

A Mining-Based System Framework for Deploying Knowledge Maps of Composite E-Services

Jia-Yuan Lee, Chun-Feng Lee, Chih-Kun Ke, Duen-Ren Liu
Institute of Information Management
National Chiao Tung University
Hsinchu 300, Taiwan
E-mail: clay@iim.nctu.edu.tw

Abstract: Providing e-services and composite e-services on the Internet is an important trend of e-business. Composite e-services are complex processes which consist of various e-services provided by different e-service providers. In such complex environments, the flexibility and success of e-business depend on effective knowledge supports to access related information and resources of composite e-services. This work proposes a knowledge map platform to provide an effective knowledge support for utilizing composite e-services. A mining-based system framework is proposed to construct the knowledge map. Moreover, the proposed knowledge map is integrated with recommendation capability to provide users customized decision support in utilizing composite e-services.

Keywords: Dynamic E-service Discovery, Composition.

I. Introduction

With the explosive growth of Internet, more enterprises are providing various e-services for collaborative commerce online to achieve competitive advantages. Composite e-services are complex processes that require the cooperation among cross-organizational e-service providers. In such complex collaborative commerce environments, online user is facing the difficulty of how to select the appropriate composite e-services that suit their needs. Accordingly, an effective knowledge support system is essential to organize and access related information resources in e-service environments.

Many researches have focused on dynamic composition of e-services and system platforms to provide composite e-services [2][4][16][17]. Casti and Shan [4] proposed a model to compose e-services dynamically. Balakrishnan [2] proposed a Service Framework Specification to compose e-services. Several e-service platforms are proposed, for example, HP e-speak [10] and Microsoft .Net [15]. Basic features of these platforms are registering, advertising, monitoring, and managing e-services. However, conventional e-service platforms do not provide effective knowledge supports for managing information resources of composite e-services.

A knowledge map is a visual display of captured information and relationships, which enables efficient

communication and learning of knowledge [20]. Knowledge Maps have been used by enterprises to manage and navigate enterprises' explicit knowledge [5][7][12]. Accordingly, this work proposes a mining-based knowledge map platform to provide effective knowledge supports for utilizing information resources of composite e-services.

A data mining approach is employed to extract knowledge patterns from the usage records of composite e-services. The extracted knowledge patterns, which represent the important subjects and associations of composite e-services, form the kernel of the knowledge map. A topic map standard [11] was adopted to develop the proposed knowledge map, providing a bridge for managing and exchanging heterogeneous resources of composite e-services. Meanwhile, the proposed knowledge map is integrated with recommendation capability to generate recommendations of composite e-services via data mining and collaborative filtering techniques. The proposed knowledge map enhanced with recommendations can provide users customized decision support to effectively utilize composite e-services.

The rest of this paper is organized as follows; Section 2 introduces related work. Section 3 describes the proposed system framework. The architecture and functionality of the system are illustrated in Section 4. Section 5 presents the integration of recommendations. Section 6 demonstrates the prototype system. Conclusions and future works are finally made in section 7.

II. Related Work

Web services and composite e-services. Web services and e-services are similar but Web services place an emphasis on Web technologies. UDDI, WSDL, and SOAP, have been accepted as the de facto standards for Web services [20]. Composite e-services are composed by several e-services, and, intuitively, it can be seen as a workflow. Casati and Shan proposed a dynamic model for composing e-services [4]. Piccinelli and Williams [17] presented a DySCO infrastructure that uses a workflow model as a basis for composing Web services.

Knowledge maps. Successive knowledge map applications take advantages of the nature of visualization and navigation to locating and publishing knowledge [6]. Chung et al. [5] proposed a knowledge map framework for discovering business intelligence to alleviate information overload on the Web. Lin and Hsueh [14] applied information retrieval algorithms to generate and maintain a

knowledge map for virtual communities of practice. Kim et al. [12] proposed a road map to develop knowledge maps in the industrial community.

Topic Maps and XTM. The ISO standard ISO/IEC 13250 Topic Maps defines a model for the semantic structuring of link networks [11]. Topic Maps provide a bridge between the domains of knowledge representation and information management and link it to existing information resources. XML topic maps (XTM) [19] specified by TopicMaps.Org is an abstract model that uses XML grammar to express Web-based topic maps.

