2012 International Conference on Industrial and Intelligent Information (ICIII 2012)
IPCSIT vol.31 (2012) © (2012) IACSIT Press, Singapore

Development of a Technology for Knowledge Evolution Course Discovery in a Professional Virtual Community

Yuh-Jen Chen[+]

Department of Accounting and Information Systems
National Kaohsiung First University of Science and Technology
Kaohsiung, Taiwan, ROC

Abstract. This work develops a technology for knowledge evolution course (KEC) discovery in professional virtual communities as a decision support mechanism to discover effectively the KEC hidden inside of a professional virtual community, which can guide community members to retrieve required empirical knowledge quickly. This objective can be obtained by performing the following tasks: (i) design of an empirical knowledge management framework for professional virtual communities, (ii) definition of a KEC model, (iii) design of a KEC discovery process, (iv) development of techniques related to the technology for KEC discovery.

Keywords: Professional virtual community, empirical knowledge, knowledge evolution, knowledge reuse

1. Introduction

As an effective means of retrieving and representing human knowledge, domain ontology can deal with scattered empirical knowledge in a professional virtual community in terms of classification, hierarchy, and relation to relieving information overloading, ultimately facilitating knowledge reasoning and problem solving. Ontology techniques have been increasingly adopted to define an entity, attribute and relationship among knowledge concepts within a specific domain using explicit descriptions and specifications that present an interoperable format understandable by both humans and machines, thereby realizing knowledge sharing and reuse. For instance, Lee and Jian (2003) [1] developed an ontology-based automatic summarization mechanism for web news. Quan et al. (2006) [2] proposed an ontology-based support platform for semantic inquiry. Chen (2007) [3] designed a knowledge construction mechanism for semantic searching. However, recent studies focus mainly on classifying and constructing domain knowledge concepts and a question-related searching instead of concentrating on knowledge evolution within empirical knowledge from dialogue among experts in professional virtual communities. Therefore, the inability to manage and reuse empirical knowledge in professional virtual communities effectively makes knowledge demanders impossible to retrieve required empirical knowledge.

This work presents a technology for KEC discovery in professional virtual communities as a decision support mechanism to discover effectively the KEC hidden inside of a professional virtual community, which can guide community members to retrieve quickly required empirical knowledge. To achieve this objective. Firstly, an empirical knowledge management framework is designed based on the empirical knowledge communication model and the knowledge management method. Subsequently, the conversation structure for community members is initially identified by performing conversation analysis. The developmental track among discussion topics is then defined through exploring interactive information and sharing process from participants. The defined developmental track is represented using the ontology method

[+] Corresponding author. Tel.: + 886-6011000 ext. 4316; fax: +886-7-6011158
E-mail address: yjchen@nkfust.edu.tw

for facilitating empirical knowledge sharing in a professional virtual community. Finally, the formal concept analysis, the probabilistic latent semantic analysis, and the Dijkstra's shortest path algorithm are adopted to develop the techniques associated with the discovery of KEC.

2. Design of Empirical Knowledge Management Framework for Professional Virtual Communities

Knowledge experts in a professional virtual community capture required information and knowledge of each other through dialogue to accumulate self-empirical knowledge. Such dialogue contents become important sources for accumulating empirical knowledge.

To effectively manage unorganized but already expressed knowledge contents among experts through a systematic means of managing tacit knowledge, this section introduces an empirical knowledge management framework for professional virtual communities, as shown in Fig. 1. Effectiveness of the proposed framework is illustrated by a knowledge management life cycle, capable of extracting, verifying, constructing, and reasoning empirical knowledge.

