ICEB 2002 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-10-2002

Towards a Framework for Discovering Project-Based Knowledge Maps

Chouyin Hsu

Duen-Ren Liu

Towards a Framework for Discovering Project-Based Knowledge Maps

Chouyin Hsu [1] Duen-Ren Liu [2]

[1,2]Institute of Information Management, National Chiao Tung University, Hsinchu, Taiwan
[1]Overseas Chinese Institute of Technology, Taichung, Taiwan
hkelly@iim.nctu.edu.tw *dliu@iim.nctu.edu.tw*

Abstract

Managing enterprise knowledge for decision support is crucial for enterprises to gain competitive advantages in knowledge-based economy. The valuable knowledge patterns hidden in numerous projects are important assets of enterprises. The management of such project knowledge is becoming increasingly important and challenging for organizational adaptation and survival in the face of continuous environmental change. This work proposes a project-based knowledge map framework to capture project knowledge and discover valuable knowledge patterns from previous projects. A collaborative two-phase data mining approach is applied to extract valuable project attributes, and discover their associations. Moreover, the discovered knowledge patterns are organized in a well-structured knowledge map, which facilitates effective navigation of project knowledge.

1. Introduction

Knowledge management is crucial to organizational adaptation and survival in the face of continuous environmental change [8]. The knowledge acquisition, storage and distribution activities in a knowledge management system enable the dynamic creation and maintenance of an enterprise's intelligence [7][8].

Knowledge management has successfully been applied in many business domains. Bolloju et al. [5] proposed an integrative model for building enterprise decision support environments using model marts and model warehouses as knowledge repositories. Massey et al. [10] proposed to reengineer the customer relationship by acquiring and disseminating knowledge to both customers and IBM's human experts. Moreover, metadata were used as a knowledge management tool for supporting user access to spatial data [13]. Rubenstein-Montano [11] surveyed knowledge-based information systems for urban planning and suggested the importance for moving towards knowledge management. The effectiveness of knowledge management has been demonstrated in these applications.

Recently, knowledge management has been considered in the field of project management. Tah et al. [12] applied knowledge management technology to identify project risk and to further improve project management. Barthès et al. [2] developed an agent-supported portal to organize knowledge in complex R&D projects. However, these applications overlook the valuable knowledge patterns and working experiences hidden in numerous projects.

Generally, a project is a 'temporary' endeavor undertaken to create a particular product or service [14]. Unstable and temporary cooperation among a project team causes several difficulties in integrating knowledge, since the project team is usually disbanded and reorganized for another new project. This kind of volatile relationship hinders the accumulation of project knowledge. Moreover, a project can be carried out at different levels of the organization or across organizations. This complex scope results in difficulty in collecting integrated project knowledge.

Consequently, knowledge support is highly required throughout project development to solve these dilemmas. The numerous historical projects are the important knowledge source. The advance of data mining techniques has inspired applications in different problem-solving domains [4][6]. Applying data mining techniques to discover various hidden knowledge is a challenge for knowledge management [7]. Therefore, data mining approach is applied here to discover project knowledge.

This work proposes a project-based knowledge map framework to discover valuable knowledge patterns (project knowledge) hidden in projects, and to integrate these discovered patterns in a well-structured knowledge map. A collaborative two-phase data mining approach is applied to extract valuable project attributes, and discover their associations. The proposed knowledge map, clearly structured and semantically expressed, not only supports effective management of discovered knowledge patterns, but also assists users in navigating project knowledge to support further project development.

This paper is organized as follows. Section 2 describes a system model for discovering project-based knowledge maps. Section 3 illustrates meta information that pertains to a project. Section 4 discusses the mining approach to discover project knowledge. Section 5 elucidates the structure of the knowledge map. Conclusions and future work are finally made in Section 6.

2. System model

This section illustrates the system model for discovering project-based knowledge maps. As shown in Figure 1, three collaborative processes work together to construct the knowledge map, including extracting project meta information, mining project knowledge and finally deploying the project-based knowledge map.