Recommendations. Collaborative filtering (CF) [3] has been successfully used in various applications [8][13][18]. The CF method utilizes preference ratings given by various users to determine recommendations to a target user based on the opinions of other similar users. Users' preference ratings are used to compute correlation coefficients among users. The correlation coefficient is a measure of the similarity between two different users' preferences.

Data mining. Data mining involves several tasks for different mining purposes, including association rule mining, clustering, classification, prediction, and time-series analysis [9]. K-means is a method commonly used to partition a set of data into K groups. Association rule mining aims to find an association between two sets of products in the transaction database [1]. The support of an association rule indicates how frequently that rule applies to the data. The apriori algorithm [1] is typically used to find association rules by discovering frequent itemsets.

III. System Framework of Kmap Platform

The proposed knowledge map (Kmap) platform includes two main sub-systems: knowledge map system and recommender system, as shown in Fig. 1.

III.1 Knowledge map system

The system includes four main modules.

Service collecting and planning. This module collects e-service information provided by various e-service providers. Predefined composite e-services are stored in the system. Moreover, the module generates service plan templates to assist users sketch a composite e-service.

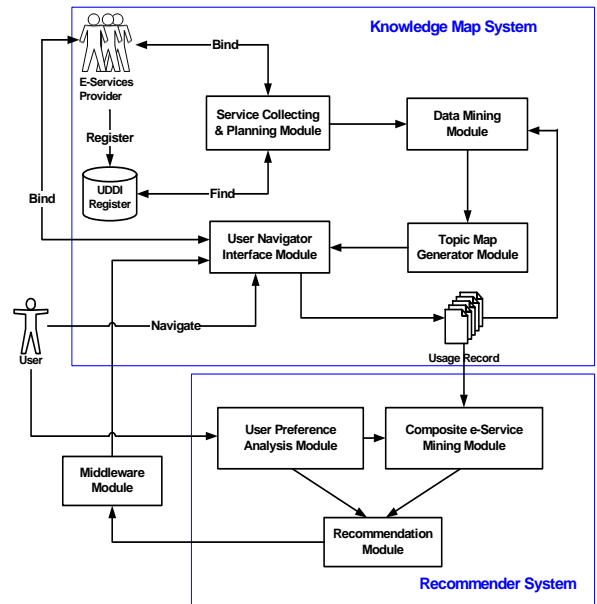
- **Data mining.** Data mining techniques are employed to discover valuable knowledge patterns from usage records and meta-information of composite e-services. The usage records log the historical executions of composite e-services conducted by various users. Important subjects (topics) and their associations of composite e-services are discovered via clustering and association rule mining techniques.

- **Topic Maps generator.** To facilitate knowledge exchange and sharing, XTM (XML topic maps) is employed to construct the knowledge map. The discovered knowledge patterns, including the valuable

subjects and their associations, are used to form the topics and topic associations of the knowledge map.

- **User navigator interface.** This module provides a navigator for user to browse composite e-services and related knowledge patterns by using the XTM-based interface.

FIG. 1. System framework of the Kmap platform



III.2 Recommender system

The proposed Kmap system can recommend the frequent attributes of each basic e-service, the ordering between e-services, and the composite e-services. Group-based recommendations are supported via analyzing the usage behaviors of interest-groups. Collaborative filtering is employed to recommending e-service providers for selected e-services.

- **User preference analysis.** This module uses *pearson* correlation coefficient to compute the similarity between users based on user's preference ratings on e-services served by various providers. The *k-NN* approach is used to derive a user's neighborhood (similar users). Moreover, a clustering approach is employed to cluster customers into interest-groups based on similarity measures.
- **Composite e-service mining.** This module uses association rule mining approach to discover *frequent attributes* of each basic e-service and *frequent ordering* between e-services from the instance execution logs. Moreover, the mining can also be conducted via considering the usage records of interest-groups to discover group-based frequent attributes and frequent orderings.
- **Composite e-service recommendation.** A scoring approach is used to recommend the top N composite e-

services according to the mining result. Group-based recommendations are achieved based on the mining result discovered from the usage records of interest-groups. Moreover, collaborative filtering (CF) approach is used to recommend e-service providers for the selected e-services. The CF approach makes recommendations via considering the preferences of neighborhood (similar users).