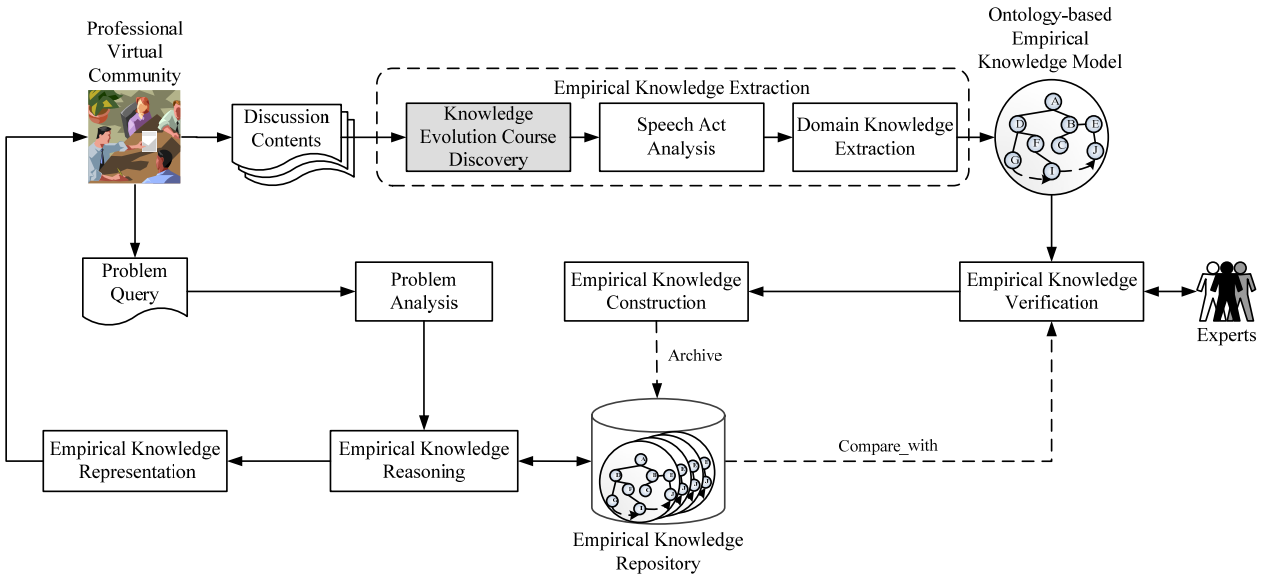

Fig. 1: Empirical Knowledge Management Framework for Professional Virtual Communities.

3. Design of KEC Discovery Scenario for Professional Virtual Communities

This section first analyzes the structure of a discussion topic for professional virtual communities as well as defines relations between topics by using the conversation analysis to establish a model for an ontology-based KEC. Based on the established model, the process of KEC discovery is then conducted.

3.1. Definition of KEC Model

This section primarily defines a KEC model, which involves analyzing the KEC structure, defining KEC relations, and designing the concept schema of KEC, as described in the following subsections.

3.1.1. Analysis of KEC Structure

A common method for constructing knowledge in professional virtual communities classifies discussion topics to provide a communication environment that is specific and with a common goal. For instance, "fund investment" can be viewed as a topic with two sub-topics according to the scope of the financial product, i.e. "domestic funds" and "offshore funds", respectively. If selected, the sub-topic "offshore funds" can be further classified into "emerging market funds" and "energy funds" based on its operational subject. These classified topics are subsequently listed into different forums, allowing community members to express a certain topic to initiate discussion in order to stimulate responses from other members. These dialogue contents become the main source for empirical knowledge in professional virtual communities.

Conversation analysis emphasizes the natural recorded dialogue of a real event, for use in analyzing topics such as “how to manage dialogue”, “how to talk in turn”, “what is the relation between words”, and “how to guide, develop, and change certain discussion topics”. Therefore, based on the three concepts of “adjacency pair”, “turn-taking” and “discussion topic” from the perspective of conversation analysis [4, 5], this study examines the involvement of each community member and their interaction with others within a professional virtual community in order to establish a KEC structure for professional virtual communities.

3.1.2. Definition of KEC Relations

According to Fig. 2, the relations [6, 7] within the established KEC structure are defined to assist community members understanding empirical knowledge. These relations include “hierarchy relation”, “mutual relation”, and “core topic relation”, as introduced as below.

(1) Hierarchy relation: A partial ordering of concept types by the type-subtype relation. Empirical knowledge in professional virtual communities can be classified effectively based on the hierarchy relation.

(2) Mutual relation: Certain similarity exists between concepts. By using the mutual relation, the strength of the relation between concepts is comprehended as well as the expanding and changing course between discussion topics is observed.

(3) Core topic relation: Relation between domain keywords and their topic concepts. Through this relation, domain keywords can search for their core-topic concepts.

