
Project-based knowledge maps: combining project mining and XML-enabled topic maps

Duen-Ren Liu and
Chouyin Hsu

The authors

Duen-Ren Liu is a Professor of the Institute of Information Management, National Chiao Tung University, Hsinchu, Taiwan. Chouyin Hsu is an Instructor of the Department of Information Management, Overseas Chinese Institute of Technology, Taichung, Taiwan and a PhD Student at the Institute of Information Management, National Chiao Tung University, Hsinchu, Taiwan.

Keywords

Knowledge management, Extensible Markup Language, Topic maps, Internet, Data handling

Abstract

Many enterprises implement various business projects on the Internet in the global knowledge economy. The task of managing distributed and heterogeneous project knowledge is very important in increasing the knowledge assets of enterprises. Accordingly, this work presents a project-based knowledge map system to properly organize project knowledge into topic maps, from which users can obtain in-depth concepts to facilitate further project development. A two-phase data mining approach involving the ISO/IEC 13250 topic maps and Extensible Markup Language (XML) is used to establish the proposed system, which can determine knowledge patterns from previous projects and transform these patterns into a navigable knowledge map. The map can help users to locate required information and also offers subject-related information easily and rapidly over the Internet.

Electronic access

The Emerald Research Register for this journal is available at
www.emeraldinsight.com/researchregister

The current issue and full text archive of this journal is available at
www.emeraldinsight.com/1066-2243.htm

Introduction

Knowledge management (KM) is crucial to the adaptation and survival of organizations in the face of continuous environmental changes (Malhotra, 1998). The activities of knowledge acquisition, storage and distribution in a KM system enable the dynamic creation and maintenance of an enterprise's intelligence (Fischer and Ostwald, 2001). According to KM research reports (KPMG Consulting, 1999, 2003), about 80 percent 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.

Knowledge management involves a thorough, systematic approach to information repository of an organization by a sequence of collaborative processes. Moreover, the rapid growth of the Internet infrastructure and information technology (IT) has increased the effectiveness of KM in many business domains. Bolloju *et al.* (2002) proposed an integrative model for building enterprise decision support environments using model marts and model warehouses as knowledge repositories. IBM took four years to reengineer their customer relationships by acquiring and disseminating knowledge to both customers and human experts (Malhotra, 1998). Moreover, metadata were used as a knowledge management tool for supporting user access to spatial data (West and Hess, 2002). Rubenstein-Montano (2000) surveyed knowledge-based information systems for urban planning and suggested the importance of moving towards knowledge management. 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 (Schwartz and Te'eni, 2000). The global Internet has clearly increased the availability of KM systems.

New IT has led to the prosperity of KM over the past decade, including artificial intelligence (AI), machine learning, international standards, data warehousing, object-oriented programming and data mining and more. For example, the best-practice knowledge base system benefits from the AI-related methods (O'Leary, 1998). Bhattacharyya *et al.* (2002) utilize evolutionary computing techniques to develop a trading model of high-frequency data on foreign exchange markets. Data warehousing functionality is

The authors gratefully acknowledge the anonymous reviewers for their valuable suggestions. This research was supported by the National Science Council of the Republic of China under the grant NSC 92-2416-H-009-010.

integrated with prescriptive-oriented data mining tools to improve churn management (Lejeune, 2001). The concept of patterns is considered as knowledge media, which is helpful for converting information between explicit and tacit knowledge (May and Taylor, 2003). These penetrating skills improve the quality of knowledge in KM systems.

Afterward, semantic expression of knowledge to facilitate rapid understanding and sharing of knowledge is increasingly important. Accordingly, this work employs topic maps, a standardized notation for interchangeably representing information (International Organization for Standardization, 2000), to improve the knowledge expression and exchangeability over the Internet. Topic maps are helpful in constructing a knowledge map to represent integrated concepts clearly and help users to navigate rapidly required and relevant project knowledge. Notably, topic maps have been referred to as the global positioning system of the information universe and is the basic technology for representing knowledge (International Organization for Standardization, 2000). Hence, the proposed knowledge map takes advantage of topic maps to accommodate the existing heterogeneity of project resources and offers a guide-like service to help knowledge seekers to find required and relevant information.

Primarily, project knowledge includes both the results of, and the work experiences of, project processes. Thus, two-phase data mining is applied to group the considerable project resources and determine nontrivial association patterns. The potential multiple-phase data mining algorithms can provide more valuable patterns for the proposed knowledge map. Moreover, Extensible Markup Language (XML) is applied to control all definitions of discovered patterns and rules to insure the consistency of the proposed knowledge map. The advantage of XML is that it represents a compromise between flexibility, simplicity, and readability by both humans and machines (World Wide Web Consortium (W3C), 2000). So XML is rapidly becoming an information-exchange standard for integrating data among various Internet-based applications (Bertino and Ferrari, 2001). Hence, the proposed project-based knowledge map, benefiting from XML-enable topic maps, is indeed practical in supporting the sharing and understanding of consistent project knowledge over the Internet.

KM in project management

Projects are a popular work-type for accomplishing different assignments (Project Management Institute, Inc., 2000). However, some unavoidable

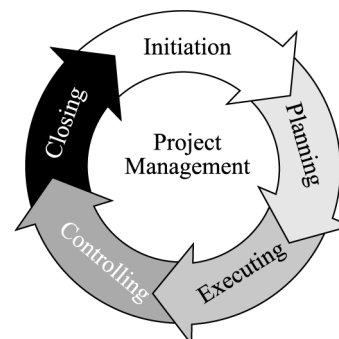
problems and changes have increased the challenges associated with project management. Therefore, KM is of great interest in the field of project management.