Middleware. The module is a bridge between the knowledge map system and recommender system. The module delivers recommendations generated by the recommender system to the navigator, while the user is browsing the knowledge map.

IV. Knowledge Map System

Fig. 2 shows the modules for deploying knowledge maps. This section introduces each module in the knowledge map system with details.

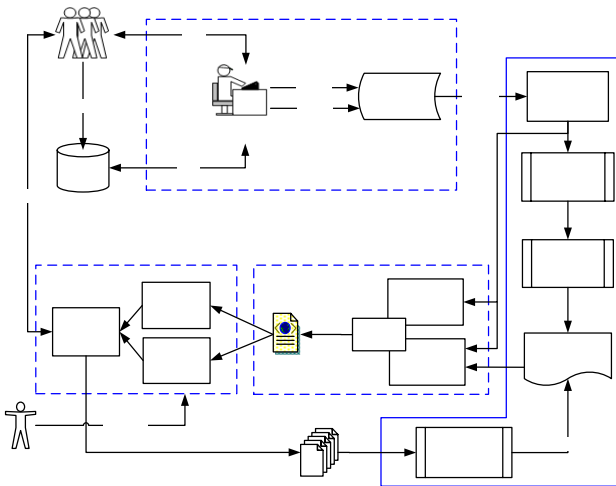


FIG. 2. Knowledge map system

IV. 1 Service collecting and planning

The module provides WSDL documents of Web services to illustrate the services and functions provided in each Web service. E-service and composite e-service information provided by the e-service provider are bind into the service database. Service plan designer can design a composite e-service by composing different e-services registered in the database. Attributes of these e-services and composite e-services are important information resources in data mining module for knowledge discovery.

IV. 2 Knowledge map template

Based on the information in the service database, attributes (meta information) of e-service and composite e-service are extracted for further discovery of knowledge patterns. Knowledge map templates are designed to record

the discovered knowledge patterns.

Meta information. Meta information is information about data which describes e-services and composite e-services. For the rest of this work, we use training programs, which contain series of training-courses (e-services), as the examples of composite e-services to illustrate the proposed work. A training-course generally consists of the following attributes: course name, provider, category, difficulty, location, time period, instructor, hours, and cost. MCS D (Microsoft Certified Solution Developers) is a certification of passing the exams for C programming, SQL server, and so on. The MCS D training program (composite e-service) includes a series of training courses (e-services) for taking MCS D exams.

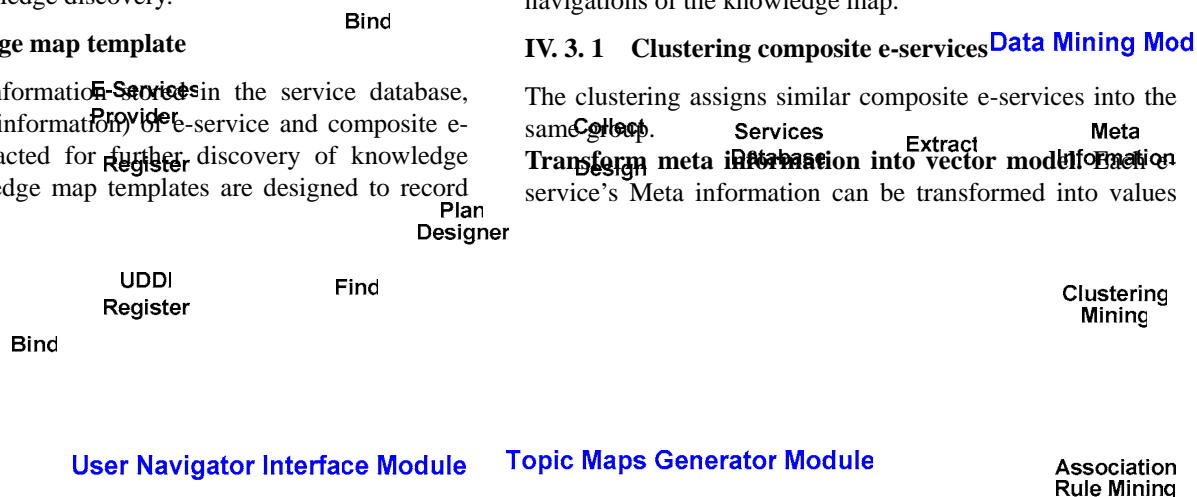
Topic Maps template. We use XML Topic Maps (XTM) [19] to construct the core structure of knowledge map templates. The general attributes for e-services are classified as follows – ServiceName, Provider, Category, Location, Cost and Rating. Moreover, extendable attributes of e-services corresponding to training courses consist of Difficulty, Instructor, TimePeriod, and Hours. Each class of the attributes is defined as a topic type in XTM. The relationships among topics are defined as associations. Several types of associations are defined as follows - Sequence: ordering relationship among e-services; Contain: structural relationship between a composite e-service and its e-services. The association type - RelateTo is defined to link e-services or composite e-services to an attribute. An association type - Relevance is defined to record the association patterns discovered from the mining results.

