-1999 (duration) R&D (department)	**DBA** (role) **Taipei** (location) 1999-2002 (duration) R&D (department) **Government**(organization)
(GA, Frank)	…..	…..
(KM,XML)	Consultant (role) Sales Manager (role) Programmer (role) 2000-2005 (duration) **Government** (organization) Hospital(organization)	Programmer (role) SA (role) 1999-(duration) Bank(organization) Hospital (organization)
⋮	⋮	

Context type	User Selection	Context-relevant Associations
role	***DBA***	(OLAP, DMBS) (ISO, ERP) (DBMS, Call center)(Call center, CRM)(John, Java)(John, Mary) (KM,XML)
location	***Taipei***	
organization type	***Government***	

5.3.7.Context-specific association mining

In the third phase, context-specific association mining applies data constraints as the major guidance for extracting internal associations from context-specific topic names. First operation selects the topic names which satisfy the context conditions of user selection from all topic names in the project-based knowledge map to generate the set of context-specific topic names. Then, Apriori algorithm is applied again for extracting the context-specific associations from the set of context-specific topic names. The procedure is shown in Figure 13.

Procedure Context_specific association mining
Input: (*US*: the set of context conditions of user selection,
(*D*: the collection of transactions of topic names)
Output: (C*SA*, the set of context-specific associations)

CSD: the collection of context specific transactions of topic names
CST = {}
for each transaction T_k of *D* do
if T_k satisfies a context condition contained in *US*
then T_k is added to *CSD*
end for
Call Apriori procedure to extract *CSA*, the set of context specific association patterns from *CSD*.
return *CSA*;

Figure 13. The procedure of context-specific association mining

Let *US* be the set of context conditions specified by the user and *D* be the collection of transactions of topic names. Let For example, the result of user selection is DBA, Taipei and Government to indicate the context conditions of role, location and organization type, respectively. The operation of context-specific association mining first generates *CSD*, the collection of context specific transactions of topic names, and then generates *CSA*, the set of context-specific associations. The advantage purposefully discovers the certain associations which directly related to user selection for efficiently user-dependent

knowledge support. The conceptual outcome is shown in Figure 14. Based on the context-specific topic names which satisfy the selected context conditions, the association mining is applied again for further mining the hidden associations.

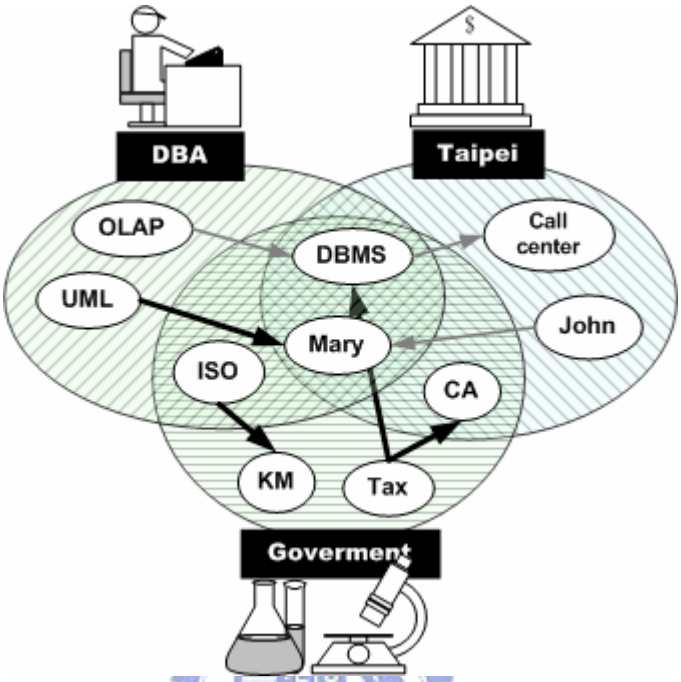


Figure 14. The conceptual outcome of context-specific associations (a)

Based on the cluster 1, the topic names whose project context satisfy the selected context conditions are collected for forming the set of context-specific topic names. Then, association rule mining method is further applied for extracting the internal association patterns. The result is given in Table 14.

Table 14. The process result of context-specific associations (a)

Topics	Project context (context type)	Topics	Project Context (context type)
OLAP,	**DBA** (role) Cube (operation) Sales (client) 1985-1995 (duration) Bank (organization)	Java	v1.3 (version no.) 1998-2000 (duration)
DBMS,	**DBA** (role) Programmer (role) **Taipei**(location) 1998- (duration) v6.0 (version no.) **Government**(organization)	ISO,	**DBA**(role) Consultant (role) 9002 (version no.) **Government (organization)**
Call center,	s Sales Manager (role) **Taipei** (location), NY (location) 2000- (duration) Bank (organization)	ERP	Consultant (role) Manager (role) 2000-(duration) Manufacture (organization)
CRM	Consultant (role) Sales Manager (role) Programmer (role) 2000-2005 (duration) Hospital(organization)	John,	SA (role) **Taipei** (location), 1998-1999 (duration) R&D (department)
UML,	SA(role) DBA(role) User case (operation) v1.4 (version no.) Bank (organization)	KM,	Consultant (role) Sales Manager (role) Programmer (role) 2000-2005 (duration) **Government** (organization) Hospital(organization)
Mary,	**DBA** (role) **Taipei** (location) 1999-2002 (duration) R&D (department) **Government**(organization)	XML	Programmer (role) DBA (role) 1999-(duration) Bank(organization) Hospital (organization)
CA	Consultant (role) **Taipei** (location) London (location) 2003-(duration) **Government**(organization)	Tax	Manager(role) Consultant (role) 1995- (duration) **Government**(organization)
:	:	:	:

Context-specific transaction with Topic Names	**Apriori algorithm**	**Context-specific Associations**
The shadow area in above table.	→	**(UML, Mary)**, **(ISO, KM)**, **(Tax, ISO)**, **(Tax, CA)** *(DBMS, Call center),(OLAP, DMBS), (John, Mary)*

Furthermore, each context condition in user selection is the criteria to select precisely interesting transactions for Apriori association mining. Therefore, data constraint is applied again for processing context-specific association mining with small size of context conditions. First operation selects the topic names which satisfy single context condition of user selection from the set of context-specific transactions of topic names. Second operation is also to apply Apriori algorithm again to extract the context-specific associations. The procedure is to repeat above operations for all context conditions in user selection.

The conceptual outcome is shown in Figure 15. We apply the role of DBA as the single criteria for separating the DBA-specific topics, including ISO, DBMS, OLAP, Mary, and UML. Next, the new association of (ISO, DBMS) is found for reminding users the specific relationship between ISO and DBMS. The process is similar in the context conditions of Taipei and Government.

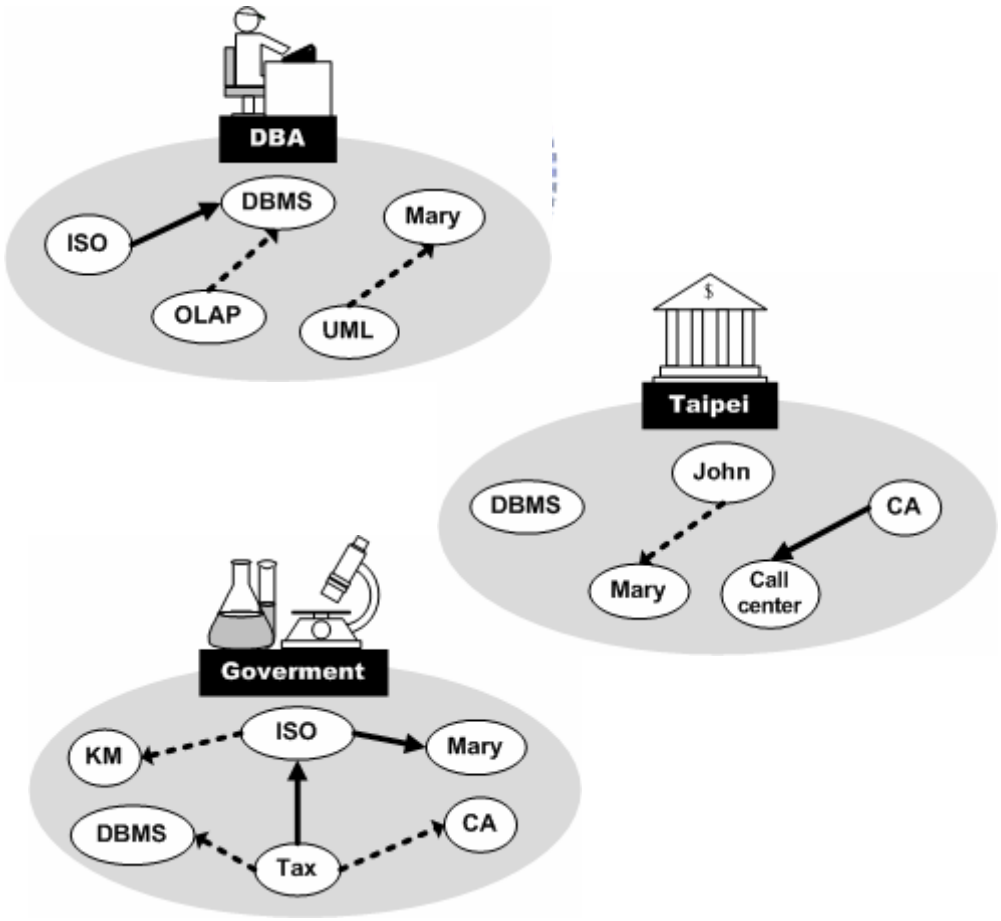


Figure 15. The conceptual outcome of context-specific associations (b)

Based on the cluster 1, the context-specific association is given in Table 15. The sets of DBA-specific topics, Taipei-specific topics and Government-specific topic are selected separately. Next, employing association rule mining in each set is helpful extracting the context-specific associations separately according to the context conditions of DBA, Taipei, and Government. Comparing with the result of context-relevant association mining, several new associations are extracted from various perspectives, such as the associations of (ISO, DBMS), (CA, Call center), (ISO, Mary) and (Tax, ISO).

The advantage is helpful for intentionally focusing on certain associations for one certain condition. The user can pay more attention to the association between ISO and DMBS if the user cares about the condition of being a DBA. Precisely locating the certain associations is useful for learning the special solutions for special condition rapidly.

Table 15. The process result of context-specific associations (b)

User Selection	Context-specific Topics	Context-specific Associations
DBA (role)	ISO, DBMS, OLAP, UML, Mary	(ISO, DBMS) *(OLAP, DBMS)* *(UML, Mary)*
Taipei (location)	DBMS, John, Mary, CA, Call center	(CA, Call center) *(John, Mary)*
Government (organization)	KM, ISO, CA, DBMS, Tax, Mary	(ISO, Mary) (Tax, ISO) *(ISO, KM)* *(Tax, DBMS)* *(Tax, CA)*

5.4. Transforming binary associations into rule statements

The association type in the framework of project-based knowledge map is proposed as binary associations, instead of complicated n-ary associations. The formal definition for binary associations is proposed in Figure 16. The irreflexive, non-transitive and strictly anti-symmetric properties reserve the consistence of the binary associations and the consistent rule transformation in this work.

A *binary relation B* over a set of topic names *T* is a subset of *TxT*.
If $(x, y) \in B$, It can also be written as *x B y*.
Let *B* be a binary relation over a set *T*.

(1) *B* is *irreflexive* means: $(x, x) \notin B$ for all $x \in T$.

(2) *B* is non-transitive means: x*By* and *y*Bz do not imply x*B*z for all x, y, z $\in T$.

(3) *B* is strictly anti-symmetric means: x*By* does not imply y*B*x for all $x, y \in T$.

Moreover, as an extensional definition of the relation *B* is given, the association becomes a meaningful statement. For example, *B* is defined as a_father_of, then x*B*y can be interpreted as: *x is a father of y*.

Figure 16. The proposed definition for binary associations

For example, *(XML, HTML)* is an extracted association, and the properties and rule transformation are explained as below:

- irreflexive: The association of *(XML, HTML)* does not imply that the associations of *(XML, XML) and (HTML, HTML)* exist. The advantage reduces the redundant rules.
- non-transitive: If the association of *(HTML, Flash*) is also extracted with the association of *(XML, HTML)*, it does not imply that the association of *(XML, Flash*) exist.
- anti-symmetric: It means the that binary associations are directional. The association of *(XML, HTML)* does not imply that the association of *(HTML, XML)*exists.
- rule transformation: Moreover, the predefined association *work_with* is meaningful in the rule transformation. The rule statement is converted: *XML works with HTML*. Moreover, the category names of *XML* and *HTML* increasingly encourage the completeness of rule statement. As a result, the formal rule statement is given as follows.
  The standard of XML works with the standard of HTML.

Consequently, transforming binary associations into natural language-like rule statements is helpful for interpreting the extracted associations for users to enhance the meaningful expression of the project-based knowledge map.

5.5. Displaying the project-based knowledge map

Visually displaying project-based knowledge map is effective for illustrating the performance of the project-based knowledge map. Based on Topic Maps, the project knowledge is represented in the hierarchy of category, topic names and project occurrences. Notably, project context is helpful for providing operational information as displaying association rule statements for users. Knowledge navigation and exploitation are efficiently fulfilled in the project-based knowledge map.

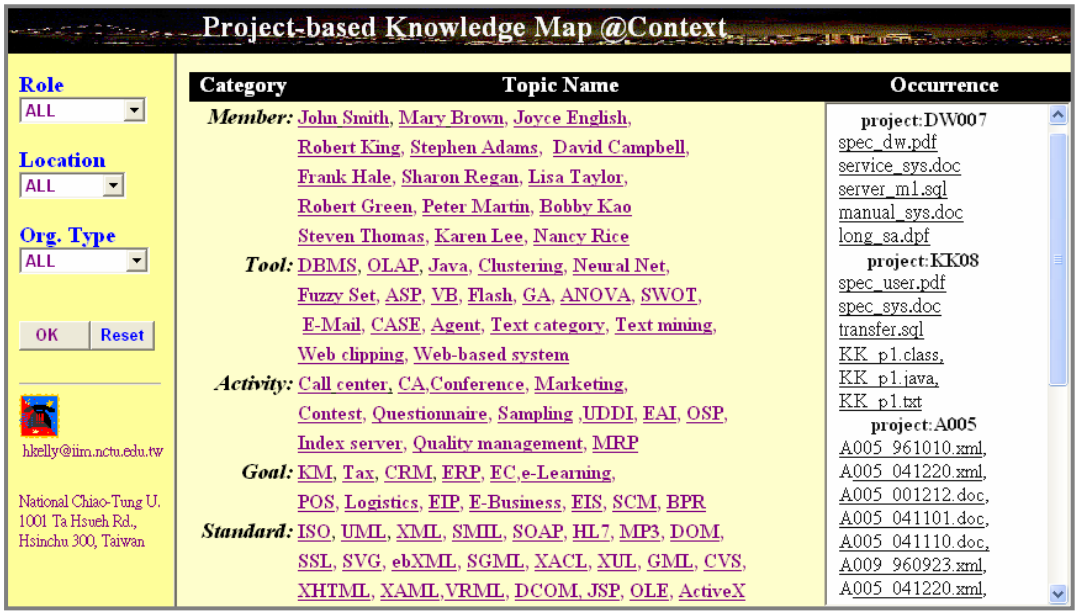


Figure 17. Displaying the project-based knowledge map

First, as shown in Figure 17, context-independent knowledge map provides the extensive topics and associations for users. The advantage is helpful for completely learning project knowledge. Then, the user can click the selection menu of role, location and organization type for explicitly expressing the user information needs. Next, the context-oriented operation is triggered to construct the relevant part of project-based knowledge map according the result of user selection, including referable project resources, relevant topics, and up way to

relevant association rule statements.