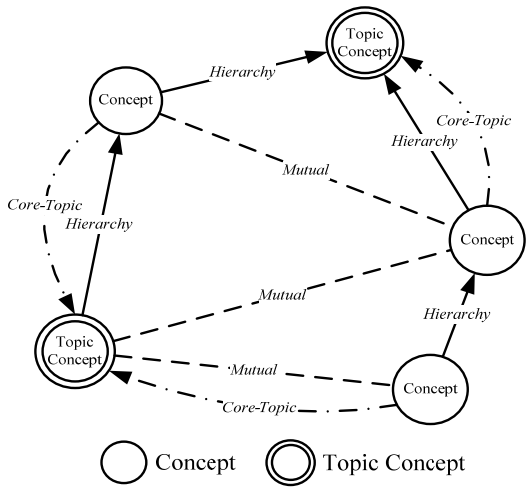

Fig. 2: Relations between Concepts in KEC.

3.1.3. Design of Concept Schema for KEC

For concepts in a KEC, this section designs a concept schema for KEC based on object-oriented techniques to facilitate the establishment of a KEC model. According to the results in Fig. 3, the concept schema for KEC consists of three elements [6, 7], i.e. concept, attribute, and relation.

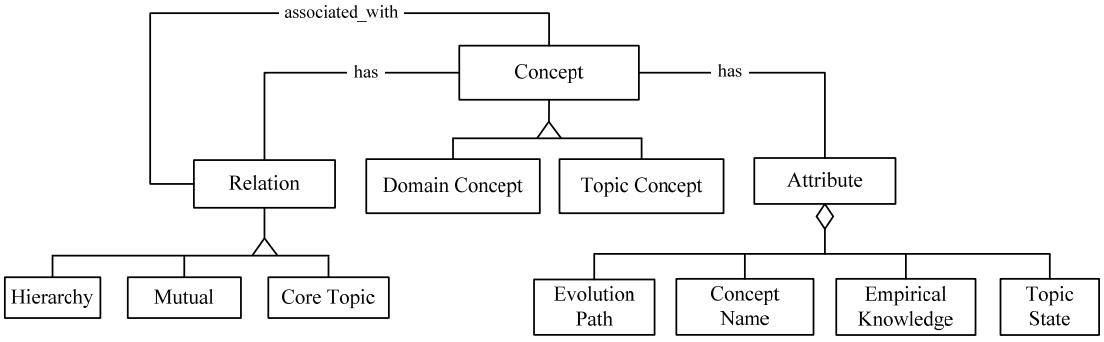

Fig. 3: Concept Schema of KEC.

3.2. Establishment of KEC Discovery Process

Based on structure analysis results, relation definition and concept schema design of KEC in Section 3.1, this section introduces a knowledge evolution discovery process for professional virtual communities to facilitate the acquisition of KEC for the expanding and changing of a discussion topic, as depicted in Fig. 4. KEC discovery as described below.

(1) Topic classification: Topics for discussion documents in a professional virtual community are inducted.

(2) Domain dictionary construction

   (a) POS tagging: Based on the document unit set, keywords for related discussion domains are determined and Parts-Of-Speech (POS) for each keyword is analyzed by conducting sentence breaking and word tagging.

   (b) Term-pair combination: The segmented keywords must be recombined through the term-pair combination to ensure the accuracy of domain keywords.

   (c) Domain keyword filtering: The Term Frequency-Inverse Document Frequency (TF-IDF) can assist experts in filtering the obtained domain keywords to ensure the accuracy of a domain dictionary.

(3) Ontology-based topic empirical knowledge model construction

   (a) Concept-set generation: Based on the inducted document units and the extracted domain dictionary, the concept set of discussion domain is analyzed by performing Formal Concept Analysis (FCA).

   (b) Hierarchy-relation generation: the mutual inclusion relation between the concept set and the document set is analyzed to establish a hierarchical classification structure, ultimately forming an empirical ontology.

   (c) Mutual-relation calculation: The strength degrees for similarity and intersection between concept sets are analyzed to facilitate the analysis of evolution course for concept expanding and changing during experts' discussion process.

(4) Topic concept extraction and representation: The document units and the domain dictionary are utilized as the data sources. The most common discussion topic concepts and their related domain keywords in a discussion domain are then identified based on Probabilistic Latent Semantic Analysis (PLSA).