First, the meta information builder is employed to extract meta information during project processing, including initialing, planning, executing, controlling and closing processes. Project meta information refers to the

The Second International Conference on Electronic Business
Taipei, Taiwan, December 10-13, 2002

important project attributes and annotations. When a project is finished to archive, the meta information is collected as well. The meta information builder appropriately annotates project attributes and converts these attributes to a consistent format for further analysis.

Second, a collaborative two-phase data mining approach is applied to extract valuable project attributes, and discover their associations. The first phase employs clustering methods to group projects into clusters according to their similarity. The second phase employs association rule mining to discover the inner knowledge patterns of the cluster of related projects, such as associations among project attributes. The discovered patterns reveal ontology-subject aspect and project domain concepts, which are very important in improving project development.

Third, the knowledge map builder integrates the discovered knowledge patterns and constructs a knowledge map to provide an information portal for accessing project knowledge.

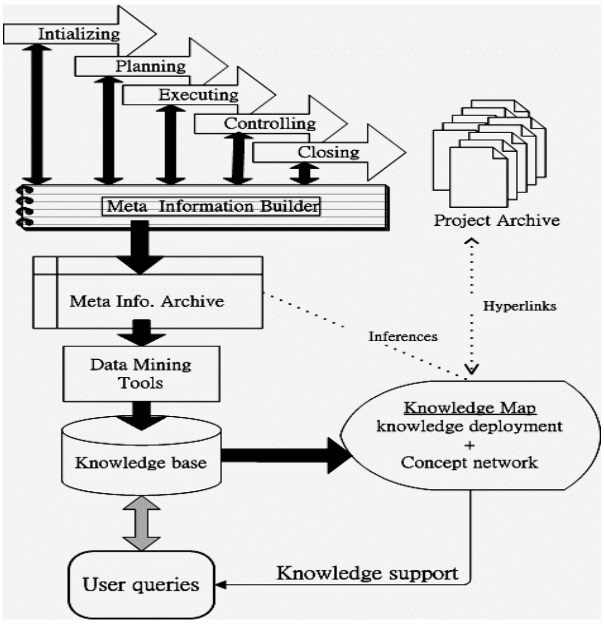

Figure 1. The system model of discovering project-based knowledge maps

3. Project meta information

A set of categories is used to classify project attributes. These categories may be pre-defined by human experts or generated ontology-based subjects for organizing project attributes consistently. This work considers five categories for software projects, Member, Tools, Activity, Goals and Cost, according to IEEE Standard for Software Project Management Plans (SPMP, IEEE Std. 1058-1998) [15].

Member : a list of the key workers in a project.
Tool : a list of major skills applied in a project.
Activity : a list of key actions in a project.
Goal : a list of objectives in a project.
Cost : the financial cost and working weeks for a project.

Table 1 presents a set of example software projects. The set of project attributes, grouped by above defined categories, constitutes the multiple project meta information. Accordingly, a vector model [3] is applied to represent these high-dimensional meta data.

Table 1. The collection of project meta information

	m1	m2	m3	m4	m5	t1	t2	t3	t4	..	g1	g2	g3	g4	g5	a1	a2	a3	..	c1	c2
P001	0	1	1	1	1	0	0	1	1	..	1	1	1	0	0	1	1	1	..	110	8
P002	0	0	1	1	1	0	1	1	1	..	1	1	1	1	1	1	1	1	..	40	6
P003	0	1	0	0	0	1	1	0	0	..	0	0	1	1	1	0	1	0	..	100	10
….																					
P100	0	1	0	1	1	0	0	0	1	..	1	1	0	0	0	1	1	0	..	60	7
P101	0	1	1	0	1	0	0	1	0	..	1	1	1	1	0	1	1	1	..	40	5
…..																					
Pn	1	1	0	0	0	1	1	1	0	..	0	0	0	0	1	0	0	0	..	10	2

==

Member : m1, m2, m3, m4, m5 denote Bob, David, John, Kim, Mary
Tool : t1,t2,t3,t4,t5,t6 denote Parallel, Java, Data Mining , XML, DBMS, CGI
Goal : g1,g2,g3,g4,g5 denote CRM, Logistics, Data Warehouse, Re-engineering, Wireless Communication
Activity : a1,a2,a3,a4,a5 denote Marketing, Agency, Consultant, Alliance, Contest
Cost : c1,c2, denote working weeks in total, total cost (Million dollars).