IV. 3 Discovering knowledge patterns

A two-phase data mining process is employed to discover valuable knowledge patters from usage records and meta-information. The usage records log the historical executions of composite e-services conducted by various users. Clustering technique is used to group composite e-services based on the similarity measure derived from meta information. Important topics (attributes) of each cluster (group) are extracted to represent the meta information of each cluster. Association rule mining is then employed to discover attribute associations (relevant relationship) in each cluster. The topics and associations form the kernel of the Kmap system. The “Relevance” associations between topics are used to recommend relevant topics during the navigations of the knowledge map.

IV. 3. 1 Clustering composite e-services

The clustering assigns similar composite e-services into the same group. The meta information of each composite service’s Meta information can be transformed into values



represented in vector space model. A composite e-service's vector contains the meta information of a composite e-service and the union of all the vectors of basic e-services included in the composite e-service.

Calculate the dissimilarity. This work uses Euclidean distance measurement to compute the proximity distance between pair-wise vectors. The Euclidean distance measures the dissimilarity between two composite e-services.

Form clusters. The K-Means clustering technique is employed to perform clustering in which Euclidean distance is used to measure the dissimilarity between composite e-services.

Determine the important topics. A cluster centroid is extracted to represent the major attributes of a cluster of composite e-services. The frequently appearing attributes will form the cluster centroid. The extracted important attributes of a cluster centroid form the topics in the discovered knowledge map of composite e-services.

IV. 3. 2 Mining association patterns

This work employs *apriori* algorithm to find association patterns, namely frequent topic sets, from the usage records and meta information of composite e-services within a cluster.

Association patterns of basic e-services. Such association patterns are discovered via applying association rule mining to the usage records of e-services. The association patterns mainly consist of associations between attributes which are derived from the frequent attribute sets of the mining result. Each attribute corresponds to a topic with a defined topic type in XTM. The associations among several types of attributes can be discovered. The association: Instructor "Nancy" \leftrightarrow Difficulty "Basic" implies that the instructor "Nancy" often lectures courses with "Basic" difficulty.

Association patterns among composite e-services. Such association patterns are discovered via applying association rule mining to the usage records of composite e-services. The association patterns mainly consist of associations between attributes or e-services which are derived from the frequent attribute sets of the mining result.

IV. 3. 3 Recommendations by association patterns

The association patterns imply relevance relationships between topics, and can be used to recommend relevant topics during users' navigations of the knowledge map.

IV. 4 Topic Maps generator

This module generates the knowledge map of composite e-services in XTM documents based on the mining results. The important attributes of basic and composite e-services extracted from each cluster form the topics in XTM. Association patterns discovered from association rule mining form the topic associations. The association type "Relevance" is used to record the association patterns discovered from the mining results. The association type

represents the relevance relationship between e-services or service attributes. Several kinds of attribute associations can be discovered via association rule mining, as described in Section 4.3.2.

V. Recommender System

The recommender system implements two approaches to generate recommendations: collaborative filtering and association rule mining.