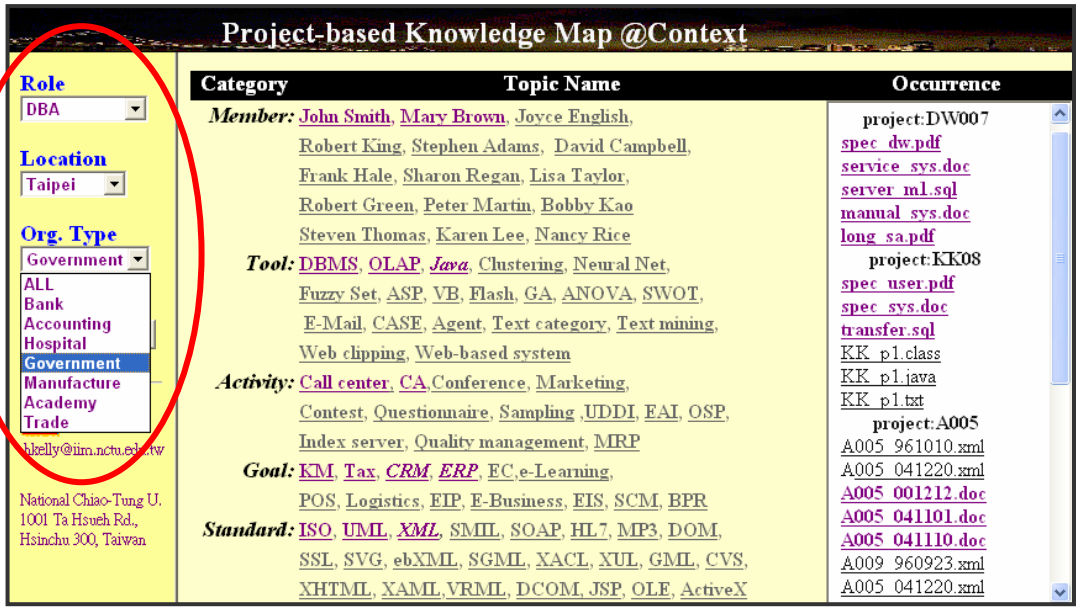


Figure 18. Displaying context-relevant topic names and occurrences

Consequently, as shown in Figure 18, context-relevant topics related to specified context conditions is determined for users, including context-specific topics and context-associated topics. The relevant project occurrences are also selected according to the relevant topic names. The advantage helps user to concentrate on the related concepts without manual separation. That is very useful as the project schedule is tight and insufficient support from human experts.

Project context is very useful in explaining association rule statements. The defined association names and the extracted associations are transformed to meaningful rule statements in natural language expression. Furthermore, project context annotated the involved topic names with operational experiences, such as the involved role, location, version number and the employment duration. As shown in Figure 19 and Figure 20, context-specific associations and context-relevant associations are purposefully extracted from various context perspectives to support different situations. The advantage classifies the rule statements for users to support different user situations. As a result, the user can learn the association rules easily from the statements and know how to apply the rules efficiently from the timely project context.

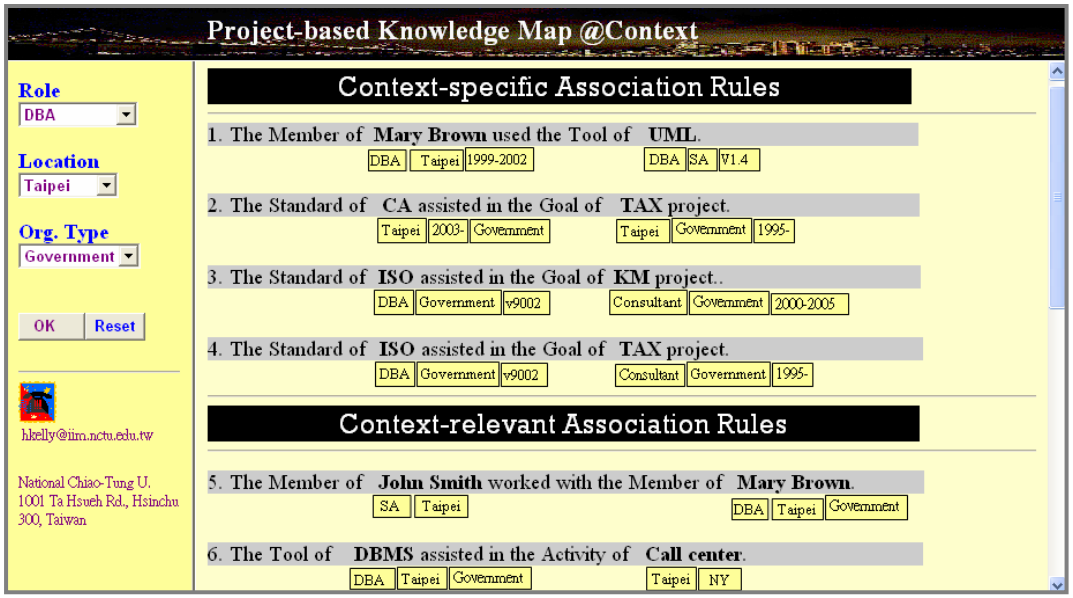


Figure 19. Displaying context-specific association rules with context

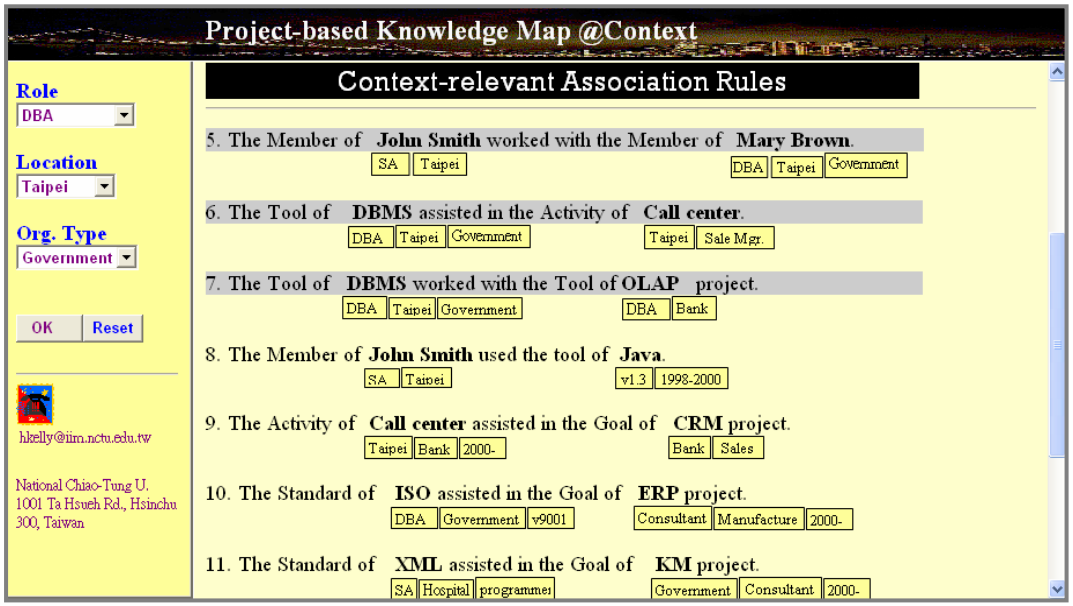


Figure 20. Displaying context-relevant association rules with context

Chapter 6. The Evolution of the Project-based Knowledge Map

With the help of Topic Maps, the representation of project-based knowledge map is meaningful to human users, including the category names, topic names and association names linking in an expressive hierarchy. With popularity of Internet, the semantics in the project-based knowledge map is rather important for disseminating the integrated project knowledge across various applications over the network. Therefore, XML/RDF syntax is proposed for improving the semantic evolution of the project-based knowledge map. The advantage not only implements the inherent hierarchy of the project-based knowledge map, but also efficiently facilitates the exploitation of the project-based knowledge map across different organizations, applications and systems over the network.

Different from the referential integrity in traditional database systems, RDF provides flexible and agreed grammar for describing binary associations. Moreover, the corresponding RDF graph visualizes the relationship and validates the RDF syntax. Accordingly, the meaningful layer of associations proposed in the framework is increasingly benefited from RDF/XML techniques.

Therefore, in the second half of this research employs XML/RDF specifications for promoting the semantic evolution of the project-based knowledge map whereas the potentiality and popularity of the XML technique is increasing in the Internet-based systems. The details are introduced in the following subsections. The outlook of the project-based knowledge is same to human users, but the inner semantics of the project-based knowledge map becomes native and interoperable in RDF/XML specifications. Therefore, the following selections explain how the important components of the project-based knowledge are represented in RDF/XML syntax.

6.1. The project context in RDF/XML

The important components of the project-based knowledge map, including topic names, project context, occurrence, and associations are appropriately

represented in the RDF/XML syntax. Unlike the referential integrity manually predefined in traditional database, the agreed structure in RDF/XML is like the natural language processing. Consequently, the typical grammar is not only readable to human users, but also exchangeable across different applications over the network.

6.1.1.Topic names in 3-deck format

According to the notation of Topic Maps, a 3-deck format topic name contains base name, display name and sort name. Therefore, as shown in Table 16, DTD syntax is helpful for governing the consistence of topic names, and few topic names are given behind for illustrating the syntax regulation.

Table 16. An excerpt of XML/DTD to declare a topic name

```
<! ELEMENT    ProjTopic (baseName, displayName, sortName) >
<! ATTLIST ProjTopic
          TopicID      ID          #REQUIRED
           Category    (Member| Tool| Standard | Goal| Activity|)     #REQUIRED   >
      <! ELEMENT    baseName          (#PCDATA)       >
      <! ELEMENT    displayName       (#PCDATA)       >
      <! ELEMENT    sortName          (#PCDATA)       >
:
<ProjTopic TopicID=”CRM”      Category=”Goal”>
       <baseName>CRM</baseName>
       <displayName>C.R.M.</displayName>
       <sortName>Customer Relationship Management</sortName>
</ProjTopic>
<ProjTopic TopicID=”John Smith”      Category=”Member”>
       <baseName>John Smith</baseName>
       <displayName>J. Smith</displayName>
       <sortName>Smith, J.</sortName>
</ProjTopic>
<ProjTopic TopicID=”OLAP”      Category=”Tool”>
       <baseName>OLAP</baseName>
       <displayName>O.L.A.P.</displayName>
       <sortName>Online Analytical Processing </sortName>
 </ProjTopic>
 :
 :
```

6.1.2.Flexible project context

Furthermore, we apply RDF graph for illustrating the one-to-many relation between the attribute and context. For example, the specification file shown in Figure 21 has two essential attributes “John Smith” and “OLAP”. Firstly, the formal declaration of topic names is helpful for preserving the consistence among topic names, as shown in Table 16. Then, an occurrence can utilize the define topic names as important attribute and annotates further operational information.

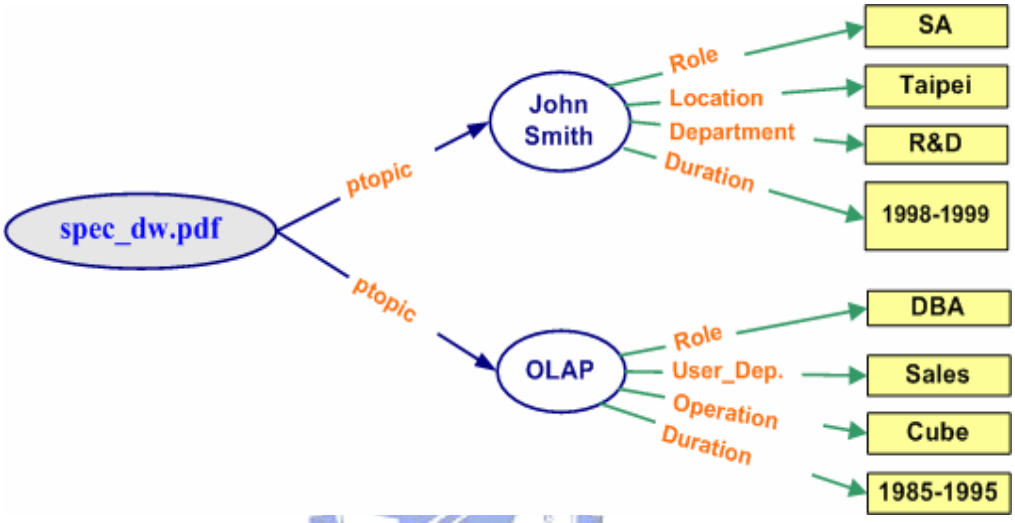


Figure 21. RDF graph illustrates the project attribute with context

RDF syntax is given in Figure 22. With the help of context, the background information of “John Smith” and the operational information of “OLAP” are provided for users to know the situation of “John Smith” was and how “OLAP” was operated as developing the specification. Moreover, context reveals the operational information for users. For example, John Smith was a SA in the department of R&D in Taipei, and had worked for the specification file from 1998 to 11999. Also, the operation of cube in OLAP is applied for the sales department, and DBA is the role of the worker who utilized OLAP.

The corresponding RDF syntax implements the semantics of. Therefore, not only the important attributes of projects are defined, but also the operational information is also timely provided and archived. The link between topic names and occurrences in the project-based knowledge map becomes expressive and directional.

```
<?xml version="1.0"?>
      <rdf:RDF xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:pkmap = "http://www.nctu.iim.tw/KM/pkmap#"
           xml:base = "http://www.nctu.iim.tw/KM/pkmap/ptopics.xml">
         <rdf:Description
rdf:about="http://www.nctu.iim.tw/PJS/DW007/spec_dw.pdf">
              <pkmap:ptopic rdf:resource="#John Smith" />
              <pkmap:ptopic rdf:resource="#OLAP" />
         </rdf:Description>
         <rdf:Description rdf:about="#John Smith">
                <pkmap:Role>SA</pkmap:Role>
                <pkmap:Location>Taipei</pkmap:Location>
                <pkmap:Department>RnD</pkmap:Department>
                <pkmap:Duration>1998-1999</pkmap:Duration>
         </rdf:Description>
         <rdf:Description     rdf:about= "#OLAP">
            <pkmap:Role>DBA</pkmap:Role>
            <pkmap:Department>Sales</pkmap:Department>
            <pkmap:Operation>Cube</pkmap:Operation>
            <pkmap:Duration>1985-1995</pkmap:Duration>
         </rdf:Description>
      </rdf:RDF>
```

Figure 22. Project context in RDF syntax

6.2. The association rule statements in RDF/XML

The semantic representation is essentially important in the proposed framework, including the definition of 3-deck topic names, triple association patterns, project context, and the transformation of rule statements. As developing the framework of project-based knowledge map, XML and RDF provided multiple choices besides relational database since 1998 and 2000 respectively.