(5) Path establishment between topic concepts: The shortest path between topic concepts is constructed through the shortest-path searching method to present the evolution course of topic expanding and changing during discussion among community experts, ultimately predicting accurately the required empirical knowledge for community members.

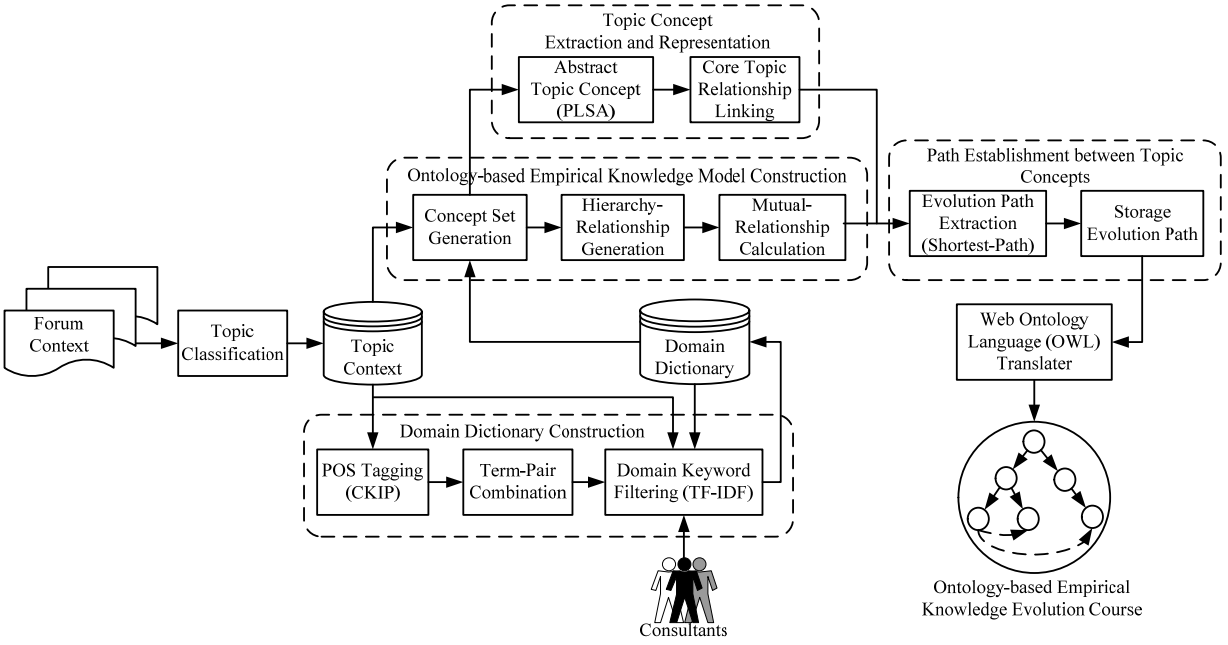

Fig. 4: KEC Discovery Process.

4. Development of KEC Discovery Techniques

Based on the designed procedure of KEC discovery for professional virtual communities in Section 3, this section develops the techniques involved in discovery of KEC to discover automatically the evolution course of topic empirical knowledge from documents in a community forum. These crucial techniques include topic classification, domain dictionary construction, ontology-based topic empirical knowledge model construction, topic concept extraction and representation, as well as path establishment between topic concepts.

(1) Topic Classification: Analyzing the KEC structure in Section 3.1.1 reveals that a speech from a community member focuses mainly on a certain subject and a topic speech. Therefore, a situation in which a speaker making a speech aims at the same document can be treated as the same document unit, and then attach the contents of responding documents to the document created by the speaker, allowing one to obtain topic concepts hidden inside of document contents in (4).

(2) Domain Dictionary Construction: Domain dictionary construction primarily involves three steps of POS tagging, term-pair combination, and domain keyword filtering. Meanwhile, the term-pair combination and the domain keyword are conducted to filter domain concepts through the assistance of experts in order to acquire a professional domain dictionary that is suitable for a certain discussion domain.