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 as shown in Figure 1 (Project Management Institute, Inc., 2000). However, a project team is usually disbanded and reorganized before another new project, meaning that such cooperation is temporary. This kind of volatile relationship hinders the accumulation 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 causes a project team to have access to diverse resources and to have a range of different backgrounds. The one-off nature of the work means that previous learning often contributes little to new projects. That is, a project manager must spend much time when looking for relevant information that may help his or her projects, and project participants are condemned often to repeat the same failures.

Knowledge management technology provides an opportunity to solve above dilemmas. Tah and Carr (2001) applied knowledge management technology to identify project risk and further improve project management. Barthès and Tacla (2002) developed an agent-supported portal to organize knowledge in complex R&D projects. Deng *et al.* (2001) developed an integrated information system based on project-specific subjects. Czuchry and Yasin (2003) offered a practical integrated informational approach to balance the strategic and operational concerns.

The ubiquitous Internet-based technologies increasingly facilitate the development of knowledge management. Li *et al.* (2002) developed a web-based knowledge management system to integrate large-scale projects in the

Figure 1 The process of project management



construction industry. Schwartz *et al.* (2000) discussed the synergy between the Internet, knowledge management and virtual organization. Internet-based technologies have great ability to support the essential tasks of knowledge management, including knowledge acquisition, organization and distribution that foster the demand of virtual organizations. Accordingly, Internet-based technologies provide an accessible, scalable, and effective platform to support knowledge management functionality, and to consolidate project management activities.

Managing the distribution of resources, instability of costs, unexpected risks and the required quality in a tight schedule is a tough mission. Internet-based KM technologies indeed have contributed many kinds of improvement. However, project management is continuously changing in a global competitive environment. Thus, efficient and innovative Internet-based KM technologies are in great demand in the field of project management.

ISO/IEC 13250 topic maps

The ISO standard ISO/IEC 13250 topic maps define a model for the semantic structuring of link networks (International Organization for Standardization, 2000). The structural information conveyed by topic maps typically includes: groupings of addressable information objects (occurrences) around topics; and relationships (associations) among topics. Thus, topic names, occurrences and associations are three basic components in topic maps as shown in Figure 2:

- (1) Topics are correspondent with the formal description of concepts or objects in the real

world. An expressive topic name is helpful to understand an object or a concept. Certainly, topic types are naturally created to classify these topics. That is, topic types represent typical class-instance relationships according to which topics are organized.

- (2) Occurrences are the original information resource connected to the meaningful topic names. A topic is an abstract label, and the occurrences are substantial references.
- (3) An association is formally a meaningful link that specifies a relationship among several topic names. These semantic association relationships help users understand the connections among these topic names.

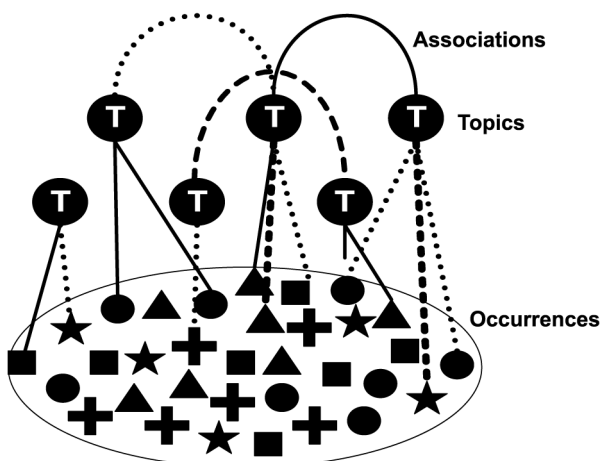
As described above, topic maps provides a clear structure in which information from topic type to information resource can be organized. Knowledge seekers can use this information to head in the right directions to obtain their required information. Therefore, topic maps is used herein to construct a navigable project-based knowledge map.

Knowledge maps

Successive knowledge map applications take advantage of the nature of visualization and navigation to locating and publishing knowledge (Davenport and Prusak, 1998). Vail (1999) highlighted the visual display of captured information; Browne *et al.* (1997) combined knowledge maps with reasoning-based methodology to elicit probability assessment for decision making; Leathrum *et al.* (2001) employed knowledge maps to design an intelligent questioning system for learning; Chung *et al.* (2003) proposed a knowledge map framework for discovering business intelligence to alleviate information overload on the Web; Lin and Hsueh (2003) applied information retrieval algorithms to generate and maintain a knowledge map system for virtual communities of practice; Kim *et al.* (2003) proposed a road map to develop knowledge maps in the industrial community.

These works employed knowledge maps to organize knowledge in various applications other than project management. In addition, they developed knowledge maps to locate required knowledge in organizations without considering the exchangeability and interoperation of knowledge maps, which however is a crucial issue for Internet-based systems. Accordingly, we employ the ISO/IEC 13250 topic map model to construct the project-based knowledge maps. The topic map model is a web-based standard to

Figure 2 Association, topics, and occurrences in topic maps



provide clear-cut hierarchical configurations. The XML technology is adopted to define the content of knowledge as well as the architecture and the representation of knowledge maps. Consequently, project-based knowledge maps, expressed in XML-enabled topic maps, can be exchanged and interoperated more easily on the Internet. Effective collaborations of project management over the Internet can thus be facilitated.

Data-mining approach

The data-mining approach 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 (Fayyad *et al.*, 1996; Berry and Linoff, 1997). This work employs the clustering method and association-rule mining to extract collaboratively knowledge patterns from historical projects as shown in Figure 3.

Clustering is an unsupervised learning method; users group data without known or specified classes, to maximize similarity within each cluster and dissimilarity among different clusters. Well-known algorithms, such as Agglomerative, K-means, CURE, CHAMELEON, have been implemented in numerous applications and successfully applied in pattern recognition, image processing, market research and other areas (Jain and Dubes, 1988; Baeza-Yates and Ribeiro-Neto, 1999).

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. The Apriori algorithm (Agrawal *et al.*, 1993) has been implemented in numerous applications and successfully applied in market basket analysis, recommender systems, user behavior analysis, and other areas.