In the first phase of this research, XML is the major syntax for constructing project-based knowledge map. Also, the flexible DTD provides many advantages for verifying the framework. Afterwards, upon the growth of knowledge map and

various context perspectives, the semantic definitions in XML become uncontrollable and inconsistent. Therefore, RDF is further employed for enhancing the semantic structure in this work. Also, semantics in XML tree is reviews below for verifying the role of RDF in the project-based knowledge map.

XML is formalism for defining a grammar, so anything can be encoded in XML if a grammar can be defined for it. The is the major limitation of XML; since XML just describes grammars there is no way of recognizing a semantic unit form a particular domain of interest. Therefore, when it comes to semantic interoperability, XML has disadvantages.

Conversely, RDF is more than XML based on a mathematical model which defines relationships. RDF offers a nested object-attribute-value structure for users to better describe the content and context relationships. Moreover, RDF model consists of schemas, components, statements, containers, statements about statements, as well as XML namespaces. The long-term goal of RDF is to link different applications and Web resources into a new global network [42][64].

In our early research, we intend to define the association name in XML custom tags. However, the adjustable DTD increases the complexity and variety of the definition declaration. The disadvantage increases the difficulty in managing the consistence and interoperation of the knowledge map is limited by the local definitions.

Next, the implementation of the binary association becomes instinctive in RDF triple model. With the help of RDF graph, the association is visible and sensible, as shown in Figure 23.

```
<rdf:Description rdf:about="#XML">
        <pkmap:work_with rdf:resource="#HTML"/>
  </rdf:Description>
</rdf:RDF>
```

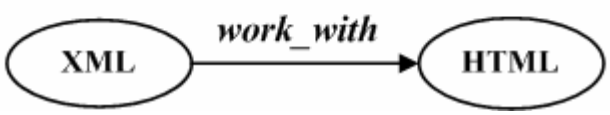


Figure 23. RDF triple syntax and the corresponding graph

RDF provides some standard ways of writing statement and produces the same effect in RDF terms. However, the same conceptual tree results from many XML trees defined in DTD. Then, many XML tree structures will easily result in many disagreements, as shown in Table 17. The drawback inevitably increases the difficulty of further rule transformation and inference.

Table 17. Various XML tree syntax for the statement

Defined DTD	XML Instance
<!ELEMENT ProjectMember(work_with)> <!ATTLIST ProjectMember id ID #REQUIRED> <!ELEMENT work_with (Project)> <!ATTLIST Project id ID #IMPLIED>	<ProjectMember id="XML"> <work_with> <Project id="HTML"/> </work_with> <ProjectMember>
<!ELEMENT work_with (PurchaseOrder, Project)> <!ELEMENT ProjectMember(#CDATA)> <!ELEMENT Project (#CDATA)>	<work_with> <ProjectMember>XML </ ProjectMember> <Project>HTML</Project> </work_with>
<!ELEMENT ProjectMember(id, work_with)> <!ELEMENT id (#CDATA)> <!ELEMENT work_with (Project)> <!ELEMENT Project (id)>	< ProjectMember> <id>XML</id> <work_with> <Project> <id>HTML</id> </Project> </work_with> </ ProjectMember >
<!ELEMENT rel EMPTY> <!ATTLIST rel src CDATA #REQUIRED type CDATA #REQUIRED dest CDATA #REQUIRED>	<rel src="XML" type="work_with" dest="HTML"/>

The rule statement is very important for increasing the semantic support of project-based knowledge map. With the help of RDF, formulating rule

transformation becomes efficiently and consistently. We explain the rule transformation and the semantics in RDF.

```
<?xml version="1.0"?>
      <rdf:RDF xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
          xmlns:pkmap = "http://www.nctu.iim.tw/KM/pkmap#"
          xml:base = "http://www.nctu.iim.tw/KM/pkmap/ptopics.xml">

   <rdf:Description rdf:about="#John Smith">
      <pkmap:engage_in rdf:resource="#CRM" />
  </rdf:Description>
<rdf:Description rdf:about="#John Smith">
   <pkmap:work_with rdf:resource="#Mary Brown"/>
     </rdf:Description>
</rdf:RDF>
```

Figure 24. RDF syntax for associations

Moreover, Rule transformation takes the advantage of statements and basic triple structure in RDF model. An excerpt of RDF syntax for two extracted associations is given in Figure 24. Therefore, we can simply insert the topic names and association names into rule statements in the position of subject, predicate and object, respectively.

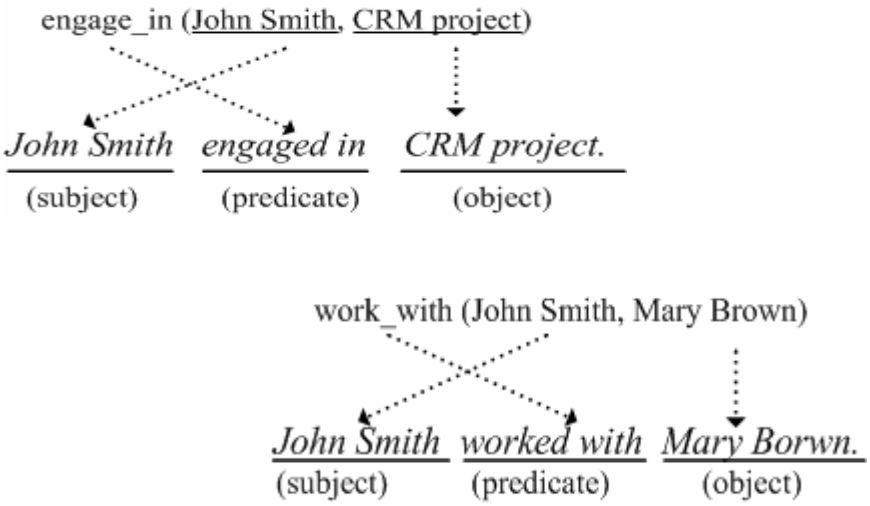


Figure 25. The mapping between associations and rule statements

As shown in Figure 25 the subject, predicate and object are naturally inherent from the association. These advantages enable RDF to represent simple statements about resources as a graph of nodes and arcs. With the help of statements, users can understand the project experience easily and establish the conceptual union of two rules, as shown in Figure 26.

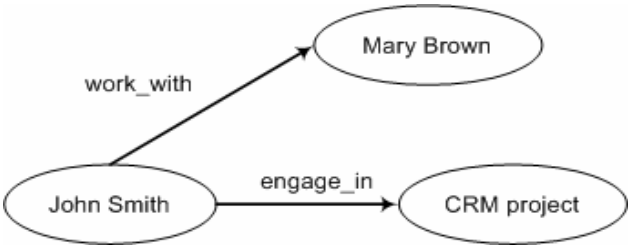


Figure 26. The conceptual union of two rules in RDF graph

6.2.1. Various forms of binary associations

Apriori association mining is employed for discovering frequent itemsets to generate association rules which satisfy minimum support and confidence. Accordingly, the corresponding RDF syntax is generated for consistently describing one-to-one binary, one-to-many binary and many-to-one binary associations. Several types of binary associations are found and shown in the following examples.

Example 7.1. Suppose the frequent itemset contain 2-itemsets: {T1, T2}. The possible resulting association rules are listed as following.

$T1 \Rightarrow T2$
$T2 \Rightarrow T1$

Example 7.2. Suppose the frequent itemset contain 3-itemsets: {T1, T2, T3}. The possible resulting associations are listed as following.

$T1 \Rightarrow T2 \wedge T3$
$T2 \Rightarrow T1 \wedge T3$
$T3 \Rightarrow T1 \wedge T2$
$T1 \wedge T2 \Rightarrow T3$
$T1 \wedge T3 \Rightarrow T2$
$T2 \wedge T3 \Rightarrow T1$

Example 7.3. Suppose the frequent itemset contain 4-itemsets {T1, T2, T3, T4}. The possible resulting associations are listed as following.

T1 ⇒ T2∧T3∧T4
T2 ⇒ T1∧T3∧T4
T3 ⇒ T1∧T2∧T4
T4 ⇒ T1∧T1∧T3
T1∧T2⇒ T3∧T4
T1∧T3⇒ T2∧T4
T1∧T4⇒ T2∧T3
T2∧T3⇒ T1∧T4
T2∧T4⇒ T1∧T3
T3∧T4⇒ T1∧T2
T2∧T3∧T4⇒T1
T1∧T3∧T4⇒T2
T1∧T2∧T4⇒T3
T1∧T1∧T3⇒T4

Basically, four types of binary associations are concluded from above examples, as shown in Table 18. If the frequent itemset contains k-itemsets and k is greater than 4, the possible resulting association types are also included in the table. However, one-to-one, one-to-many, many-to-many are three acceptable binary relations in this work, since these binary associations can provide sufficient relationship for a particular topic. The advantage is very important in the project-based knowledge map. However, composite-to-composite binary relations that complicate the rules and probably overwhelm the inference operations are not acceptable in this work. Particularly RDF can offer various grammars of bags (bag), sequences (seq) and alternatives (alt), shown as below, for describing different aggregation of topics and context.

- Bag is an unordered list of resources or literals.
- Sequence is an ordered list of resources or literals.
- Alternative is a list of resources or literals that represent alternatives for the value of a property.

Table 18. The acceptable binary types

Binary Type	Example	Acceptable
One-to-one	T1⇒T2 T2⇒T1	**Yes**
One-to-many	T1 ⇒ T2∧T3 T2 ⇒ T1∧T3 T3 ⇒ T1∧T2 T1 ⇒ T2∧T3∧T4 T2 ⇒ T1∧T3∧T4 T3 ⇒ T1∧T2∧T4 T4 ⇒ T1∧T1∧T3	**Yes**
Many-to-one	T1 ∧ T2⇒ T3 T1 ∧ T3⇒ T2 T2 ∧ T3⇒ T1 T2∧T3∧T4⇒T1 T1∧T3∧T4⇒T2 T1∧T2∧T4⇒T3 T1∧T1∧T3⇒T4	**Yes**
Many-to-Many	T1∧T2⇒ T3∧T4 T1∧T3⇒ T2∧T4 T1∧T4⇒ T2∧T3 T2∧T3⇒ T1∧T4 T2∧T4⇒ T1∧T3 T3∧T4⇒ T1∧T2	***No***

Situation 7.1. Let (John, Java, DBMS) be one of the extracted frequent itemsets and the resulting rule of (John Smith ⇒ Java ∧ DBMS) is a strong rule. Then grammar <Bag> in RDF model is helpful for grouping topics of Java and DBMS as composite for associating with the topic John Smith together. The valid RDF syntax is given Figure 27, and the corresponding RDF graph is shown in Figure 28. The blank node in the RDF graph means a RDF bag container for covering a set of relevant topic names.

```
<?xml version="1.0"?>
      <rdf:RDF xmlns:rdf =
"http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:pkmap = "http://www.nctu.iim.tw/KM/pkmap#"
           xml:base = "http://www.nctu.iim.tw/KM/pkmap/ptopics.xml">

   <rdf:Description rdf:about="#John Smith">
      <pkmap:engage_in>
         <rdf:Bag>
                  <rdf:li rdf:resource = "#Java" />
                  <rdf:li rdf:resource = "#DBMS" />
         </rdf:Bag>
      </pkmap:engage_in>
    </rdf:Description>
</rdf:RDF>
```

Figure 27. The RDF syntax for one-to-many binary associations

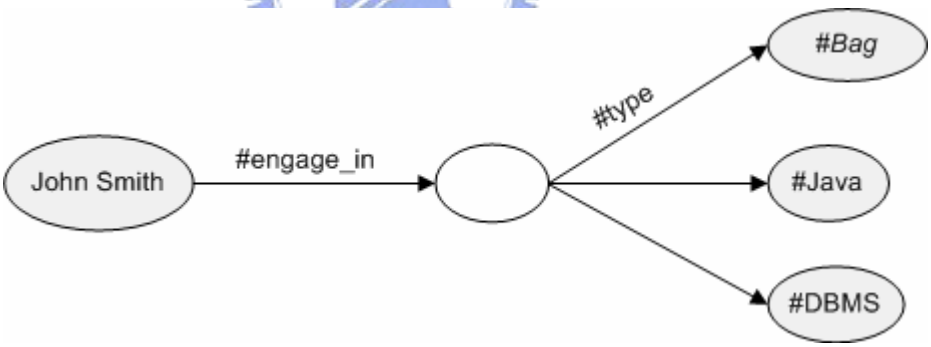


Figure 28. The corresponding RDF graph for one-to-many associations

Situation 7.2. Let (ISO, XML, KM) be an extracted frequent itemset and the resulting rule of (ISO ∧ XML ⇒ KM) is a strong rule. Then grammar <Bag> in RDF model is helpful for grouping topics of ISO and XML as composite for associating with the topic KM together. The RDF syntax is given in Figure 29, and corresponding RDF graph is given in Figure 30. The blank node in the RDF graph means the bag container for covering a set of relevant topic names.

```
<?xml version="1.0"?>
      <rdf:RDF xmlns:rdf =
"http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:pkmap = "http://www.nctu.iim.tw/KM/pkmap#"
           xml:base = "http://www.nctu.iim.tw/KM/pkmap/ptopics.xml">
         <rdf:Description>
         <pkmap:group>
         <rdf:Bag>
                  <rdf:li rdf:resource = "#ISO" />
                  <rdf:li rdf:resource = "#XML" />
         </rdf:Bag>
         </pkmap:group>
        <pkmap:engage_in rdf:resource ="#KM"/>
        </rdf:Description>
</rdf:RDF>
```

Figure 29. The RDF syntax for many-to-one binary association

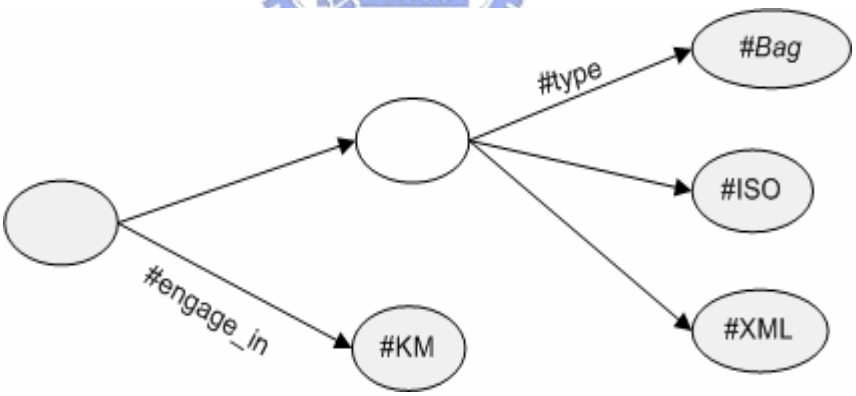


Figure 30. The corresponding RDF graph for many-to-one associations

6.3. The project description in RDF/XML

6.3.1.The implementation for a project and project objects

RDF model is naturally suitable for implementing the relationship between project and project objects. For example, project DWProject contains many project objects, including PA001.pdf, PB011.doc, PD070.pdf, PP071.class, PP082.java, and DW_Server. As shown in Figure 31, the RDF syntax indicates the complicated one-to-many relationship between the project and the relevant project objects, and the corresponding RDF graph is shown in Figure 32. The genid node in the RDF graph means the bag container for covering different project objects [57].