(3) Construction of Ontology-based Topic Empirical Knowledge Model: Given the ability of the Formal Concept Analysis (FCA) to extract useful concept hierarchy from a data set and use it as the basis for ontology development [8, 9], this section utilizes formal concept analysis to construct concept hierarchy relations based on the document units inducted from (1) and the domain dictionary established from (2). Additionally, the course of expanding and changing between two topic concepts is observed by adopting the concept inter-relations generation method [10] to determine the degree of intersection between two concept topics in order to redeem the inability to present a partial concept intersection through formal concept analysis.

(4) Topic Concept Extraction and Representation: Probabilistic Latent Semantic Analysis (PLSA) [11, 12] is a novel statistical technique for the analysis of two-mode and co-occurrence data, which has applications in information retrieval and filtering, natural language processing, machine learning from text, and in related areas. Thus, this section using the Probabilistic Latent Semantic Analysis (PLSA) to identify topic concepts in the generated document units in (1) as well as relevant concepts attached under one topic concept.

(5) Path Establishment between Topic Concepts: As mentioned in Section 3.2, the course of empirical knowledge evolution can be treated as the expanding or changing of a topic concept in discussion. Therefore, the mutual relation obtained from (3) is viewed as the path-weight between concepts to display the evolution track of discussion topic. Multiple paths likely exist between two concepts. The shortest path between two random topic concepts is estimated using Dijkstra's algorithm [13, 14] for use as the evolution path for expanding the concept, ultimately providing topic evolution references for community members. The shortest path between topic concepts is calculated based on the relation between topic concepts and domain concepts in (4) as well as the mutual relation in (3).

5. Conclusions and Perspectives

This work developed a technology for KEC discovery for professional virtual communities to retrieve and represent the KEC in a professional virtual community. The proposed technology can also orient knowledge demanders on how to retrieve required empirical knowledge quickly.

Results of this study significantly contribute to efforts to achieve KEC discovery in order to accurately predict and supply required empirical knowledge for future knowledge demanders, shorten empirical knowledge searching time, increase empirical knowledge reusability, and ultimately reach empirical knowledge sharing in an enterprise.

6. Acknowledgements

The authors would like to thank the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract Nos. NSC98-2221-E-327-039 and NSC99-2221-E-327-036.

7. References

[1] C. S. Lee, Z. W. Jian, Ontology-based fuzzy event extraction agent for Chinese e-news summarization, Expert Systems with Applications. 2003, 25(3): 431-447.

[2] T. T. Quan, S. C. Hui, A. C. M. Fong, Automatic fuzzy ontology generation for semantic help-desk support, IEEE Transactions on Industrial Informatics. 2006, 2(3): 155-164.

[3] C. L. Chen, Partially constructed knowledge for semantic query, in Engineering Science, National Cheng Kung University, 2007, pp. 58-60.

[4] D. Schiffrin, Approaches to discourse: language as social interaction, Blackwell, 1994.

[5] R. Wardhaugh, Introduction to sociolinguistics, Blackwell, 1997.

[6] Y. J. Chen, Y. M. Chen, Y. S. Su, C. C. Wen, Ontology-based distributed case-based reasoning in virtual enterprises, International Journal of Software Engineering and Knowledge Engineering, 2009, 19(8): 1039-1082.

[7] Y. J. Chen, Y. M. Chen, H. C. Chu, Development of a mechanism for ontology-based product lifecycle knowledge integration, Expert Systems with Applications, 2009, 36(2): 2759-2779.

[8] B. Ganter, R. Wille, Formal concept analysis: mathematical foundations, Berlin: Springer-Verlag, 1998.

[9] R. Wille, Conceptual graphs and formal concept analysis, in ICCS' 97, 1997, pp. 290-303.

[10] S. S. Weng, H. J. Tsai, S. C. Liu, C. H. Hsu, Ontology construction for information classification, Expert System with Applications. 2006, 31(1): 1-12.

[11] L. Cai, T. Hofmann, Text categorization by boosting automatically extracted concepts, in Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, 2003, pp. 182-189.

[12] T. Hofmann, Probabilistic latent semantic analysis, in Proceedings of the National Academy of Science, 1999, pp. 289-296.

[13] J. A. Bondy, Graph theory with applications, Elsevier Science Ltd, 1976.

[14] G. Chartrand, O. R. Oellermann, Applied and algorithmic graph theory, Mcgraw-Hill College, 1992.