Group-based recommendations. The recommendation can be conducted by considering interest-groups to provide group-based recommendations. Customers with similar preferences are grouped into interest-groups. For a customer u , the group-based recommendation proceeds as follows. The instance execution logs of u 's interest-group are used to discover the frequent attribute sets of e-services and frequent ordering sets between e-services. Such discovered frequent attribute and ordering sets are then used to derive the scores of composite e-services and make recommendations.

E-service provider recommendations. Collaborative filtering (CF) approach is used to recommend e-service providers for the selected e-services. The CF approach finds the k -nearest neighbors (most similar users) of u , and then makes recommendations via considering the k -nearest neighbors' preferences on e-service providers and their similarity to u . Through the user's k -nearest neighbors, the system can predict user's preference on an e-service served by a particular provider.

VI. System Implementation and Demonstration

The implementation is conducted using several software tools, including ASP .NET(C#), JSP, Microsoft Visual Studio .NET and Borland J-Builder. Web server is setup on Microsoft IIS 6.0 and Apache Tomcat 5. Microsoft SQL server 2000 is used as the database system for storing related data of e-services and composite e-services. Microsoft UDDI Service is used as the UDDI engine for e-service search and registrations. Moreover, the data mining tool - Weka 3.4 is employed to discover the knowledge patterns as well as frequent attribute and ordering sets of composite e-services. The XTM files for storing the discovered knowledge patterns including topics and associations in each cluster is generated according to the XTM defined in section 4.2 and 4.4. ASP.NET is employed to develop the Web-based navigator which allows users to browse and navigate the XTM information of composite e-services. The Kmap system is integrated with the recommendation capability to provide more effective knowledge support for managing and browsing composite e-services.

System demonstration. The Kmap navigator shows the clusters of composite e-services with attributes of cluster

centroid attached. Once a user selects a cluster to browse, the attribute values of the selected cluster can be browsed further in the navigator. User can browse the topics in the navigator through topics of clusters, "RelateTo" and "Relevance" associations. Once a user selects an e-service "Programming a MS SQL Server 2000 Database", the χ map system displays the detailed information, as shown in Fig. 3. The relevant e-services are also suggested to provide further navigations of relevant e-services. When a user Mike logs in, the system identifies Mike's interest-group. Mike browses the cluster A of composite e-services to decide needed services. The navigator provides the recommended composite e-services which are derived from the usage records of composite e-services conducted by Mike's interest group. Mike can click on any e-service in selected composite e-service to obtain the details (frequent attribute values) of the clicked e-service. The system can also suggest providers to serve the e-service via the collaborative filtering approach.

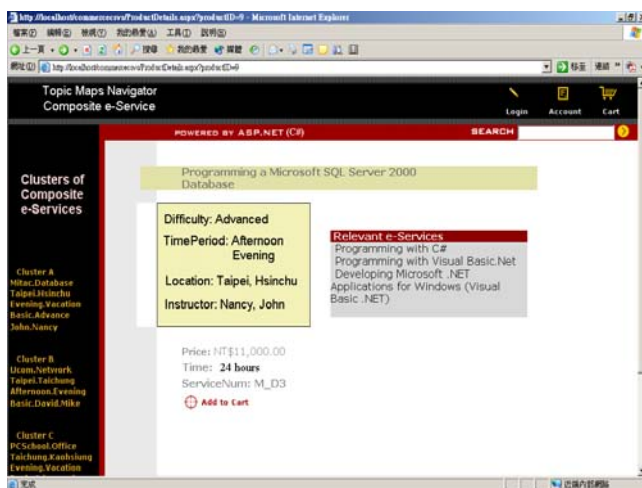


FIG. 3. Navigation with relevant e-services

VII. Conclusions

This work mainly develops a χ map system to provide knowledge supports for browsing and managing composite e-services. Data mining techniques are employed to discover valuable knowledge patterns of composite e-services. The discovered important subjects and association patterns are used as the kernel to generate the knowledge map. This work employs the XML topic maps to construct the knowledge maps of composite e-services. Consequently, the knowledge maps of composite e-services, expressed in XML Topic Maps, can be exchanged and interoperated more easily on the Internet. Effective collaborations of conducting composite e-services over the Internet can thus be facilitated. Moreover, the proposed system provides integrated browsing and recommendations of composite e-services. The proposed knowledge map enhanced with recommend-

ations can provide user customized decision support to effectively utilize composite e-services.