Vector model

Each project is annotated as a set of key attributes, and a vector model is used to represent the attributes of a project. A vector is defined as follows. A weight value $w_{ij} \geq 0$ denotes the importance of project attribute j on project i. 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. A project P_j is associated an attribute vector $P_j = (w_{j1}, w_{j2}, \ldots w_{jk})$.

According to the meta information given in Table 1, project P001 is represented by a vector model (0,1,1,1,1,0,0,1,1,1,1,1,1,1,0,0,1,1,1,0,0,110,8), implying that project P001 includes the key team members David, John, Kim and Mary; employs the tools, Data Mining, XML, DBMS and Java; is aimed at CRM, Logistics, Data Warehouse, Re-engineering and Wireless networks; and involves the activities of Marketing, Agency, Council, Alliance and Contest. The final two items imply that this project cost 8 million dollars over 110 working weeks.

4. Mining project knowledge

The mining of project knowledge contains two phases. The first phase employs clustering methods to group projects into clusters according to their similarity. The second phase employs association rule mining to discover the inner knowledge patterns such as associations among project attributes.

4.1 Clustering projects

This work employs an agglomerative algorithm that conducts hierarchical clustering to group projects in a bottom-up way. The agglomerative 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 [9]. The steps of the agglomerative algorithm include the following steps.

Constructing a dissimilarity matrix: The dissimilarity between projects depends on the distance between projects. This work uses Euclidean distance measurement to compute the proximity distance between i*th* and k*th* vectors. The pair-wise distance is useful to measure the dissimilarity and build a dissimilarity matrix for all attributes.

Forming clusters: The second step is to from clusters based on the dissimilarity matrix. A threshold value is an important parameter to decide how 'close' the projects will form a cluster. Projects with distance values less than a threshold value are grouped into the same cluster. For example, project P001, P100, P101 and P002 form a cluster labeled as 'Cluster#012', which is the main cluster example considered in the rest of this paper.

Determining Cluster centroid: A cluster of projects is a group of 'similar' projects, and a centroid is used to represent the key attributes of a cluster of projects. The centroid vector, $Cj = (r_{j1}, r_{j2}, ..., r_{jk})$ of a cluster j is determined according to the frequency (importance) of attributes. If the frequency (importance) of attribute i in cluster j is greater than a given threshold value, then r_{ji} is set to 1, otherwise it is set to zero.

The vector model is also used in the query module to find similar projects that match the partial concept or initial idea provided by users. The query includes multiple attributes from various categories, such as ' John, CRM', or a single attribute, such as ERP. Users could issue query using a partial concept or restricted conditions. A user query is represented as a vector model to indicate the query criteria of project attributes. The query result is determined according to the similarity between the query vector and the centroid vector. The similarity is quantified by the cosine of the angle between these two vectors. By ranking the similarity value, the system can find the most relevant cluster or find clusters with similarity greater than a given threshold value.

4.2 Mining associations among project attributes

Association rule mining is employed to discover relationships among project attributes. This work applies *Apriori* algorithm [1] to discover associations among project attributes. Association rule mining extracts association rules that satisfy a user-specified minimum support and confidence. The support for an association rule, X⇒Y, is defined as the percentage of project instances in cluster D that contain both attribute sets X and Y, while the confidence is defined as the proportion of project instances that contain attribute set X that also contain attribute set Y.

The *Apriori* algorithm discovers large itemsets by means of multiple passes over the data. An itemset is a set of project attributes. A large itemset is an itemset with support greater than minimum support. The algorithm is briefly illustrated as follows [1].