The inherent RDF model is advantageous for representing the belonging relationship between multiple project objects in a project. Also, the resource URL provides the timely links for accessing the objects.

```
<?xml version="1.0"?>
      <rdf:RDF xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
           xmlns:pkmap = "http://www.nctu.iim.tw/KM/pkmap#"
           xml:base = "http://www.nctu.iim.tw/KM/pkmap/ptopics.xml">

<rdf:Description rdf:about="http://www.abcd.com.tw/DWProject">
 <pkmap:projectobjects>
      <rdf:Bag>
          <rdf:li   rdf:resource= "http://www.nctu.iim.tw/PJS/PA/PA001.pdf" />
          <rdf:li   rdf:resource= "http://www.nctu.iim.tw/PJS/PA/PB011.doc"/>
          <rdf:li   rdf:resource= "http://www.nctu.iim.tw/PJS/PA/PD070.pdf"/>
           <rdf:li   rdf:resource= "http://www.nctu.iim.tw/PJS/PA/PP071.class"/>
           <rdf:li   rdf:resource= "http://www.nctu.iim.tw/PJS/PA/PP072.java"/>
           <rdf:li   rdf:resource= "http://www.nctu.iim.tw/PJS/PA/DW_Server"/>
      </rdf :Bag>
  </pkmap:projectobjects>
</rdf:Description>
</rdf:RDF>
```

Figure 31. The RDF syntax for project and project objects

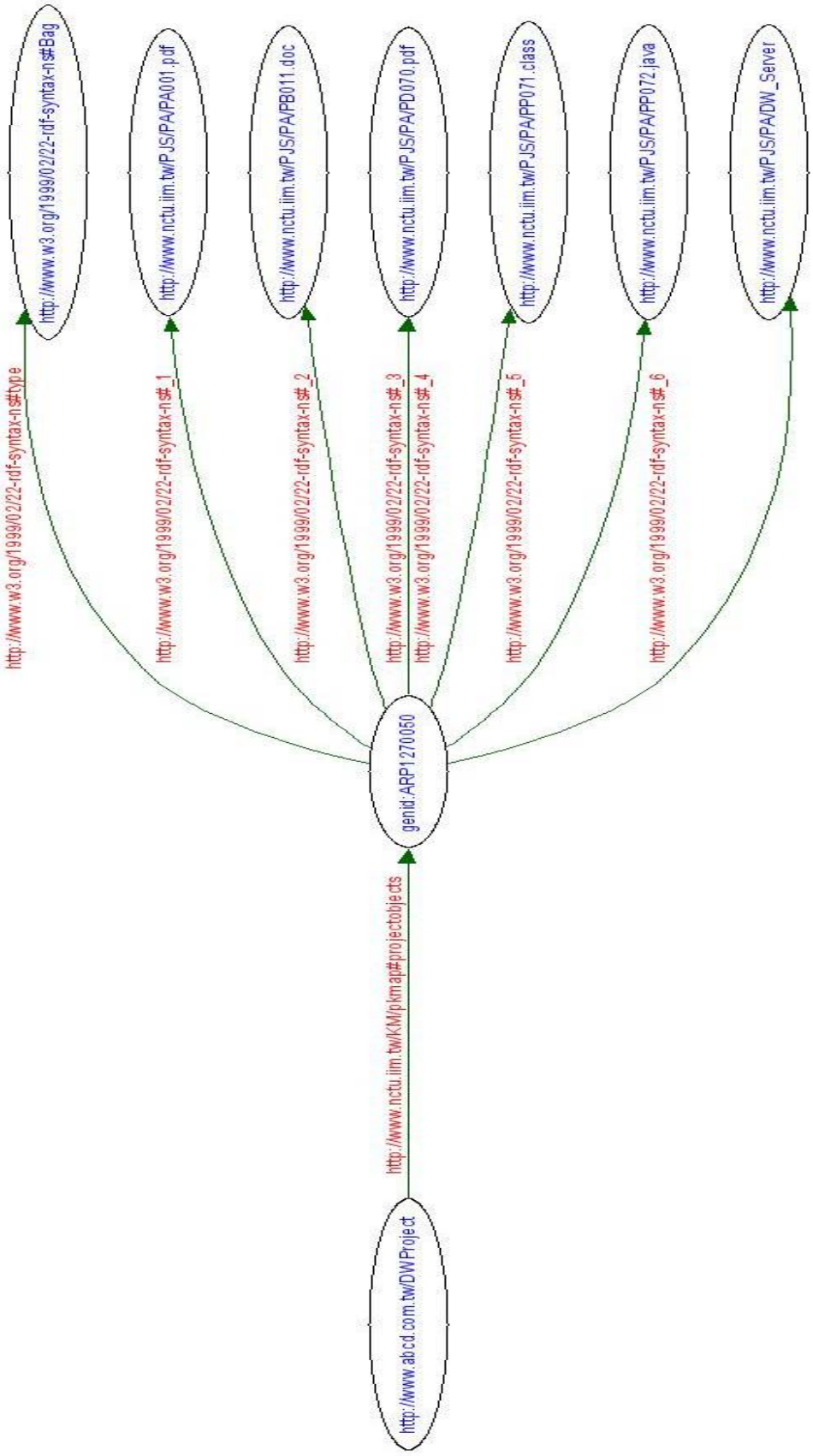


Figure 32. The corresponding RDF graph

6.3.2.The implementation for project attributes of a project object

As describing attribute details of a project object, RDF is also suitable for explaining the meaningful topic names and the belonging category name. For example, a project object PP072.java is describing with the following topic names, including the Member of John Smith and David Campbell, the Tool of Java and JDBC, the Standard of ISO, and the Goal of CRM. Therefore, the RDF syntax is applied for explaining the dependence and shown in Figure 33. Also, the corresponding RDF graph is given in Figure 34 [57]. As describing the project attributes for a project objects, the formal topic names with proper category name are helpful for illustrating the important features in a consistent and agreed manner. Simply, RDF model is advantageous for representing the implementation for project attributes of project objects.

As a result, the structure of the project-based knowledge map is naturally implemented in RDF model and the corresponding RDF graph explained the conceptual outcome for users. The semantic evolution of project-based knowledge map is promising in RDF technology. The consequence improves the evolution of project-based knowledge map in the system and the dissemination of project knowledge extensively over the Internet.

```
<?xml version="1.0"?>
    <rdf:RDF xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
        xmlns:pkmap = "http://www.nctu.iim.tw/KM/pkmap#"
        xml:base =
            "http://www.nctu.iim.tw/KM/pkmap/ptopics.xml">

 <rdf:Description rdf:about="http://www.nctu.iim.tw/PJS/PA/PP072.java" >
      <pkmap:Member    rdf:resource= "#John Smith" />
      <pkmap:Member    rdf:resource= "#David Campbell" />
      <pkmap:Tool      rdf:resource= "#Java " />
      <pkmap:Tool      rdf:resource= "#JDBC" />
      <pkmap:Standard  rdf:resource= "#ISO" />
      <pkmap:Goal      rdf:resource= "#CRM" />
 </rdf:Description>
```

Figure 33. The RDF syntax for project attributes

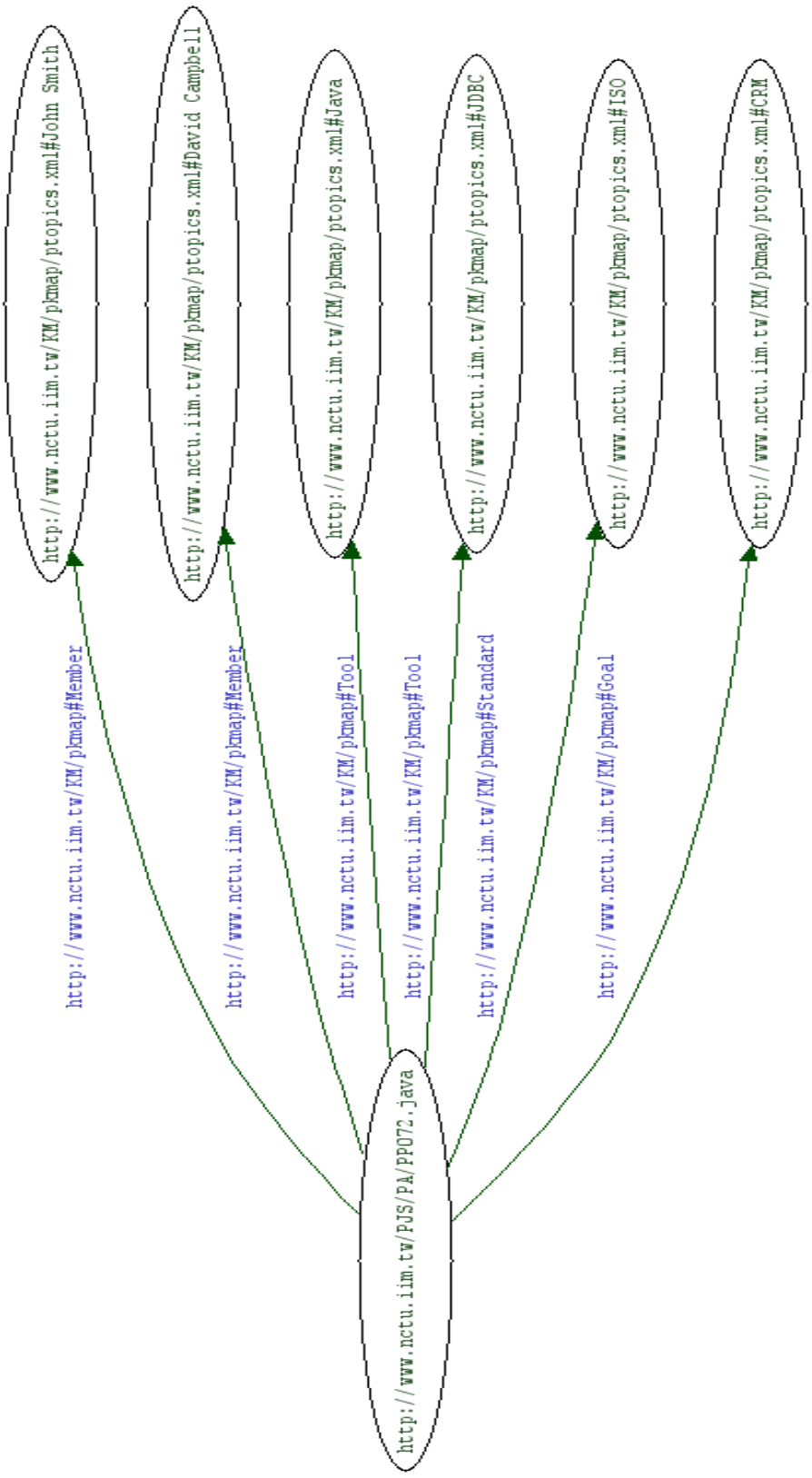


Figure 34. The corresponding RDF graph

Chapter 7. Conclusions and Future Works

7.1. Summary

The framework of the project-based knowledge map presents an important advancement in unifying project resources, discovering useful association patterns dependent on context constraints and systemizing all project knowledge into a meaningful structure to human users. Notably, context information deliberately increases the flexible description from the operation of integrating project resources up way to representing the project-based knowledge map. Furthermore, the semantic evolution of the system is particularly developed in RDF/XML technology for extending the practicability of the project-based knowledge map on the web service in the near future. Accordingly, many advantages and advancements are conveyed in this research. Summaries are herein explained from the aspects of knowledge management and information technology.

In the aspect of knowledge management, the proposed framework captured and organized the project knowledge from historical projects extensively. The intelligent assets are purposefully accumulated as developing the project-based knowledge map in an organization. Therefore, many advantages facilitate the performance of knowledge management as follows.

- Topic names are useful portals of the project-based knowledge map since various topic names are associated with different accessible links to relevant project resources. Not only providing the important topics, but also offering the links for accessing relevant references. Users can learn and understand the project knowledge effectively.
- Association rules are consistently converted into natural language statements for efficient knowledge support. Previous project experiences are extracted and expressed in the rule statements for users to easily understand and apply the associations in further project development.
- Project context notably improves the understanding of knowledge representation. The timely annotation actively leads users to know and

to apply the project knowledge easily and efficiently.

- Project context facilitates knowledge filtering as interacting with users for constructing the user-dependent project-based knowledge map.
- Project context facilitates the performance of knowledge discovery. With complete and flexible descriptions from project resources, more data mining methods are therefore workable in the knowledge discovery operation from different perspectives.
- Context-relevant and context-specific associations offer general and particular rules for supporting various user situations.

In the aspect of information technology, the framework of project-based knowledge map considers the extensive support and for human users and various systems on the network. Topic Maps, data mining methods, web-based system and RDF/XML work collaboratively for developing project-based knowledge map. Furthermore, the proposed project context promotes the influence of information technology in this work as follows.

- The instinctive structure of the project-based knowledge map in Topic Maps is interoperable and exchangeable over the Internet.
- Data mining is highly effective in discovering knowledge, but contributes less to representing knowledge. The project-based knowledge map employs Topic Maps to compensate for this shortcoming.
- Context enhancing the semantic links between topic names and occurrences improves the expression facility of Topic Maps.
- Constraint-based data mining in various context perspectives is helpful for facilitating the user-dependent knowledge support.
- RDF/XML is typically applied for the semantic transformation of project-based knowledge map. Therefore, the evolution of the project-based knowledge map facilitates the dissemination of project knowledge across different applications over the network.

7.2. Future works

The primitive system illustrates the efficiency of the framework of project-based knowledge map. Surely, we also find some challenges during the experimental process. Therefore, many research issues are continuously proposed for improving the performance of the framework of project-based knowledge map, including semantic evolution, system advancement and user interaction.