Acknowledgements

This research was supported by the National Science Council of Taiwan under the grant NSC 92-2416-H-009-010.

References

- [1] R. Agrawal, T. Imielinski, and A. N. Swami, "Mining Association Rules between Sets of Items in Large Databases", *Proc. of the 1993 ACM SIGMOD Intl Conf. on Management of Data*, 1993.
- [2] R. Balakrishnan, "A Service Framework Specification for Dynamic E-Services Interaction", *Proc. of the 4th Intl Enterprise Distributed Objects Computing Conf. (EDOC'00)*, 2000.
- [3] J. S. Breese, D. Heckerman, C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", *Proc. of the 14th Conference on Uncertainty in Artificial Intelligence*, pp. 43-52, 1998.
- [4] F. Casati and M.-C. Shan, "Dynamic and Adaptive Composition of E-Services", *Information Systems*, 26(3), pp. 143-163, 2001.
- [5] W. Chung, H. Chen, J. Nunamaker, "Business Intelligence Explorer: A Knowledge Map Framework for Discovering Business Intelligence on the Web", *Proc. of the 36th Hawaii Intl. Conf. on System Science*, 2003.
- [6] T. H. Davenport, L. Prusak, *Working Knowledge: How Organizations Manage What They Know*, Harvard Business School Press, Boston, MA, 1998.
- [7] M. J. Eppler, "Making Knowledge Visible Through Intranet Knowledge Maps: Concepts, Elements, Cases", *Proc. of the 34th Hawaii Intl. Conf. on System Sciences*, 2001.
- [8] W. Hill, L. Stead, M. Rosenstein, G. Furnas, "Recommending and evaluating choices in a virtual community of use", *Proc. of the ACM CHI'95*, 1995.
- [9] J. Han, M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, 2000.
- [10] HP (Hewlett Packard), E-Speak, <http://www.e-speak.hp.com/>.
- [11] ISO (International Organization for standardization), International Electro-technical Commission (IEC), "ISO/IEC 13250:2000, Topic Maps", Available <http://www.iso.org>.
- [12] S. Kim, E. Suh, H. Hwang, "Building the Knowledge Map: An Industrial Case Study", *Journal of Knowledge Management*, Vol. 7, No. 2, pp. 34-45, 2003.
- [13] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and R. Riedl, "GroupLens: Applying Collaborative Filtering to Usenet News", *Communications of the ACM*, 40(3), pp. 77-87, 1997.
- [14] F.-R. Lin, C.-M. Hsueh, "Knowledge Map Creation and Maintenance for Virtual Communities of Practice", *Proc. of the 36th Hawaii Intl. Conf. on System Sciences*, 2003.
- [15] Microsoft, .Net, <http://www.microsoft.com/net>.
- [16] G. Piccinelli, G.-D. Vitantonio, and L. Mokrushin, "Dynamic Service Aggregation in Electronic Marketplaces", *Computer Networks*, 37(2), pp. 95-109, 2001.
- [17] G. Piccinelli and S. L. Williams, "Workflow: A Language for Composing Web Services", W.M.P. van der Aalst et al. (Eds.): *BPM 2003*, LNCS 2678, 2003.
- [18] U. Shardanand, P. Maes, "Social information filtering: algorithms for automating world of mouth", *Proc. of the ACM (CHI'95)*, pp. 210-217, 1995.
- [19] TopicMaps.Org, XML Topic Maps (XTM) 1.0, <http://www.topicmaps.org/xtm/index.html>
- [20] A. Tsalgatidou and T. Pilioura, "An Overview of Standards and Related Technology in Web Services", *Distributed and Parallel Databases*, 12(2-3), pp. 135-162, 2002.
- [21] E.F. Vail, "Mapping Organizational Knowledge", *Knowledge Management Review*, Issue 8, May/June, pp. 10-15, 1999.