- In the first pass, *Apriori* counts the support of individual items and determine which of them have minimum support.
- Each subsequent pass starts with a seed set represented by the itemsets found to be large itemsets in the previous pass. From this set it generates the new potentially large itemsets, called candidate itemsets.
- At the end of the pass, it determines which of the candidate itemsets are actually large. This process continues until no large itemsets are found.

WEKA 3.0 software, an open source software issued under the GNU General Public License, is used for implementing *Apriori* algorithm [1]. Table 2 shows the mining result, and the support rules are explained as follows.

Table 2. The association rule of Cluster#012

```
WEKA 3.0.associations.Apriori
Minimum support: 0.5
Number of cycles performed: 189
=== Run information === …….
Best rules found:
  1. g3 ==> a3  …………………(1)
  2. a3 ==> g3
  3. a1 ==> g1  …………………(2)
  4. t5 g3 ==> a3  …………...(3)
  5. t5 a3 ==> g3
  6. t3 a1 ==> g1  ………………(4)
  7. t2 g3 ==> a3
  8. t2 a3 ==> g3  ……….……...(5)
  9. m5 t3 ==> g1  …….………….(6)
 10. m5 g1 ==> t3
 11. m3 t2 ==> t5  ……..…… …(7)
 12. m3 t5 ==> t2
 13. t2 t5 ==> m3
 14. m2 ==> m5  ……………….(8)
```

Support rules

<u>Support#1: Team-member dispatch</u>

A team of workers is the most commonly employed to complete a project. Based on rule (8) in Table 2, m2 and m5 are associated team members, as follows.

➢ *David and Mary are suitable team members.*

<u>Support #2: Complementary relations</u>

The relationship between cross-category attributes is useful to understand what kind of attribute complements each other. This kind of cross-category information is helpful to select suitable tools or appropriate activities to meet certain goals.

(i) *Data Warehouse* and *Consultant* are two associated concepts according to rule (1) in Table 2.

➢ *The Consultant activity works for the goal of Data Warehouse.*

(ii) *CRM* and *Marketing* are two associated concepts derived from rule (2).

➢ *Achieving the goal of CRM requires the Marketing activity.*

(iii) *Java*, *DBMS* and *Data Warehouse* are associated concepts derived from rules (3) and (5).

- *Java and DBMS tools assist in the goal of Data Warehouse.*

(iv) *Data Mining*, *CRM* and *Marketing* are associated concepts according to rule (4) in Table 2.

- *The Data Mining tool and Marketing activity assist in the goal of CRM.*

Support#3: Members' skills

Knowledge of members' skills is basic knowledge for supporting work assignments and determining what trainings needed. For example, rules (6) and (7) show members' skills.

- *Mary has the skill of Data Mining.*
- *John has the skills of Java programming and DBMS.*

Assigning team members, complementary relations among project attributes, and members' skills are very important tasks in managing projects. These support rules save much effort in initial planning of projects.

5. Deploying project-based Knowledge map

A knowledge map is developed to integrate the fragmental knowledge patterns discovered in the mining process. Such a map provides navigation facility to assist users in locating requested project knowledge.

The knowledge map developed here is a hierarchical structure that represents relationships among a cluster, categories, features, and objects. The cluster-node, at the first-level of the knowledge map, represents a cluster derived from the clustering analysis phase. The category-nodes, at the second level, denote pre-defined categories of project attributes. The feature nodes describe the important attributes of a cluster. The object nodes denote project resources.

Moreover, the map shows a particular object instance, which is the finance sheet, including average cost estimates and time taken. This is a notable consideration for project development, since time and cost are two important factors that require knowledge support.

A clear project-based knowledge map is shown in Figure 2 to illustrate the integrated knowledge patterns of the Cluster#012. This unique map provides the following supports for knowledge seekers.

(i) Domain concept: employed to provide a basic overview of relevant projects.

(ii) Information portal: used to access detailed project resources for reference.