Semantic evolution is on going for developing the project-based knowledge map on the web service. RDF and OWL are Semantic Web standards that provide a framework for asset management, enterprise integration and the sharing and reuse of data on the Web. These standard formats for data sharing span application, enterprise, and community boundaries - all of these different types of "user" can share the same information, even if they don't share the same software. Today, the World Wide Web Consortium announced final approval of two key Semantic Web technologies, the revised RDF and OWL (Web Ontology Language). Basically, the primitive system was first developed in 2003, and the semantic evolution is successively developed in RDF/XML since 2005. The effort on semantic evolution is evident to disseminate the project-based knowledge map extensively across various systems on the Internet.

System advancement is an interesting challenge to extend the performance of the framework of project-based knowledge map. With the benefit of project context, constraint-based data mining is therefore effective from various perspectives. Thus, the employment of various data mining methods for extracting useful knowledge patterns is valuable and creative effort in this research. The advantage can raise the significance of the project-based knowledge map. Also, various evaluation measures for data mining algorithms are continuously proposed to support user decision making. Furthermore, the manipulation of comparison, synthesis or coordination among project-based knowledge maps is arranged in the future. The advantage is helpful for integrating the project knowledge to accumulate intelligent assets extensively.

User interaction is important for increasing the communication and understanding in the system to perceive user information needs. User behavior and requirement shown in user interaction is certainly important in the knowledge

management system. Beyond the representation of traditional keywords, the proposed project context is helpful for summarizing user selection for interaction. The advantage attracts users to use the system frequently. However, the simple, novel and useful interaction for user context is always inviting in the continuous research.

Also, user interaction is helpful for processing user evaluation which is valuable for the improvement of this research work. In the real word, interviews and questionnaires are mostly useful for perceiving users' opinions to adjust the data mining algorithms and user interaction in the system. Primarily, both of the suggestions from end users and domain experts are valuable reference in this research toward the promotion of project knowledge in the framework of project-based knowledge map.

References

[1] Agrawal, R., T. Imielinski, T. and Swami, A.N. (1993), Mining association rules between sets of items in large databases, *Proceedings of the ACM SIGMOD International Conference on Management of Data*, IGMOD 93, pp. 207-216.

[2] Allen, B. and Kim, K.-S. (2001), Person and context in information seeking: interactions between cognitive and task variables, *The New Review of Information Behavior Research: Studies of Information Seeking in Context*, vol. 2, pp.1-16.

[3] Baeza-Yates, R. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval. MA*, USA: Addison-Wesley.

[4] Barthès, J.-P. A. and Tacla, C. A. (2002), Agent-supported portals and knowledge management in complex R&D projects, Computers in Industry, Vol. 48 No. 1, pp. 3-16.

[5] Beckett, D. (2001), The design and implementation of the Redland RDF application framework, *Proceedings of WWWIO conference*.

[6] Berry, M.J. A., and Linoff, G. (1997). *Data mining techniques: for marketing, sales, and customer support*, USA: John Wiley and Sons.

[7] Bolloju, N, Khalifa, M. and Turban, E. (2002), Integrating knowledge management into enterprise environments for the next generation decision support, *Decision Support Systems*, Vol. 3 No.2, pp.163-176.

[8] Browne, G.J., Curley S.P. and Benson, P.G. (1997), Evoking Information in Probability Assessment: Knowledge Maps and Reasoning-Based Directed Questions, *Management Science*, Vol.43, No.2, pp.1-14.

[9] Cadastre Reengineering Project, http://www.ctg.albany.edu/publications/online/new_models/cases/cadastre

[10] Candan, K. S., Liu, H., and Suvarna, R. (2001). Resource Description Framework: Metadata and Its Applications. *ACM SIGKDD Exploration Newsletter*, Vol.3, pp.6–19.

[11] Chung, W., Chen, H., and Nunamaker, J. (2003), Business Intelligence Explorer: A knowledge map framework for discovering business intelligence on the Web, *Proceedings of the 36th Hawaii International Conference on System Sciences*, HICSS 2003.

[12] Czuchry, A.J. and Yasin, M.M. (2003), Managing the project management process, *Industrial Management & Data Systems*, Vol.103 No 1, pp. 39-46.

[13] Davenport, T., and Prusak, L. (2000), *Working Knowledge: How Organizations Manage What They Know*, USA: Harvard Business School Press.

[14] Deng, Z.M., Li, H., Tam, C.M., Shen, Q.P. and Love, P.E.D. (2001), An application of the Internet-based project management system, *Automation in Construction*, Vol. 10 No. 2, pp.

[15] 239-246.

[16] Dervin, B., Harlock, S., Atwood, R., and Garzona, C. (1980), The human side of information: An exploration in a health communication context, *Communication yearbook ,USA: Transaction Books*, Vol.4, pp.591–608.

[17] Dey, A.K., and Abowd, G.D. (1999). Towards a Better Understanding of Context and Context-Awareness, *Technical Report GIT-GVU-99-22. College of Computing*, Georgia Institute of Technology, USA.

[18] Fayyad, U.M., Piatetsky-Shapiro, G., and Smyth, P. (1996), From Data Mining to Knowledge Discovery: An Overview, A*dvances in Knowledge Discovery and Data Mining*, pp.1-34.

[19] Fischer, G. and Ostwald, J. (2001), Knowledge management: Problems, promises, realities, and challenges, *IEEE Intelligent systems*, Vol.16, No.1, pp. 60-72.

[20] International Organization for Standardization, ISO/IEC 13250, Information technology SGML Applications--- Topic Maps, ISO, Geneva (2000).

[21] Han, J., and Kamber, M. (2000), *Data Mining: concepts and technique*, USA: Morgan Kaufmann.

[22] IEEE Std. 1058-1998 (1998), IEEE Standard for Software Project Management Plans (SPMP), Institute of Electrical and Electronics Engineers, Inc., NJ.

[23] IsaViz, http://www.w3.org/2001/11/IsaViz/

[24] Jain, A. K., and Dubes, R. C. (1988), *Algorithms for Clustering Data*, USA: Prentice Hall, Englewood Cliffs.

[25] Johnson, J. D. (2003), On contexts of information seeking, *Information Processing and Management*, Vol.39, No.5, pp.735-760.

[26] Kim, S., Suh, E., and Hwang, H. (2003), Building the Knowledge Map: An Industrial case study, *Journal of Knowledge Management*, Vol.7, No.2, pp.34-45.

[27] KPMG Consulting (2003), *Insights from KPMG's European Knowledge Management Survey 2002/2003*, London, UK.

[28] Lawrence, S. (2000), Context in Web Search, *IEEE Data Engineering*, Vol.23, No.3, pp.25-32.

[29] Leathrum, J., González, O., Zahorian, S. and Lakdawala, V. (2001), Knowledge Maps for Intelligent Questioning Systems in Engineering Education, *Proceedings of the 2001 American Society for Engineering Education Annual Conference & Exposition.*

[30] Li, H., Tang, S., Man, K.F. and Love, P.E.D. (2002), VHBuild.com: a Web-based system for managing knowledge in projects, *Internet Research:Electornic Networking Applications and Policy*, Vol.12, pp.371-37.

[31] Lin, F.-r., & Hsueh, C.-m. (2003), Knowledge Map Creation and Maintenance for Virtual Communities of Practice, *Proceeding of the 36th Annual Hawaii International Conference on System Sciences*, HICSS 2003.

[32] Liu, D.-r., and Hsu, C. (2004), Project-based knowledge maps: combining project mining and XML-enabled topic maps. *Internet Research: Electronic Networking Applications and Policy*, Vol.14, No.3, 254-266.

[33] Norman, K. L. (1991), The psychology of menu selection: designing cognitive control at the Human-Computer Interface, USA: Ablex publishers.

[34] Malhotra, Y. (1998), Deciphering the knowledge management hype, *Journal for Quality & Participation*, Vol.21, No.4, pp. 58-60.

[35] Martin, D., Birbeck, M., Kay, M., Loesgen, B., Pinnock, J., Livingstone, S., Stark, P., Williams, K., Anderson, R., Mohr, S., Baliles, D., Peat, B. and Ozu, N. (2001), *Professional XML 2nd Edition*, USA: Wrox Press.

[36] Massey, A.P., Montoya-Weiss, M. and Holcom, K. (2001), Re-engineering the customer relationship: leveraging knowledge assets at IBM, *Decision Support Systems*, Vol.32, No.2, pp.155-170.

[37] Ng, R., Lakshmanan, L. V. S., Han, J., and Pang. A. (1998), Exploratory mining and pruning optimizations of constrained associations rules, *Proceeding of ACM-SIGMOD'98 Conference on Management of Data*, Seattle, USA, pp.13–24.

[38] O' Leary, D.E. (1998), Using AI in Knowledge Management: Knowledge Bases and Ontologies, *IEEE Intelligent Systems*, Vol.13, No.3, pp.33-39.

[39] Piatetsky-Shapiro, G. (1991), Discovery, analysis, and presentation of strong rules, *Knowledge Discovery in Databases*, pp. 229-248.

[40] PMI standard (1996), *A Guide to the Project Management Body of Knowledge*, Project Management Institute, Inc, PA, USA.

[41] Rath, H.H., and Pepper, S. (1999), Topic Maps: Introduction and Allegro, *Proceedings of Markup Technologies 99 Conference*, VA, USA.

[42] RDF Primer, W3C Recommendation 10 February 2004, http://www.w3.org/TR/rdf-primer/.

[43] Schwartz, D.G. and Te'eni, D. (2000), Tying Knowledge to Action with kMail, *IEEE Intelligent Systems*, Vol. 15 No. 3, pp. 33-39.

[44] Steinbach, M., Karypis, G., and Kumar, V., (1999), A comparison of document clustering techniques, *KDD Workshop on Text Mining*.

[45] Rubenstein-Montano, B. (2000), A survey of knowledge-based information systems for urban planning: moving towards knowledge management, *Computers, Environment and Urban Systems*, Vol. 24.

[46] Rüger, S.M. and Gauch, S.E. (2000), Feature reduction for document clustering and classification, *Technical report Computing Department*, Imperial College, London, UK.

[47] Salminen, A., Lyytikäinen, V., and Tiitinen, P. (2000), Putting documents into their work context in document analysis, *Information Processing & Management*, Vol.36, No.4, pp.623-641.

[48] Schwartz, D.G., Divitini, M. and Brasethvik, T. (2000), On Knowledge Management in the Internet Age, *Internet-based Knowledge Management and Organizational Memories*, Schwartz, D.G., Divitini, M., Brasethvik, T. (Editors), Idea Group Publishing, pp.1-23.

[49] Schwotzer, T. (2002), Context Driven Spontaneous Knowledge Exchange, *German Workshop on Experience Management*, pp.131-138.

[50] Shneiderman, B. (1997), *Designing the User Interface: Strategies for Effective Human-computer Interaction*, Addision-Wesley, Reading, MA.

[51] Smyth, P., Goodman, R.M.(1992), An information theoretic approach to rule induction from databases, *IEEE Transaction on Knowledge and Data Engineering*, Vol. 4, pp.301-316.

[52] Tah, JHM. and Carr, V. (2001), Towards a framework for project risk knowledge management in the construction supply chain, *Advances in Engineering Software*, Vol.32, No.10-11, pp.835-846.

[53] Turner, J.R.(1992), *The Handbook of Project Based Management: Improving Processes for Achieving Your Strategic Objectives*.

[54] Vail, E.F. (1999), Mapping Organizational Knowledge. *Knowledge Management Review*, Vo.8, pp.10-15.

[55] Voida, S., Mynatt, E. D., MacIntyre, B., and Corso, G. M. (2002), Integrating virtual and physical context to support knowledge workers, *IEEE Pervasive Computing*, Vol.1, No.3, pp.73-79.

[56] Voorhees, E. (1986), Implementing agglomerative hierarchic clustering algorithms for use in document retrieval, *Information Processing and Management*, Vol.22, No.6, pp.465-476.

[57] W3C RDF Validation service: http://www.w3.org/RDF/Validator/direct

[58] Wang, S. (2002), Knowledge maps for managing Web-based business, *Industrial Management & Data Systems*, Vol.102, No.7, pp.357–364.

[59] Wang, S. and Ariguzo, G. (2004), Knowledge management through the development of information schema, I*nformation & Management*, Vol.41, pp. 445-456.

[60] Weka software, http://www.cs.waikato.ac.nz/ml/weka/

[61] West, L. A., and Hess, T. J. (2002), Metadata as a knowledge management tool: Supporting intelligent agent and end user access to spatial data, *Decision Support Systems,* Vol.32, No.3, pp.247– 264.

[62] Wideman, Max, Wideman Comparative Glossary of Common Project Management Terms, http://maxwideman.com/pmglossary/PMG_P10.htm

[63] Willett, P. (1998), Recent trends in hierarchic document clustering: a critical review, *Information Processing and Managemen*t, Vol.24, pp.577-597.

[64] XML (Extensible Markup Language 1.0) (Second Edition), W3C Recommendation, available at http://www.w3.org/TR/REC-xml REC-xml-20001006 (2000).

Appendix A. The system of the project-based knowledge map

The project-based knowledge map system is developed as the web-based system. As connecting the system web site, the proper user interaction is helpful for facilitating user navigation on the project-based knowledge map. First, user is required to choose the cluster number to view the project-based knowledge map. Then, the project-based knowledge map in the structure of Topic Maps is displayed for the user, as shown in Figure 35.