XML specifications

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 W3C and the second edition was published in 2000 (W3C, 2000). Unlike Hyper Text Markup Language (HTML), XML allows users to define their own set of markup tags that relate to their documents.

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. A “document type definitions” (DTD) describes the custom tags and rules for XML data (Martin *et al.*, 2001). Notably, a DTD is a set of declarations that incorporate XML data to describe their structure and permissible format. 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 knowledge map in project management.

However, XML exclusively focuses on manipulating data, and Extensible Stylesheet Language (XSL) is useful for displaying data. XML formalism separates the structure from the representation of data, so users can use XSL to design different output formats of a single XML data (W3C, 1999). Moreover, users can employ XSL to implement different templates for HTML, PDF, Palm and WAP view types. Thus, the same XML data may be shared in many Internet-based applications, to increase the exchangeability of information over the Internet.

Consequently, an adaptive DTD with corresponding XML data and multiple XSL files are used to establish the content and the representation of the proposed project-based knowledge map, which is a network-accessible knowledge map, in which project participants and managers can obtain the information on the Internet in a convenient manner, as shown in Figure 4.

Figure 3 Two-phase data mining for extracting project knowledge

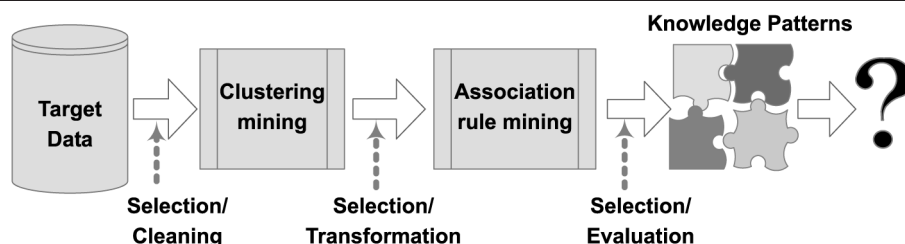
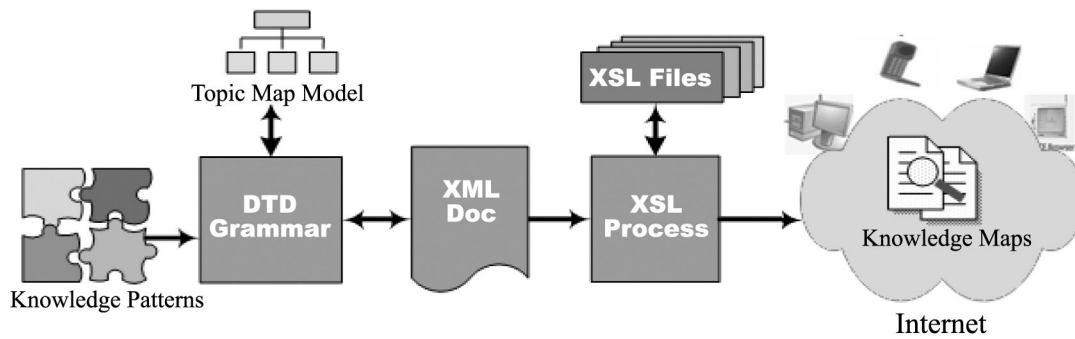


Figure 4 The process of deploying Internet-based knowledge maps



System framework

This section illustrates the system framework for building project-based knowledge maps. As shown in Figure 5, four cooperative modules are involved in constructing a knowledge map, including meta information builder for collecting meta information about projects, knowledge discovering module for extracting knowledge patterns, knowledge map generator for deploying a required knowledge map, and the module of knowledge base is designed to accumulate all important products generated from other modules in a timely manner.

Meta information builder

A project is composed of various project objects, each of which is associated with some certain objectives, particular manipulations, and even special performance measures or technical annotations. Meta information refers to important attributes and annotations of projects. The meta-information builder is employed to extract meta information during project processing. When a project is finished and archived, the meta information is collected afterward. This builder appropriately collects the important attributes of projects and converts these attributes into agreeing meta information, which is composed of

multi-dimensional facts. Thus, each project is given a set of multi-dimensional data, which is a computational format for advanced analysis.

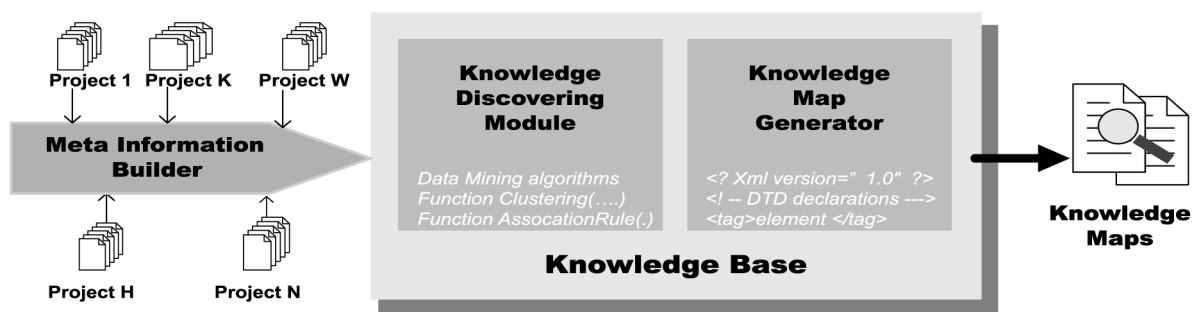
Knowledge-discovering module

A collection of consistent meta information delivered from the meta-information builder is ready for knowledge discovering task (Figure 5). A two-phase data mining approach is applied to extract valuable project attributes, and discover their association relationships. The first phase involves clustering methods to group projects according to similarities. Projects in the same cluster are approximately associated with similar topic names. The second phase employs association rule mining to discover association patterns for each cluster. An association relationship in a cluster is determined by coexistences of several topics. Most association relationship can be converted into support rules, which are very helpful in making decisions, including about assigning work, selecting teammate, or preparing activities.