(iii) Budget control: used to estimate the time taken and financial cost of a new project.

(iv) Efficient query: a clear summary of a group of related projects in response to a query.

Conceptual network for support rules

The support rules discovered from the association rule mining are integrated in a conceptual network as shown in Figure 3. A conceptual network contains two types of primitive element, nodes and links. Nodes are used to represent project features (attributes), while links with labels are used to describe the associations between project features.

The generated concepts may imply important project knowledge, which accumulates working experience of senior experts. The clear structure promotes the reuse of the discovered knowledge, and the conceptual network helps users to understand project knowledge.

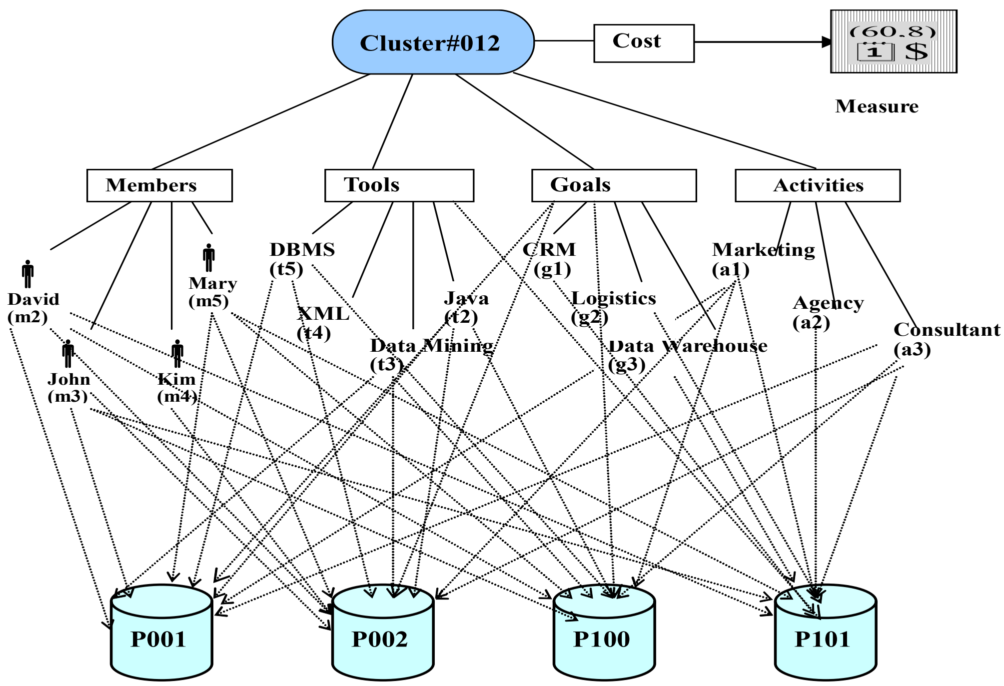

Figure 2. The knowledge map structure of Cluster#012

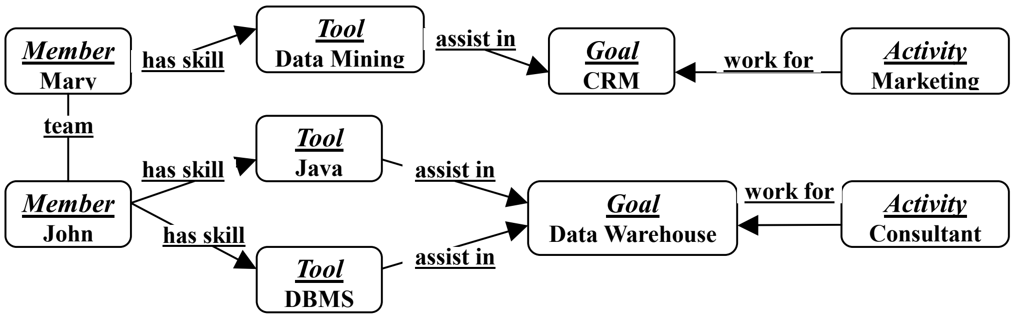

Figure 3. Conceptual network of Cluster#012

6. Conclusion and future work

The data mining approach is highly effective for extracting knowledge, but contributes less to knowledge sharing. The knowledge map, however, compensates for this lack. The proposed project-based knowledge map framework employs data mining approach to discover valuable project knowledge, and most importantly, integrates the discovered knowledge patterns in a knowledge map.