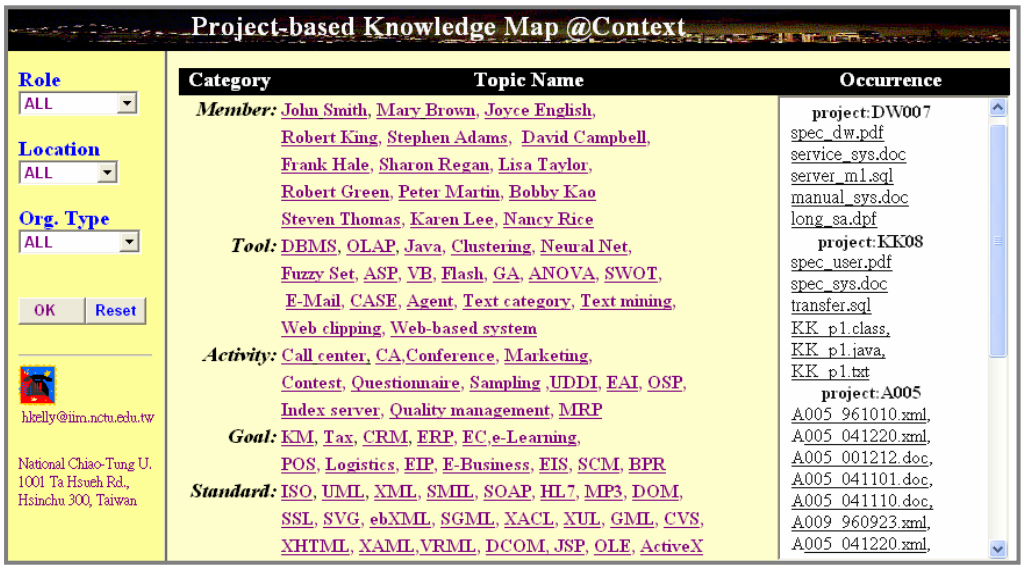


Figure 35. Displaying the project-based knowledge map

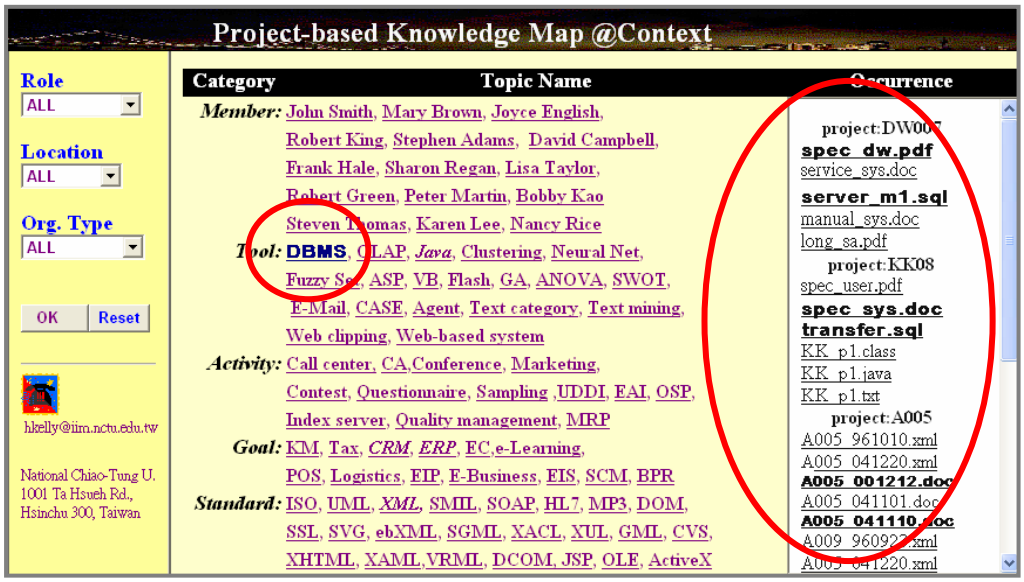


Figure 36. The selected topic name and the referable project occurrences

Next, as the user clicks the interesting topic, the referable project occurrences are intelligently separated for the user. As shown in Figure 36, the topic of DBMS is related to several project occurrences, such as spec_dw.pdf, server_m1.sql spec_sys.doc and transfer.sql. The useful hyperlinks lead users to access the file. For example, the user can open the file of spec_dw.pdf to learn the details of DBMS, as shown in Figure 37.

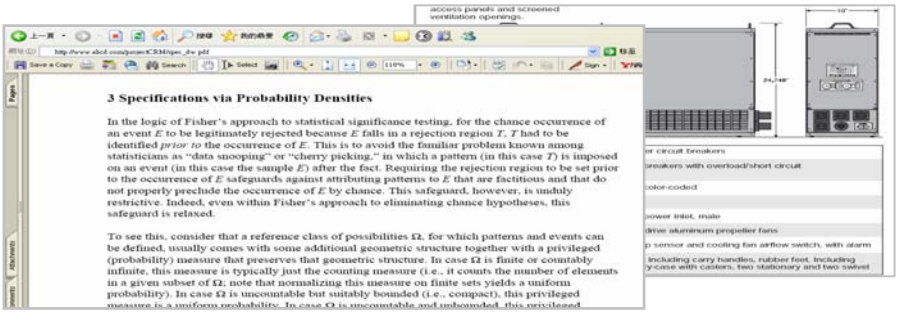


Figure 37. An example of referable project occurrences

However, each topic is usually connected with several project occurrences. Learning all topics and the details in the connecting project occurrences is really a heavy burden on users. Therefore, the selection menu is helpful for determining the relevant project knowledge. As shown in Figure 38, the result of the user selection is DBA, Taipei and Government, fourteen context-relevant topic names with referable project occurrences are further separated for users.

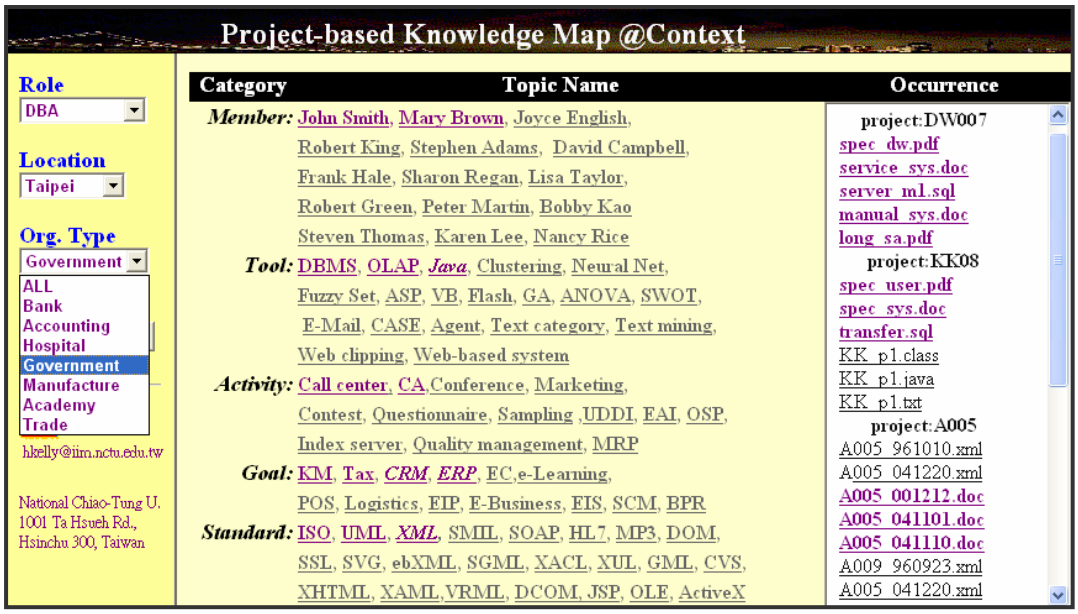


Figure 38. Displaying the context-relevant result for the first selection

Moreover, the user can indicate different context conditions to learn different part of project knowledge. As shown in Figure 39, the user selection indicates the role, location and organization type are consultant, New York and Bank, and different set of context-relevant topics names with referable project occurrences are separated for users. Furthermore, the rule statements with project context are useful for learning project knowledge. The screenshots are shown in Figure 19 and Figure 20 previously.

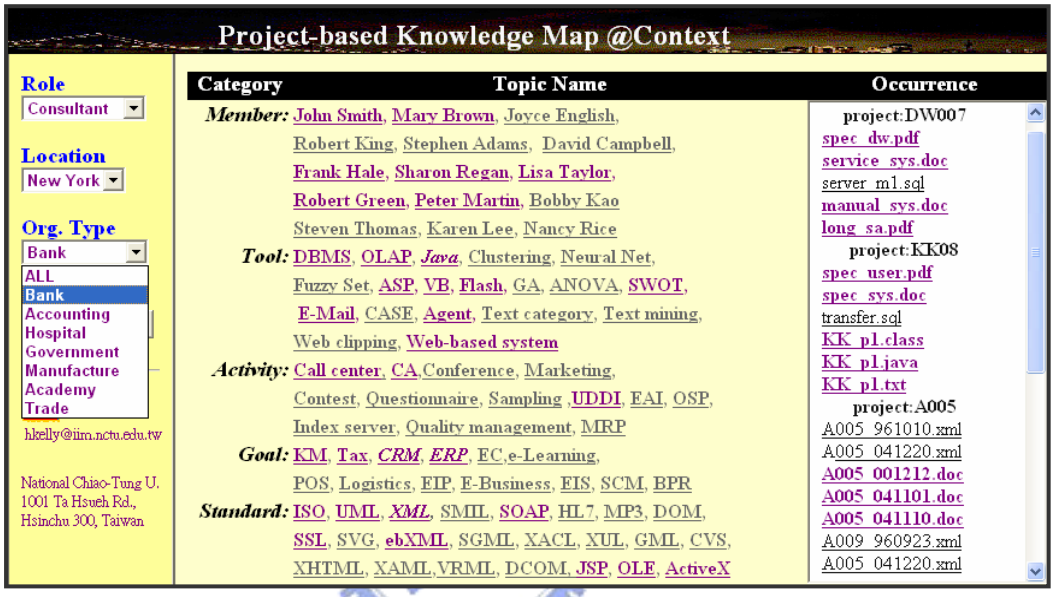


Figure 39. Displaying the context-relevant result for the second selection

Figure 40 shows the three-tier system architecture developed in this work. Moreover, the referential integrity of system tables is also given in Figure 41 for explaining the implementation of the project-based knowledge map.

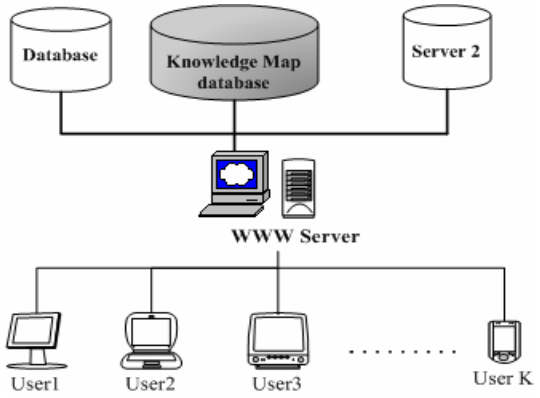


Figure 40. The three-tier system architecture

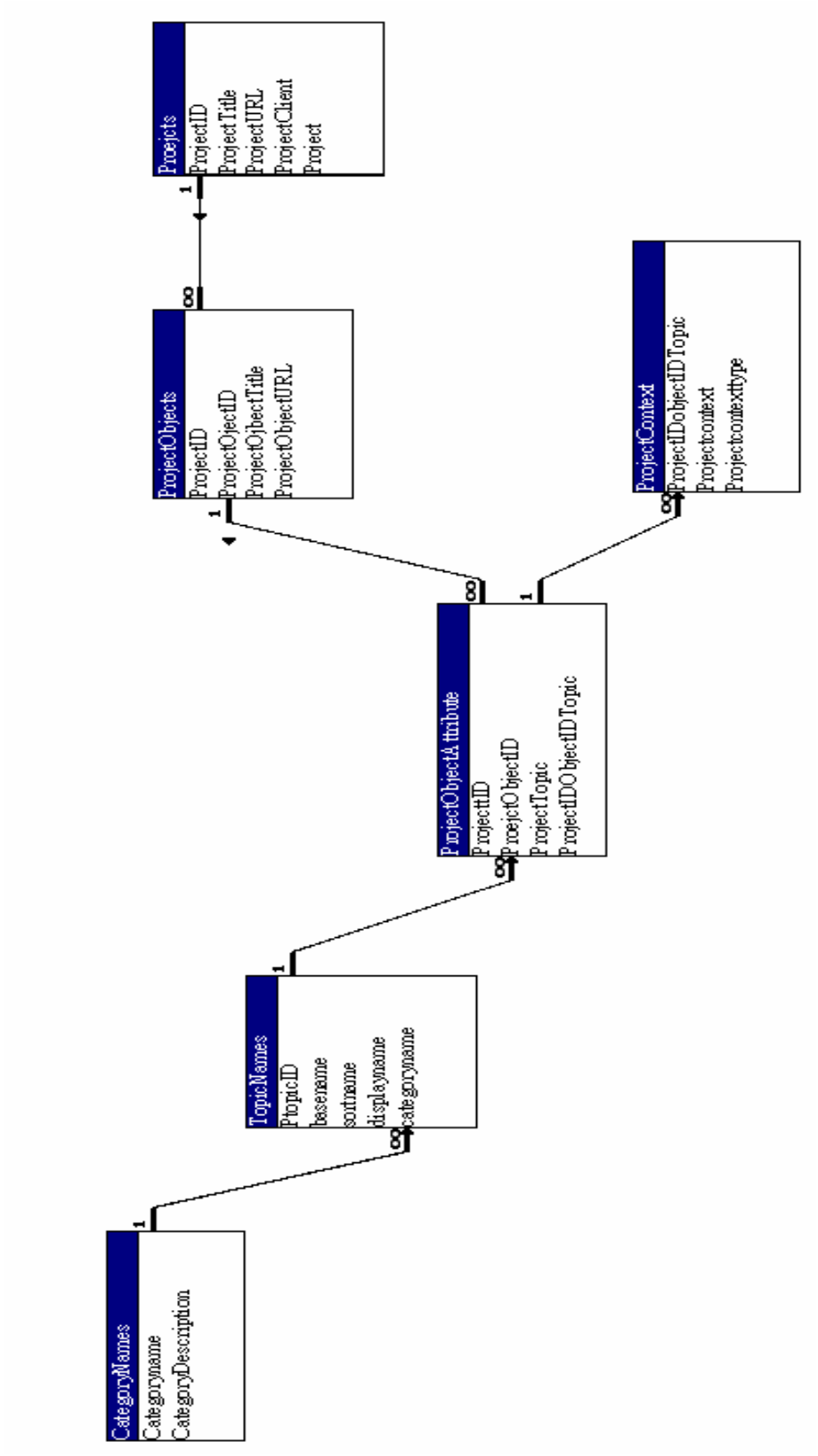


Figure 41. The referential integrity of the system tables

Appendix B. Evaluation on Data Mining Methods

B.1. Evaluation measures on association rule mining

Association rules mining aims to find frequent patterns, associations, correlations, or causal structures among sets of items or objects in transaction databases, relational databases, and other information repositories [1]. Applications include basket data analysis, cross-marketing, catalog design, clustering, and classification, etc.

Association rules are implications of the form:

$$\mathbf{A \rightarrow B}$$

where A and B are two disjoint subsets of all available items.

A is called the antecedent or LHS (left hand side)

B is called the consequent or RHS (right hand side).

When mining association rules, how can the data mining system tell which rules are likely to be interesting to the user? The answer is that only the user can judge if a given rule is interesting or not [21]. Support and confidence are usually applied in most research works. Therefore, the details are also discussed in the following paragraphs. However, many measures are helpful for choosing the 'interesting' association rules, as shown in Table 19.

Support is defined on itemsets and gives the proportion of transactions which contain antecedent and consequent. It is used as a measure of significance (importance) of an itemset. Since it basically uses the count of transactions it is often called a frequency constraint. Supports main feature is that it possesses the down-ward closure property (antimonotonicity) which means that all sub sets of a frequent set (support greater than minimun support threshold) are also frequent. This property (actually, the fact that no super set of an infrequent set can be frequent) is used to prune the search space in level-wise algorithms, such as Apriori algorithm. The disadvantage of support is the rare item problem. Items that occur very infrequently in the data set are pruned although they would still produce interesting and potentially valuable rules. The rare item problem is

important for transaction data which usually have a very uneven distribution of support for the individual items; that is, few items are used all the time and most item are rarely used.