Knowledge map generator and knowledge base

The knowledge map generator mainly provides two services – converting the discovered patterns into XML format, and representing the complete

Figure 5 System framework of deploying project-based knowledge maps



knowledge map. The predefined DTD sets up the block format of topic maps to describe the conclusive project knowledge, and corresponding XML data clearly describe the detailed content of knowledge maps. Moreover, the various XSL file sheets are prepared to display knowledge maps over the Internet. As shown in Figure 4, users can access the required knowledge maps over the Internet easily and rapidly.

Knowledge base primarily stores and manages nontrivial products of all modules to prevent users from overusing or double defining any concept, such as a topic name or association relationship. The advantage is helpful to sustain the uniqueness and effectiveness of each terminology, as it benefiting the maintenance of knowledge maps. Furthermore, it governs the input and output of data for each module, such as forwarding compatible meta data to knowledge discovering module, or the discovered knowledge patterns to map generator.

Structure of project-based knowledge maps

Based on topic maps, the proposed knowledge map contains four layers: the topmost layer is a cluster entrance; the second layer contains the predefined category names; the third layer contains several topic names of project attributes to describe the major features of this cluster; the bottom layer collects project resources. Bidirectional links are used to properly connect topic names to related project objects. The hierarchy outlook of a knowledge map is shown in Figure 6.

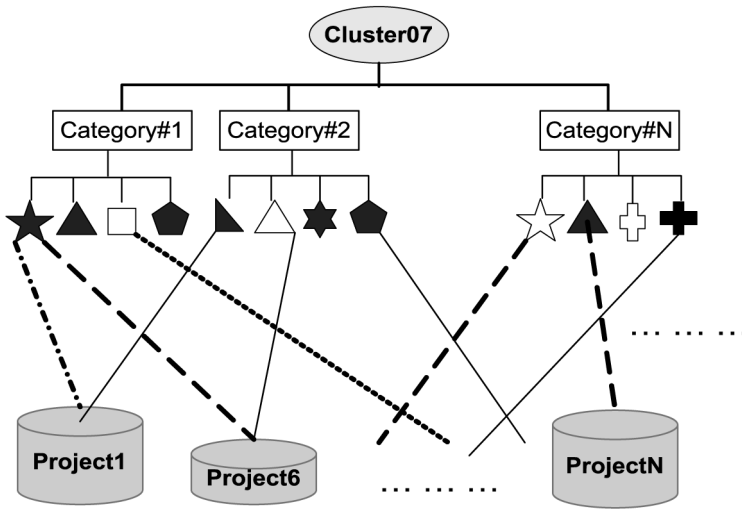
Category names basically correspond to the predefined topic types in this work, which are helpful to classify topic names. Hence, a set of clear-cut category names pre-defined by human experts or ontology-subject way are used to organize topic names. We summarize the experts' opinions and the IEEE Standard for Software Project Management Plans (Institute of Electrical and Electronics Engineers, Inc., 1998) and propose five category names as topic types – member, tool, goal, activity and cost – to classify the project knowledge.

Topic names practicably correspond to project attributes. A standard definition format for each attribute is helpful to unify the concepts. According to the notation of topic maps, a topic is designed to three parts: the base name, the display name and the sort name (International Organization for Standardization, 2000). For example, eXML/XML/Extensible Markup Language, is the base/display/sort name, respectively.

Occurrences practically correspond to the instances of topic names, that is, the related project objects of a topic name. Project objects are the original resource of project knowledge. An electronic representation is helpful to hyperlink to the proper topic names. The custom DTD declaration, shown in Figure 7, defines the syntax rules to describe the basic elements of category names, topic names and the related occurrences.

Associations are used to note the certain relationships between few topics in this work. Moreover, an association relationship indicates some working experiences that are valuable for future support. In order to reuse these experiences, we simplify the DTD declaration to represent an association as shown in Figure 8.

Figure 6 The structure of a project-based knowledge map



Meta information and project attributes

This section gives a set of pertinent data to demonstrate the building of a knowledge map and several important algorithms are explained. First of all, a vector model is used to record meta information of project objects and the important project attributes.

Vector models are intrinsically suitable to stand for multi-dimensional data in a computational format (Jain and Dubes, 1988; Baeza-Yates and Ribeiro-Neto, 1999). Thus, a vector model is used to describe the original multiple attributes of project objects, and to select the important attributes of a project by different feature selection algorithms or heuristic evaluation as shown in Figure 9. Moreover, the conclusive project

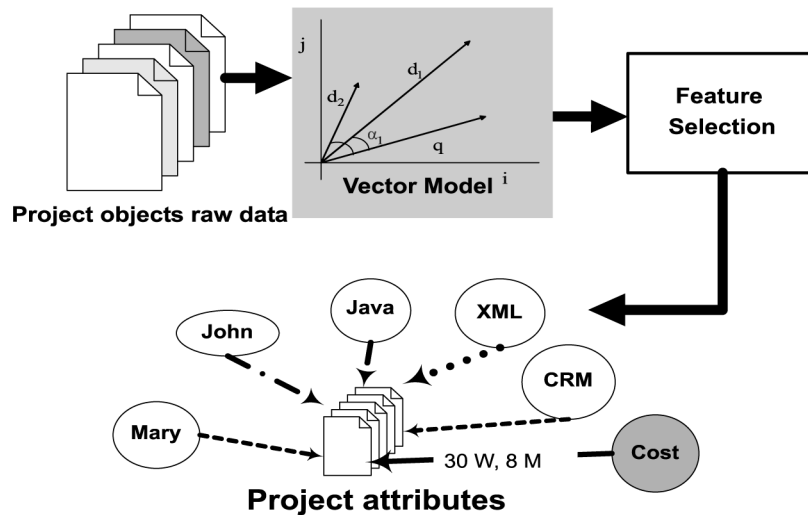
Figure 7 An excerpt of DTD to declare a topic name and its occurrences

```
<! ELEMENT ProjTopic (baseName, displayName, sortName, OccurrenceRef+ ) >
<! ATTLIST ProjTopic
    TopicID      ID      #REQUIRED
    Category     (Member | Tool | Goal | Activity | Cost) ... #REQUIRED >
<! ELEMENT baseName      (#PCDATA) >
<! ELEMENT displayName   (#PCDATA) >
<! ELEMENT sortName      (#PCDATA) >
<! ELEMENT OccurrenceRef EMPTY >
<! ATTLIST OccurrenceRef
    xlink:type      NMTOKEN      #FIXED 'simple'
    xlink:href      CDATA          #REQUIRED >
```