The approach provides effective knowledge support as follows. First, the knowledge map provides a simple but clear guide to clarify the distribution of project knowledge. Second, the extracted key attributes provide the project domain clearly. Third, each related project object is accessible by hyperlinks. Finally, the conceptual network captures important support rules discovered from the association rule mining.

The knowledge map is very helpful in supporting the management of project knowledge. Seeking innovative information technology to support other knowledge management tasks is an interesting future work. Moreover, the structure of the map must be further enhanced to increase the power of the knowledge representation. Finally, applying innovative data mining approaches in different phases to discover more valuable patterns remains a challenge.

Acknowledgements

This research was supported in part by the National Science Council of the Republic of China under the grant NSC 91-2416-H-009-008.

Reference

[1] R. Agrawal, T. Imielinski, A. N. Swami, Mining association rules between sets of items in large databases, *Proc. of the ACM SIGMOD Intl. Conf. on Management of Data*, Washington, May 1993.

[2] Jean-Paul A. Barthès, C. A. Tacla, Agent-supported portals and knowledge management in complex R&D projects, *Computers in Industry,* Vol: 48 (1), 2002, pp. 3-16

[3] Ricardo Baeza-Yates, Berthier Ribeiro-Neto, *Modern Information Retrieval*, Addison-Wesley, 1999,pp15-25

[4] M. Berry, G. Linoff. *Data mining techniques: for marketing, sales, and customer support*, Wiley Computer Pub., New York, 1997

[5] N. Bolloju, M. Khalifa, E. Turban, Integrating knowledge management into enterprise environments for the next generation decision support, *Decision Support Systems* Vol: 33 (2), June, 2002, pp. 163-176

[6] U.M. Fayyad, G. Piatetsky-Shapiro, P.Smyth, R. Uthurusamy, *Advances in Knowledge Discovery and Data Mining*, AAAI/MIT Press, 1996.

[7] Gerhard Fischer, Jonathan Ostwald, Knowledge Management: problems, promises, realities, and challenges, *IEEE Intelligent Systems*, Vol: 16(1), January/February 2001. pp. 60-72

[8] K. Harris, M. Fleming, R.Hunter, B. Rosser, A. Cushman, Gartner Group Report, The Knowledge Management Scenario : Trends and Directions for 1998-2003, Gartner Group Strategic Report 18, 1999

[9] Anil K. Jain, Richard C. Dubes, *Algorithms for clustering data*, Prentice Hall 1988, pp55-69

[10] Anne P. Massey, Mitzi M. Montoya-Weiss, Kent Holcom, Re-engineering the customer relationship: leveraging knowledge assets at IBM, *Decision Support Systems* Vol: 32 (2), December, 2001, pp. 155-170

[11] B. Rubenstein-Montano, A survey of knowledge-based information systems for urban planning: moving towards knowledge management, *Computers, Environment and Urban Systems*, 24 (2000), pp155-172

[12] J.H..M. Tah, V. Carr, Towards a framework for project risk knowledge management in the construction supply chain, *Advances in Engineering Software,* Vol: 32 (10-11), October - November, 2001, pp. 835-84

[13] Lawrence A. West Jr. , Traci J. Hess, Metadata as a knowledge management tool: supporting intelligent agent and end user access to spatial data, *Decision Support Systems,* Vol: 32 (3), January 2002 pp247– 264

[14] Project Management Institute, A guide to the Project Management Body of Knowledge, 2000 Edition.

[15] IEEE Std. 1058-1998, IEEE Standard for Software Project Management Plans (SPMP)