Confidence is defined as the probability of seeing the consequent under the condition that the transactions also contain the antecedent. Confidence is directed and gives different values for the rules X $\rightarrow$Y and Y $\rightarrow$ X. Confidence is not down-ward closed and was developed together with support by Agrawal et al. (the so-called support-confidence framework) [1]. Support is first used to find frequent (significant) itemsets exploiting its down-ward closure property to prune the search space. Then confidence is used in a second step to produce rules from the frequent itemsets that exceed a minimum confidence threshold. A problem with confidence is that it is sensitive to the frequency of the consequent (B) in the database. Caused by the way confidence is calculated, consequents with higher support will automatically produce higher confidence values even if there exists no association between the items.

Table 19. Evaluation measures of association rules

Measures	Advantage
Support	the probability of A,B appearing together
Confidence	the probability of seeing the consequent under the condition that the transactions also contain the antecedent.
Coverage	shows that what part of itemsets from consequent is covered by a rule
Lift	shows proportion of conditional probability B (under condition of A) to unconditional probability of B.
Interest	the extension of lift measure.
Conviction	an alternative to confidence which was found to not capture direction of associations adequately.
Leverage (Piatetsky – Shapiro	how many more units (items A and B together) are sold than expected from the independent sells.
J – Measure	how much information is contained in rule.
Chi-square test	verify the independence or correlation of attributes.

B.2. Evaluation and distance measures on clustering

Several measures are used for computing the similarity or distance between two clusters. The similarity in a cluster of objects is typically computed based on the distance between each pair of objects. Given two feature vectors for project objects p1 and p2, we want to measure the similarity between them. Note that any distance measure can be converted into a similarity metric as follows:

$$similarity(p1, p2) = \frac{1}{1 + dis\tan ce(p1, p2)}$$

Therefore, distance measures are such important for computing the similarity. Euclidean distance is well-know. Also, cosine metric is commonly used for commonly used metric in natural language programming.

$$\cos(p1, p2) = \frac{p1 \bullet p2}{\|p2\| \|p2\|}$$

dot product of d1 and d2 is divided by the length of d1 and length of d2.

The cosine similarity measure is useful for indicating the degree of similarity in the range [0,1]. In a user query system, the advantage can rank the resultant documents even with the measure of partially matching with user query. However, similarity measure in agglomerative clustering is simply used for determining the elements and the corresponding clusters, instead of ranking the order. Therefore, Euclidean distance is simply selected in agglomerative clustering in many works since the Euclidean distance is related to the cosine similarity measure owing to the process of normalization [46].

Moreover, if p1 and p2 are two binary vectors which have only two states: 0 or 1, where 0 means that the variable is absent, and 1 means that it is present. Euclidean distance is one of common measures. Also, many well-known metrics are used in many applications, including Dice, Matching, Jaccard, Overlap and Cosine. Each of them has its own characteristics and limitations. The details of these measures are shown in Table 20.

Table 20. The distance measures for binary vector space model

Distance	Formula	Characteristics
Dice	$\frac{2\lvert A \cap B \rvert}{\lvert A \rvert + \lvert B \rvert}$	normalizes for length.
Overlap	$\frac{\lvert A \cap B \rvert}{\min(\lvert A \rvert, \lvert B \rvert)}$	measure of inclusion.
Jaccard	$\frac{\lvert A \cap B \rvert}{\lvert A \cup B \rvert}$	penalizes more than Dice when there are a small number of shared dimensions.
Matching	$\lvert A \cap B \rvert$	counts the dimensions that are non-zero (no normalization for length).
Cosine	$\frac{\lvert A \cap B \rvert}{\sqrt{\lvert A \rvert \times \lvert B \rvert}}$	same as Dice when vectors have same number of non-zero entries. But, it is less sensitive to 0's than other measures.

Next, evaluation of clustering is rather important for confirming the clustering results. Two widely used measures in text mining to evaluate the quality of the clustering algorithms are entropy and F-measure. Entropy measures the homogeneity of the clusters.

The lower the entropy is the more homogenous the clusters are:

$$E_j = -\sum_i p(i, j) \cdot \log P(i, j)$$

where P(i,j) is the probability that a document has class label i and is assigned to cluster j.

The total entropy for a set of clusters is obtained by summing the entropies of each cluster weighted by the size of each cluster:

$$E = \sum_j \frac{n_j}{n} E_j$$

where n_j is size of cluster j and n is total number of documents

F-measure cluster evaluation method combines the precision and recall ideas from information retrieval. In this technique, each cluster is considered as if it were the result of a query and each class as if it were the desired set of documents for the query.

Precision and recall for each cluster j and class i is calculated as follows:

$$recall\ (i, j) = \frac{n_{ij}}{n_i} \qquad precision\ (i, j) = \frac{n_{ij}}{nj}$$

where n_{ij} is the number of members of class *i* in cluster *j*, n_j is the number of members of cluster *j* and n_i is the number of members of class *i*.

The F measure of cluster *j* and class *i* is then given below.

$$F\ (i, j) = \frac{(2 * recall\ (i, j) * precision\ (i, j))}{precision\ (i, j) + recall\ (i, j)}$$

For an entire hierarchical clustering the F-measure of any class is the maximum value it attains at any node in the tree. An overall value for the F-measure is calculated by taking the weighted average of all values for the F-measure, and n is the number of document.

$$F = \sum_i \frac{n_i}{n} \max_j F\ (i, j)$$

Appendix C. The Development and Evaluation on RDF

C.1. Online RDF validation parser

Currently, RDF is suggested for the implementation of associations. Since the basic triple model formulated in RDF is worldwide agreed, the extended models and grammars support more operations. The advantage improves the further operation of incorporation, exchange and synthesis among knowledge maps. Currently, RDF Validation Parser and IsaViz RDF editor are employed in this work for RDF development and syntax validation. The former is to edit the RDF file for parser validation, and the latter is to draw RDF model for generating compatible RDF file.

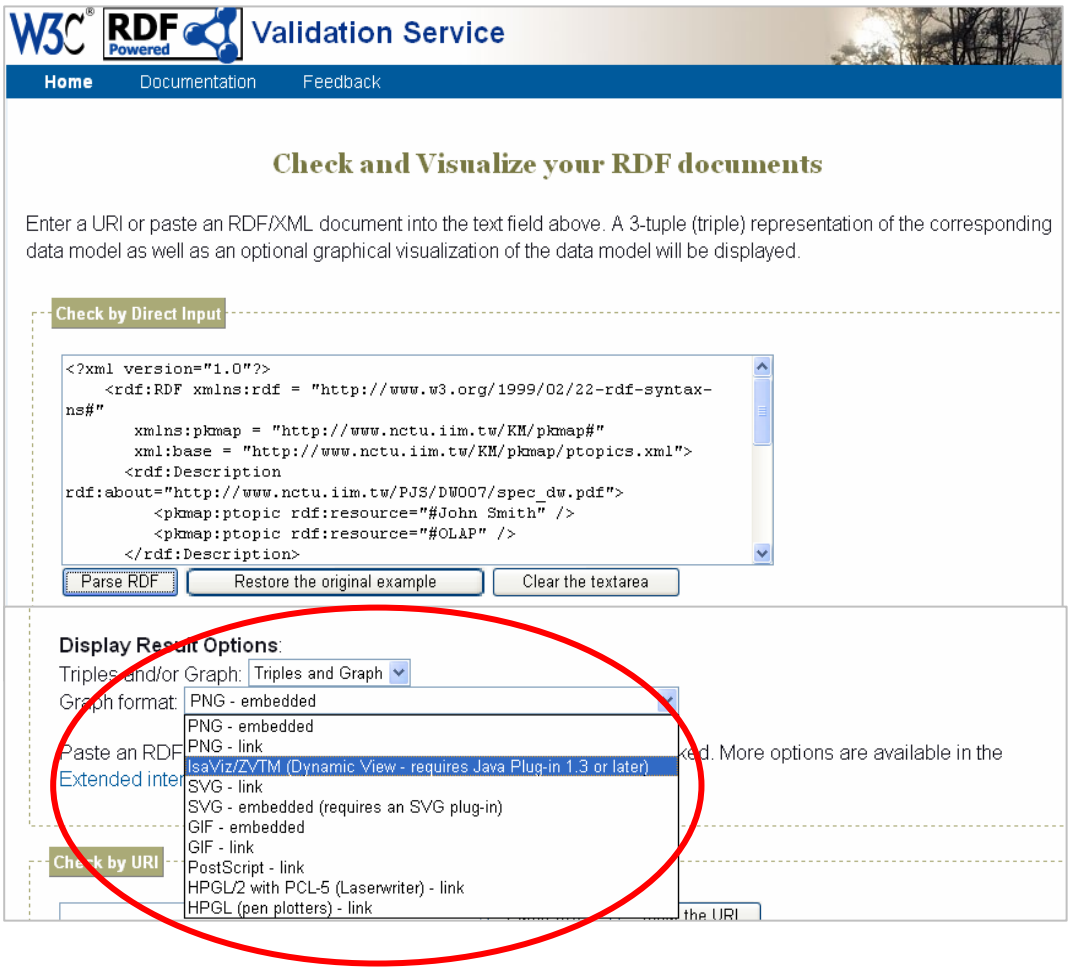
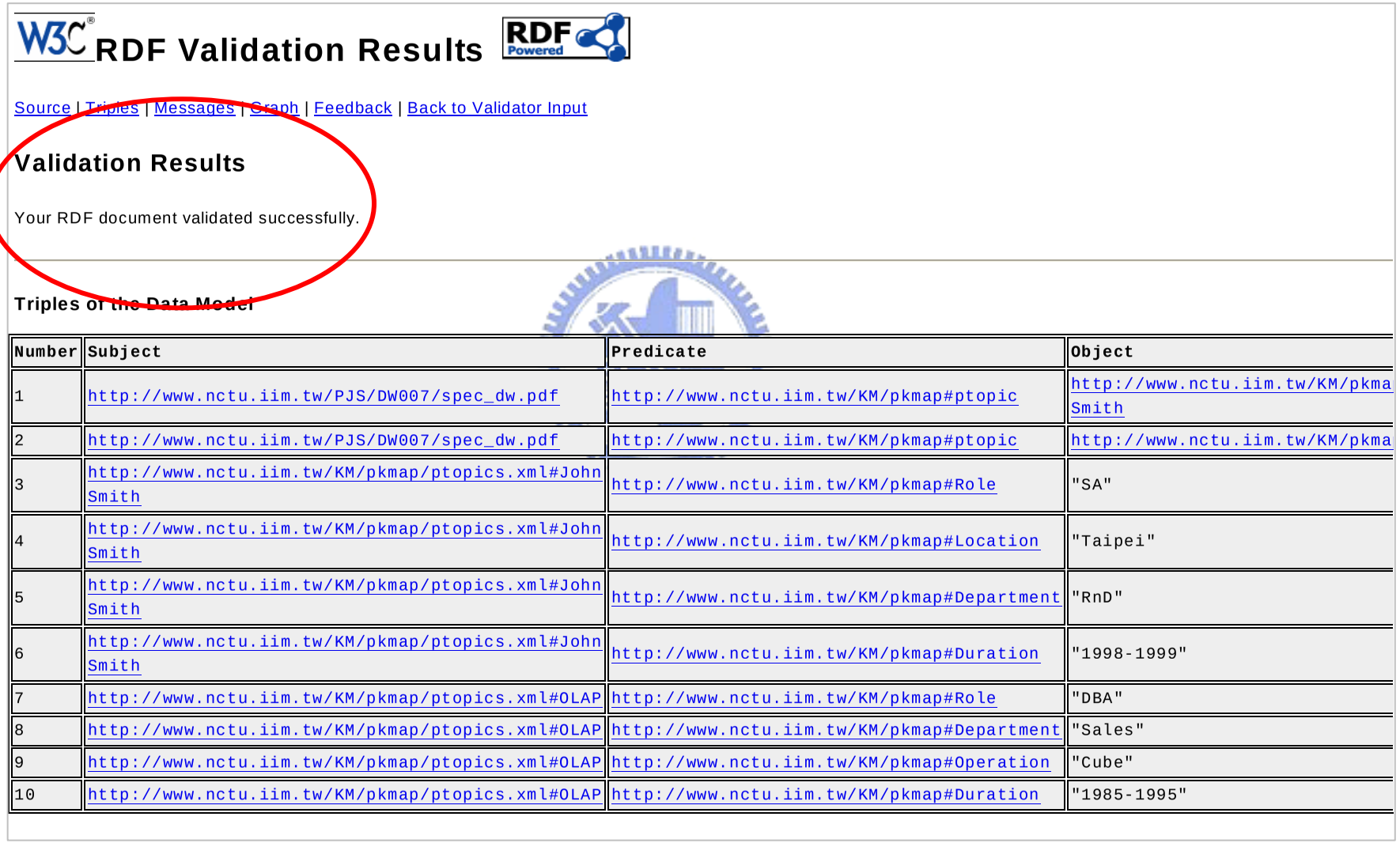


Figure 42. W3C online RDF Validation Parser with RDF graph



RDF Validation Results

Source | Triples | Messages | Graph | Feedback | Back to Validator Input

Validation Results

Your RDF document validated successfully.