Figure 8 An excerpt of DTD to declare an association

```
<! ELEMENT ProjAssociation (AssociationTopic)+ >
<! ATTLIST ProjAssociation
    AssociationName ID      #REQUIRED >
<! ELEMENT AssociationTopic (#PCDATA) >
```

Figure 9 Project attributes are selected from meta information in the vector model



attributes are clearly represented in the vector model.

A project O_j is associated with a multi-dimensional vector, that is, $O_j = (w_{1j}, w_{2j}, \dots, w_{kj})$, 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. For example, a set of project vectors are given in Table I; project P001 is presented as a vector of

$(1,1,1,1,1,0,0,1,1,1,1,1,1,0,0,1,1,1,0,0,100,9)$, such that according to the definition of each dimension in the list below, project P001 involves Bob, David, John, Kim and Mary; it involve the tools of data mining, XML, DBMS and CGI; it involves the goals of CRM, logistics, and data warehouse; it involves the activities of marketing, agency work, and consulting; it takes around 100 weeks and cost 9 million dollars:

- *Member* (key workers or project leaders): Bob (m1), John (m2), David (m3), Kim (m4), Mary (m2).

Table I A collection of projects with proper attributes in vector models

	m1	m2	m3	m4	m5	t1	t2	t3	t4	t5	t6	g1	g2	g3	g4	g5	a1	a2	a3	a4	a5	c1	c2
P001	1	1	1	1	1	0	0	1	1	1	0	1	1	1	0	0	1	1	1	0	0	100	9
P002	0	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	1	1	0	50	8
...
P003	0	1	0	0	0	1	1	0	0	1	1	0	0	1	1	1	0	1	0	1	1	100	10
P100	0	1	0	1	1	0	0	0	1	1	0	1	1	0	0	0	1	1	0	0	0	40	6
P101	0	1	1	0	1	0	0	1	0	1	1	0	1	1	1	0	1	1	1	1	0	50	5
...
Pn	1	1	0	0	0	1	1	1	0	0	0	1	0	0	0	1	0	0	0	0	1	10	2

- *Tool: packages* (standards, languages, applied in projects): Parallel (t1), Java (t2), data mining (t3), XML (t4), DBMS (t5), CGI (t6).
- *Goal* (key purposes, objectives of projects): CRM (g1), logistics (g2), data warehouse (g3), ERP (g4), LAN (g5).
- *Activity* (held actions, events, contests in projects): marketing (a1), agency (a2), consultant (a3), alliance (a4), contest (a5);
- *Cost* (average working weeks and finance expense).

The last two appended dimensions record expenses and period of the project. This work mainly considers these two dimensions to estimate the average cost of a cluster of projects.

Mining project knowledge

Two-phase data mining is employed to discover project knowledge. The first phase employs an Agglomerative algorithm to cluster projects and then extract important project attributes from each cluster. The second phase uses an Apriori algorithm to discover association relationships.

Clustering projects

The Agglomerative algorithm conducts hierarchical clustering to group projects in a bottom-up way (Jain and Dubes, 1988). The algorithm places each object in its own cluster and gradually merges these atomic clusters into larger and larger clusters until all objects are in a single cluster. The major steps are constructing a dissimilarity matrix, forming clusters, and determining centroids.

Constructing a dissimilarity matrix is to measure the distance between projects. This work uses Euclidean distance measurement to compute the proximity distance between pair-wise vectors. The pair-wise dissimilarity matrix of the vectors is shown in Table II.

Forming clusters is to form clusters based on the dissimilarity matrix. A threshold value is an important parameter to decide how “close” the

Table II A dissimilarity matrix using the measure of Euclidean distance

	P001	P002	P003	...	P100	P101	...	Pn
P001	0	2.6	3.9	...	2.4	2.2	...	3.9
P002	2.6	0	3.3	...	2.8	2.4	...	4.1
P003	3.9	3.3	0	...	3.9	3.2	...	2.8
...
P100	2.4	2.8	3.9	...	0	2.8	...	4.0
P101	2.2	2.6	3.2	...	2.8	0	...	4.0
...
Pn	3.9	4.1	2.8	...	4.0	4.0	...	0

projects will form a cluster. Projects with distance values less than a threshold value are grouped into the same cluster. Given threshold value 3.0, project P001, P100, P101 and P002 could form a cluster and be labeled as “Cluster07”, which is the main illustrative example in this paper.

Determining centroid involves choosing a cluster centroid, which represents the major attributes of a cluster of projects. The frequently appearing attributes will form the cluster centroid. A vector model is also applied to represent a cluster centroid. If an attribute was present more often than the threshold value in a cluster, then this attribute is set to 1 in the centroid vector; otherwise, it is set to zero. According to Cluster07, the given example above, the centroid vector is (0, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 0, 0, 60, 7) by the threshold value 0.5. That is, the important attributes of Cluster07 are David, John, Kim, Mary, Data Mining, XML, DBMS, Java, CRM, logistics, data warehouse, marketing, agency, consultant, 70 average working weeks and 7 million dollars. The last two items represent the average cost of this cluster and are used to estimate the cost of new projects.

The extracted important attributes of a cluster centroid form the topics in the discovered project-based knowledge map. Moreover, hyperlinks are created to link each topic name to relevant references (occurrences) as shown in Figure 10.

Mining associations among project attributes

Mining association patterns is helpful to derive the practical rules, which could give users immediate

Figure 10 An excerpt of XML document of a new topic

```

<ProjTopic TopicID="CRM" Category="Goal">
  <baseName>CRM</baseName>
  <displayName>C.R.M.</displayName>
  <sortName>Customer Relationship Management</sortName>
  <OccurrenceRef
    xlink:href="http://w2.iim.nctu.edu.tw/Proj/Project19990510/K2.doc"
  />
  <OccurrenceRef
    xlink:href="http://www.iim.nctu.edu.tw/PP/Project200012/B23.doc"
  />
  <OccurrenceRef
    xlink:href="http://w3.ocit.edu.tw/Projects/Project699/A2.doc"
  />
  <OccurrenceRef
    xlink:href="http://ww2.ocit.edu.tw/Server5/Project366/P3.doc"
  />
</ProjTopic>

```

decision-making support. The formal statement of the association rule is described following (Agrawal *et al.*, 1993). An association rule is an implication of the form $X \Rightarrow Y$, where X and Y are sets of items. $X \Rightarrow Y$ holds in the transaction set D with confidence c if c percent of transactions in D that contain X also contain Y . $X \Rightarrow Y$ has support s in the transaction set D if s percent of transactions in D contain $X \cup Y$.

The principle of association rule mining is to find out the co-occurrence patterns of items and summarize interesting rules from them. The Apriori algorithm is typically used to find association rules by discovering large itemsets. An itemset (set of items) is considered large if the support of that itemset exceeds a user-specified minimum support. The main procedures are given as follows (Agrawal *et al.*, 1993). The algorithm counts the support of individual items and determines which of them have minimum support. Each subsequent step starts with a seed set of large itemsets to generate the new potentially large itemsets, called candidate itemsets. The algorithm then determines which of the candidate itemsets are actually large. This process continues until no large itemsets are found.

This work employs Apriori algorithm to find association patterns satisfying minimum support from the meta information of project objects. Table III gives the supposed vector model of meta information for Cluster07. Based on the minimum support 0.2 and minimum confidence 0.5, we got interesting association patterns as shown in Figure 11.

Usually, a set of large itemsets are found, some are useful, some are not. The criteria are dependent on user's opinions. For example, the connection between Mary and John shown at line 3

in Figure 11 will be a helpful hint for selecting a member. The knowledge base centrally stores many valuable associations for further use. Each association has a unique association name and few related topic names as shown in Figure 12.

Notably, the discovered association patterns form the associations among topics in the project-based knowledge map.

Deploying project-based knowledge maps

A knowledge map is developed to display the project knowledge, generated from the mining process. Such a map provides users with a navigable architecture to find useful and relevant information. Figure 13 shows the knowledge map of Cluster07, including categories, topic names and association rules. Each topic name is equipped with hyperlinks to get the reference resources.

For example, the "member" category lists the acceptable participants to be considered; the "tool" category shows users the necessary training; the "cost" category help users to estimate the expense and arrange the schedule. Each category includes several important topic names, and a customary hyperlink for each topic name will connect to the relevant project references. These advantages help project managers to develop new projects, while taking advantage of prior knowledge, and help project participants to obtain various kinds of relevant information.

Furthermore, the proposed knowledge map summarizes associations to a set of support rules to increase the efficiency of this system. These formal and semantic rules, which are valuable expertise, can support users immediately instead of taking

Table III The attributes of project objects

		John	David	Kim	Mary	Java	DM	XML	DBMS	CRM	LG	DW	MK	AG	CT
P001	A1	1	1	0	1	0	0	1	0	0	1	0	0	0	1
	A2	1	0	0	1	0	1	0	0	0	0	1	0	1	0
	A3	1	0	0	1	0	1	0	0	1	0	0	1	0	0
P002	B1	1	0	0	1	1	0	0	0	0	1	0	0	0	0
	B2	0	0	0	0	0	1	0	0	1	0	1	0	0	1
	B3	0	1	0	0	0	0	0	1	0	0	1	0	0	1
P100	C1	1	0	0	0	1	0	0	1	0	0	0	0	1	0
	C2	1	0	0	1	1	0	0	0	0	0	0	0	0	1
	C3	0	0	1	1	0	1	0	0	1	0	0	0	0	0
P101	D1	0	1	0	0	0	1	1	0	1	0	0	1	0	0
	D2	0	0	1	0	0	1	0	0	0	1	1	0	0	1

Figure 11 The association result of Apriori algorithm

1. CRM=1 4 ==> DM=1 4 conf:(1)
2. DM=1 6 ==> CRM=1 4 conf:(0.67)
3. Mary=1 6 ==> John=1 5 conf:(0.83)
4. John=1 6 ==> Mary=1 5 conf:(0.83)
5. XML=1 2 ==> David=1 2 conf:(1)
6. David=1 3 ==> XML=1 2 conf:(0.67)
7. Java=1 3 ==> John=1 3 conf:(1)
8. John=1 6 ==> Java=1 3 conf:(0.5)
9. DM=1 Consultant=1 2 ==> DataWs=1 2 conf:(1)
10. DataWs=1 4 ==> DM=1 Consultant=1 2 conf:(0.5)
11. DM=1 MarKeting=1 2 ==> CRM=1 2 conf:(1)
12. MarKeting=1 2 ==> DM=1 CRM=1 2 conf:(1)
13. DM=1 CRM=1 4 ==> MarKeting=1 2 conf:(0.5)
14. CRM=1 4 ==> DM=1 MarKeting=1 2 conf:(0.5)
15. MarKeting=1 2 ==> CRM=1 2 conf:(1)
16. CRM=1 4 ==> MarKeting=1 2 conf:(0.5)
17. Consultant=1 5 ==> DataWs=1 3 conf:(0.6)
18. DataWs=1 4 ==> Consultant=1 3 conf:(0.75)