Triples of the Data Model

Number	Subject	Predicate	Object
1	http://www.nctu.iim.tw/PJS/DW007/spec_dw.pdf	http://www.nctu.iim.tw/KM/pkmap#ptopic	http://www.nctu.iim.tw/KM/pkma Smith
2	http://www.nctu.iim.tw/PJS/DW007/spec_dw.pdf	http://www.nctu.iim.tw/KM/pkmap#ptopic	http://www.nctu.iim.tw/KM/pkma
3	http://www.nctu.iim.tw/KM/pkmap/ptopics.xml#John Smith	http://www.nctu.iim.tw/KM/pkmap#Role	"SA"
4	http://www.nctu.iim.tw/KM/pkmap/ptopics.xml#John Smith	http://www.nctu.iim.tw/KM/pkmap#Location	"Taipei"
5	http://www.nctu.iim.tw/KM/pkmap/ptopics.xml#John Smith	http://www.nctu.iim.tw/KM/pkmap#Department	"RnD"
6	http://www.nctu.iim.tw/KM/pkmap/ptopics.xml#John Smith	http://www.nctu.iim.tw/KM/pkmap#Duration	"1998-1999"
7	http://www.nctu.iim.tw/KM/pkmap/ptopics.xml#OLAP	http://www.nctu.iim.tw/KM/pkmap#Role	"DBA"
8	http://www.nctu.iim.tw/KM/pkmap/ptopics.xml#OLAP	http://www.nctu.iim.tw/KM/pkmap#Department	"Sales"
9	http://www.nctu.iim.tw/KM/pkmap/ptopics.xml#OLAP	http://www.nctu.iim.tw/KM/pkmap#Operation	"Cube"
10	http://www.nctu.iim.tw/KM/pkmap/ptopics.xml#OLAP	http://www.nctu.iim.tw/KM/pkmap#Duration	"1985-1995"

Figure 43. The validation results of W3C Validation Parser

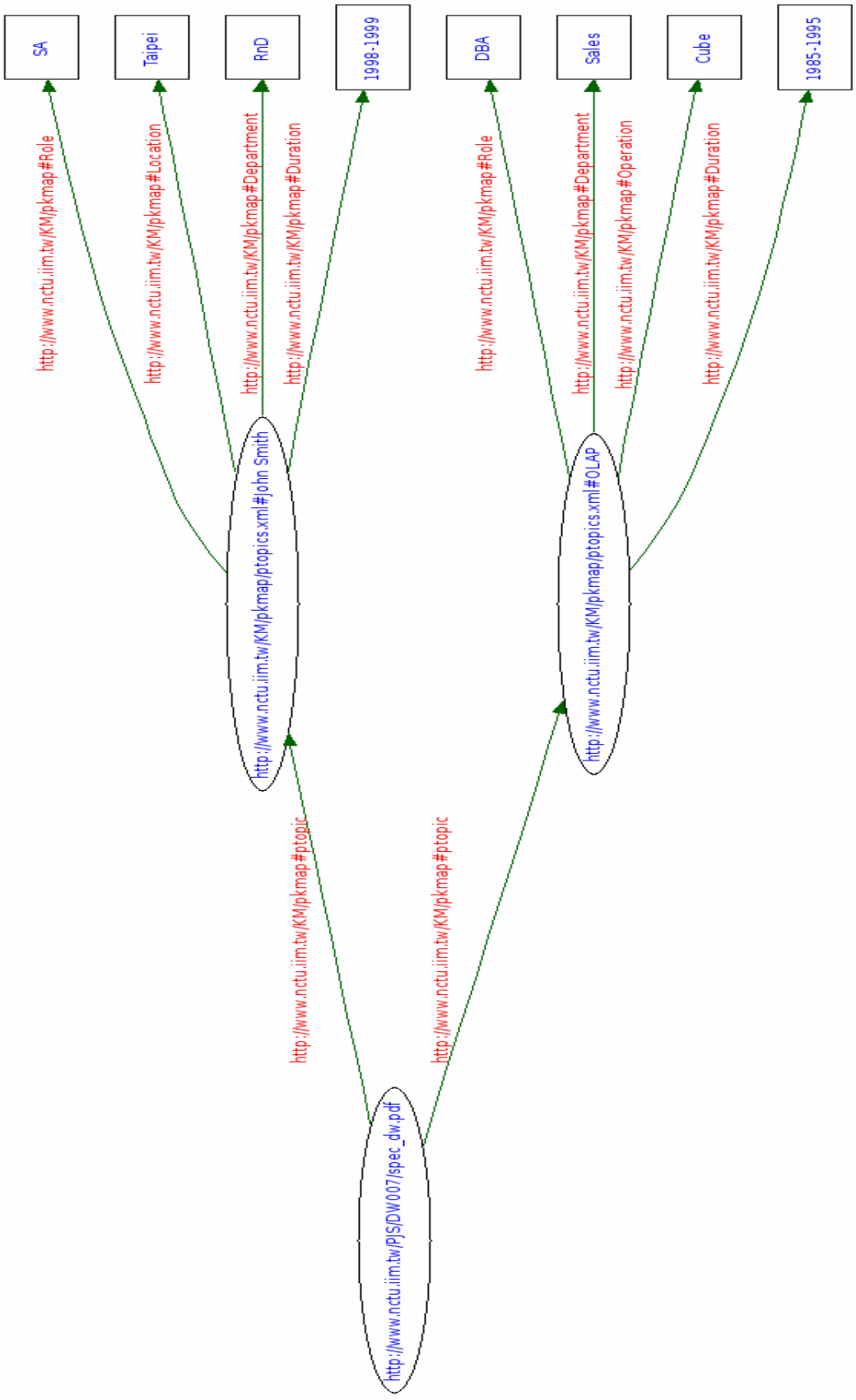


Figure 44. The RDF graph generated by W3C RDF Validation Parser

The RDF Validation Parser is demonstrated on the W3C site. It allows you to paste an RDF document into a form and get a visualization of graph, triples, and a validation of the RDF [57]. As shown in Figure 42,Figure 43 and Figure 44, the main page for pasting RDF file and selecting graph format, the successful validation result and RDF graph are shown separately.

C.2. IsaViz RDF editor

In the other way, we may draw RDF model in graph for generating compatible RDF grammar in the tool of IsaViz RDF editor. IsaViz RDF editor is a visual environment for browsing and authoring RDF models, represented as directed graphs [23]. Current stable version is V.2.1 on October 2004, and the most recent development version is V.3.0 alpha on December 2005. Moreover, we also have to download the package of Graphviz/dot for the specific platform (version 1.8.9 or later is mandatory for installing IsaViz 2). As IsaViz is implemented in Java, it requires a JVM (Java Virtual Machine) 1.3.0 or later (1.4.0 or later strongly recommended) since the GUI makes extensive use of Java2D which featured poor performances in JVM 1.2.x. Bascally, IsaViz RDF editor features:

- a 2.5D user interface allowing smooth zooming and navigation in the graph.
- creation and editing of graphs by drawing ellipses, boxes and arcs.
- RDF/XML, Notation 3 and N-Triple import.
- RDF/XML, Notation 3 and N-Triple export, but also SVG and PNG export.

As using IsaViz RDF editor, several settings operations are essential. First, the environment parameters are required in the preferences window, as shown in Figure 45. The important directories are set for proper file import and export function in the system, including project directory, GraphViz DOT executable directory, GraphViz font directory, RDF directory and temporary directory.

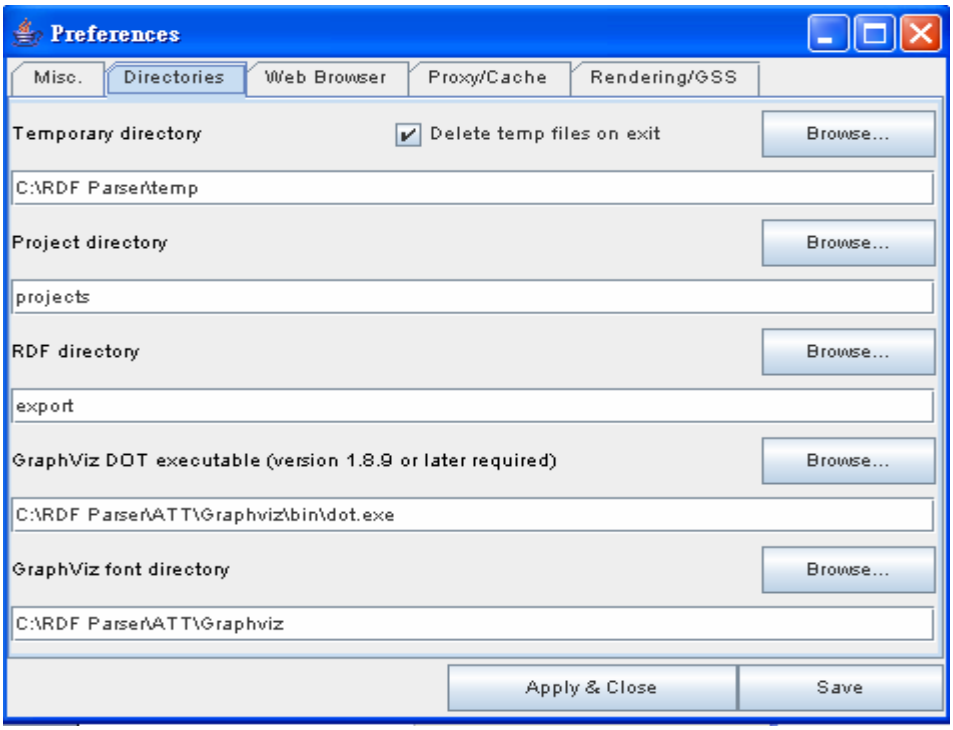


Figure 45. The parameters in the preferences in IsViz RDF editor

Next, the namespaces setting is important for creating the unique RDF model of project-based knowledge map. As shown in Figure 46, the prefix of pkmap is defined and

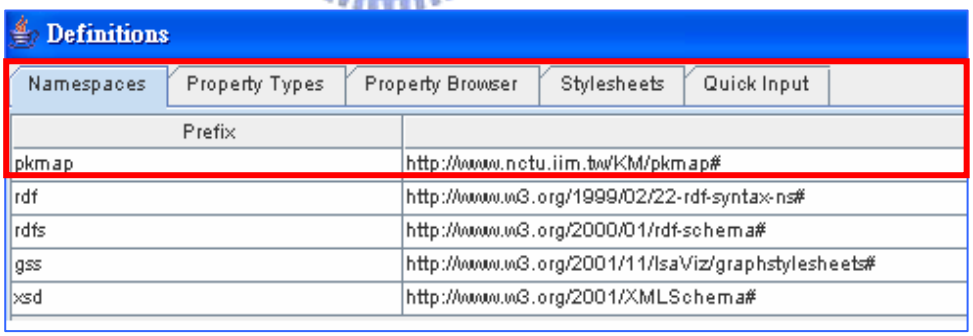


Figure 46. The definitions for namespaces

Next, the main menu of IsViz RDF editor is helpful for creating the node, arc or square on the Graph board for RDF model development. As shown in Figure 47, the window Attributes is used for entering the value for nodes on the window Graph, including subjects, objects and predicates.

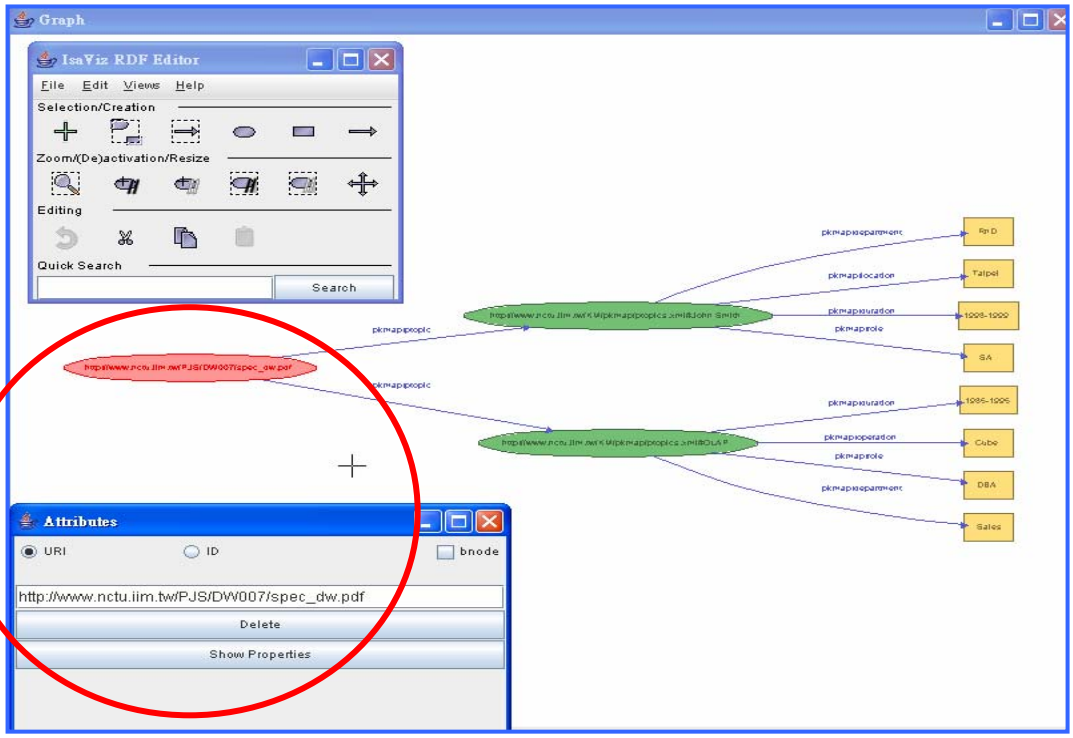


Figure 47. The screenshot of project context in IsViz RDF editor

As building the RDF graph successfully, the corresponding RDF/XML file is also created automatically. The selection of view RDF/XML is used for displaying the RDF/XML file on the screen, as shown in Figure 48.

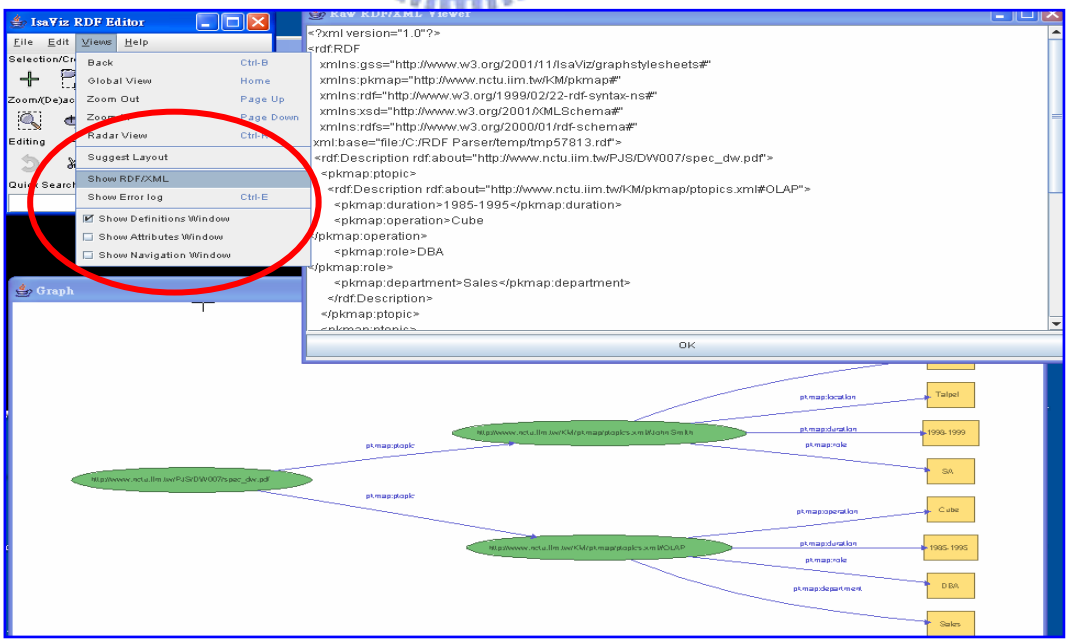


Figure 48. The screenshot of RDF/XML file for the model