Figure 12 An excerpt of XML data defining an association

```

<ProjAssociation AssociationName="has_skill">
  <AssociationTopic>John</AssociationTopic>
  <AssociationTopic>Java </AssociationTopic>
</ProjAssociation>
<ProjAssociation AssociationName="team up">
  <AssociationTopic>John</AssociationTopic>
  <AssociationTopic>Mary </AssociationTopic>
</ProjAssociation>
<ProjAssociation AssociationName="support for">
  <AssociationTopic>Consultant</AssociationTopic>
  <AssociationTopic>Data Mining</AssociationTopic>
  <AssociationTopic>Data Warehouse</AssociationTopic>
</ProjAssociation>

```

many working years. As shown in Figure 14, rules 1 and 2 indicate the member's skills; rule 3 indicates suitable teammates; rules 4, 5, 6 and 7 indicate some specific activities and tools that are

used to reach different goals. This system discovers these experiences to improve project-based knowledge sharing, rather than being hidden in the mind of participants or being lost.

Figure 13 The project-based knowledge map

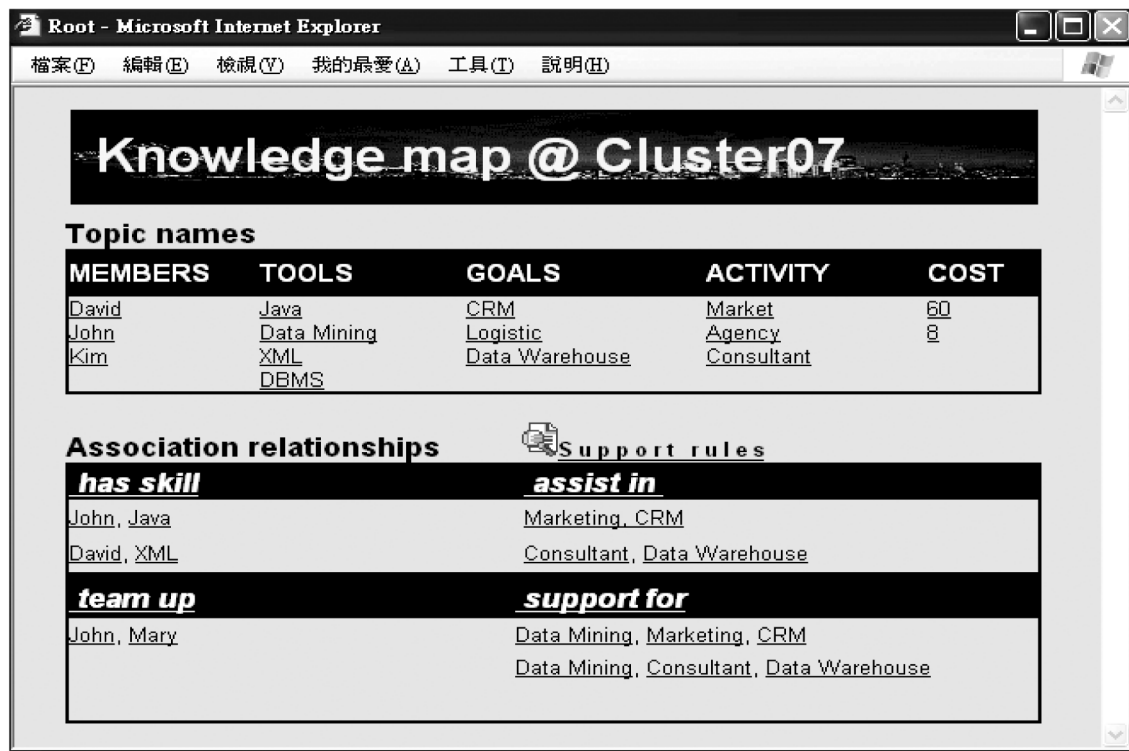


Figure 14 The derived support rules



Conclusion and future work

Data mining is highly effective in discovering knowledge, but contributes less to representing knowledge. The knowledge map, however, compensates for this shortcoming. The generated concepts or rules capture experience of senior experts, which may take many years to accumulate. The knowledge map provides users an

integrated concept and a convenient portal with a set of proper topic names, through which to access detailed project information. Moreover, the support rules save users from working for many years to get the learning experience in developing projects. The project-based knowledge map expressed in topic maps presents an important advancement in the field of Internet-based project management in the following aspects:

- *Information portals*: different topic names are associated with different accessible links for accessing relevant information resources.
- *Prior knowledge*: a group of proper topic names and the derived support rules form the prior knowledge for further projects. The prior knowledge help user to quickly understand the important attributes and association relationships in different phases of project development, such as reference materials or staffing plans.
- *Ontology concepts*: an insightful interpretation of project knowledge yields the benefits of experiences and domain expertise. This system incorporates ontology by means of unifying concepts, terminology, constraints on knowledge patterns, and systemizing all results into a unique representation.

Focusing on an easy-to-use XML syntactical grammar for implementing the project-based knowledge map is crucial. However, improving the semantic expression of the map in XML specifications is an evitable work. Ongoing development will also concern innovative data mining algorithms to extract different knowledge patterns, and to derive more practical support rules for knowledge sharing.

References

- Agrawal, R.T., Imielinski, T. and Swami, A.N. (1993), "Mining association rules between sets of items in large databases", in Buneman, P. and Jajodia, S. (Eds), *Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, Washington, DC, May 26-28*, ACM Press, New York, NY, pp. 207-16.
- Baeza-Yates, R. and Ribeiro-Neto, B. (1999), *Modern Information Retrieval*, Addison-Wesley, Harlow.
- 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.
- Berry, M.J.A. and Linoff, G. (1997), *Data Mining Techniques: For Marketing, Sales, and Customer Support*, John Wiley & Sons, New York, NY.
- Bertino, E. and Ferrari, E. (2001), "XML and data integration", *IEEE Internet Computing*, Vol. 5 No. 6, pp. 75-6.
- Bhattacharyya, S., Pictet, O.V. and Zumbach, G. (2002), "Knowledge-intensive genetic discovery in foreign exchange markets", *IEEE Transactions on Evolutionary Computation*, Vol. 6 No. 2, pp. 169-81.
- Bolloju, N., Khalifa, M. and Turban, E. (2002), "Integrating knowledge management into enterprise environments for the next generation decision support", *Decision Support Systems*, Vol. 33 No. 2, pp. 163-76.
- 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.
- 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 Science, Big Island, Hawaii*, available on CD-ROM.
- Czuchry, A.J. and Yasin, M.M. (2003), "Managing the project management process", *Industrial Management & Data Systems*, Vol. 103 No. 1, pp. 39-46.
- Davenport, T.H. and Prusak, L. (1998), *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, Boston, MA.
- 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. 239-46.
- Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. and Uthurusamy, R. (1996), *Advances in Knowledge Discovery and Data Mining*, MIT Press and the AAAI Press, Cambridge, MA.
- Fischer, G. and Ostwald, J. (2001), "Knowledge management: problems, promises, realities, and challenges", *IEEE Intelligent Systems*, Vol. 16 No. 1, pp. 60-72.
- Institute of Electrical and Electronics Engineers, Inc. (1998), *IEEE Standard for Software Project Management Plans (SPMP)*, Vol. IEEE Std. 1058-1998, Institute of Electrical and Electronics Engineers, Inc., Piscataway, NJ.
- International Organization for Standardization (2000), *Information Technology SGML Applications – Topic Maps*, Vol. /IEC 13250, ISO, Geneva.
- Jain, A.K. and Dubes, R.C. (1988), *Algorithms for Clustering Data*, Prentice-Hall, Englewood Cliffs, NJ.
- 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.
- KPMG Consulting (1999), *Knowledge Management Research Report 1999*, KPMG Consulting, London.
- KPMG Consulting (2003), *Insights from KPMG's European Knowledge Management Survey 2002/2003*, KPMG Consulting, London.
- 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, Albuquerque, NM*.
- Lejeune, M.A.P.M. (2001), "Measuring the impact of data mining on churn management", *Internet Research*, Vol. 11 No. 5, pp. 375-87.
- 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: Electronic Networking Applications and Policy*, Vol. 12, pp. 371-7.
- Lin, F.-R. and Hsueh, C.-M. (2003), "Knowledge map creation and maintenance for virtual communities of practice", *Proceedings of the 36th Hawaii International Conference on System Science, Big Island, Hawaii*, available on CD-ROM.
- Malhotra, Y. (1998), "Deciphering the knowledge management hype", *Journal for Quality & Participation*, Vol. 21 No. 4, pp. 58-60.
- Martin, D., Birkbeck, M., Kay, M., Loesgen, B., Pinnock, J., Livingstone, S., Williams, P.S.K., Anderson, R., Mohr, S., Baliles, D., Peat, B. and Ozu, N. (2001), *Professional XML 2nd ed.*, Wrox Press, Indianapolis, IN.
- May, D. and Taylor, P. (2003), "Knowledge management with patterns", *Communications of the ACM*, Vol. 46 No. 7, pp. 94-9.
- O'Leary, D.E. (1998), "Using AI in knowledge management: knowledge bases and ontologies", *IEEE Intelligent Systems*, Vol. 13 No. 3, pp. 33-9.

- Project Management Institute, Inc. (2000), *A Guide to the Project Management Body of Knowledge*, Project Management Institute, Inc., Newtown Square, PA.
- 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 No. 3, pp. 155-72.
- Schwartz, D.G. and Te'eni, D. (2000), "Tying knowledge to action with kMail", *IEEE Intelligent Systems*, Vol. 15 No. 3, pp. 33-9.
- Schwartz, D.G., Divitini, M. and Brasethvik, T. (2000), "On knowledge management in the Internet age", in Schwartz, D.G., Divitini, M. and Brasethvik, T. (Eds), *Internet-based Knowledge Management and Organizational Memories*, Idea Group Publishing, Hershey, PA, pp. 1-23.
- Tah, J.H.M. 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. 835-46.
- Vail, E.F. (1999), "Mapping organizational knowledge", *Knowledge Management Review*, No. 8, pp. 10-15.
- 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. 247-64.
- World Wide Web Consortium (1999), *Extensible Stylesheet Language Transformations (XSLT) Version 1.0*, available at: www.w3.org/TR/xslt Vol. W3C Recommendation, World Wide Web Consortium.
- World Wide Web Consortium (2000), *Extensible Markup Language (XML) 1.0*, 2nd ed., available at: www.w3.org/TR/REC-xml (1st ed. published in 1998) Vol. W3C Recommendation, World Wide Web Consortium.

Further reading

- Harris, K., Fleming, M., Hunter, R., Rosser, B. and Cushman, A. (1999), "The knowledge management scenario: trends and directions for 1998-2003", Gartner Group Strategic Analysis Report, Gartner Group Inc., Stamford, CT.
- Holzner, S. (2001), *Inside XML*, New Riders Publishing, Indianapolis, IN.
- 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. 155-70.
- Punj, G. and Stewart, D. (1983), "Cluster analysis in marketing research: review and suggestions for application", *Journal of Marketing Research*, pp. 134-48.
- Rath, H.H. (2000), "Topic maps: templates, topology, and type hierarchies. markup languages: theory and practice", in Sperberg-McQueen, C.M. and Usdin, B.T. (Eds), *Markup Languages*, MIT Press, Cambridge, MA.