Manuscript submitted to

Journal

A Framework for Designing Knowledge Map using Text-mining method

December 2002

Intae Kang

• Yongtae Park

School of Engineering

Department of Industrial Engineering

Seoul National University

San 56-1, Shillim-Dong, Kwanak-Gu, Seoul 151-742, KOREA

• TEL: 82-2-880-8358, FAX: 82-2-889-8560, E-Mail: **parkyt@cybernet.snu.ac.kr**

It is confirmed that this item has not been published nor is currently being submitted elsewhere.

A Framework for Designing Knowledge Map using Text-mining method

Keywords *Knowledge Map, Text-mining, Structured knowledge, Unstructured knowledge, Knowledge Management*

Abstract *Knowledge map is a representation tool to visualize knowledge sources and relationships among knowledge artifacts, which is considered as a core element of knowledge management system. To construct a knowledge map, we need to identify which category each newly registered knowledge artifact is mapped into, called structured knowledge. However, in view of fully utilizing obtainable knowledge, it is necessary to consider existing knowledge artifacts that are stored in previously established DB but not categorized. Thus an approach to incorporate unstructured knowledge as well as structured one into knowledge map turns out to be critical. With this end in view, we identified relationship between structured knowledge and unstructured knowledge using text-mining method. In our framework, we extract keywords set from each knowledge artifact by means of text-mining and compare it with each other to calculate similarity between knowledge artifacts. With this similarity, we can determine which knowledge artifacts are mapped into a certain location of knowledge map. To embody our framework, we also designed technology based knowledge map and developed web-based system in the case of a system integration company.*

1. Instruction

With the growth and evolution of an organization, it is necessary to manage multidisciplinary and huge amount of knowledge [3]. Thus, we need a method to classify and categorize knowledge artifacts generated and stored in an organization. From this point of view, *knowledge map*, a representation tool to visualize knowledge sources and relationships among knowledge artifacts, attracts more and more attention from both practitioners and academicians [4]. However, so far the research outcomes of knowledge map mainly focus on its abstract concept and explanation. But to construct a real knowledge map and incorporate it into *KMS* (*knowledge management system*), we need more detailed and concrete framework to develop knowledge map based on the needs of users and developers of KMS.

Construction of a knowledge map is generally composed of two phases; building up map structure and mapping knowledge artifacts into it. First, in accordance with the necessities of an organization, the structure of knowledge map can be organized by various criteria. For example knowledge map can take form of business model, workflow, technology tree, or etc. Once the structure of a knowledge map is established, we need to identify into which location knowledge artifacts are mapped. In other words, we should assign each knowledge artifact to specific category of knowledge map previously defined. In case of registration of new knowledge artifact, we can appoint its category and map it into knowledge map. This newly registered knowledge with assigned category is called *structured knowledge* in this research. However, for full utilization of obtainable knowledge, it is necessary to consider existing knowledge artifacts stored in knowledge repository established beforehand, but not designated to a certain category as well. In this study, we call this obtainable but not categorized existing knowledge *unstructured knowledge.* To classify and assign unstructured knowledge, the relationship between structured knowledge and unstructured knowledge is required to be identified. To this end, we developed a framework based-on text-mining method to design knowledge map encompassing not only structured knowledge but also unstructured knowledge. In our framework, keywords sets of each knowledge artifact are obtained by means of text-mining, and they are compared with each other to calculate similarity between them. With this similarity, we can assign unstructured knowledge to a location of associated structured knowledge. Using our framework, we can flexibly incorporate existing knowledge items into knowledge map of various forms. To embody this framework, we developed web-based knowledge map system based on it, called TUM (text-mining used map), which is for the case of a system integration company.

2. Background

2.1 Research on knowledge map

The terms knowledge map is defined as a visual architecture of knowledge domain that enables us to examine the knowledge on a global scale and from different perspectives [6]. For knowledge users' perspective, it is a basic tool to retrieve necessary knowledge and to analyze the relationship among knowledge artifacts, and for the case of knowledge managers, it is an essential method to grasp the status of knowledge and plan knowledge management strategies [4]. However, so far, the majority of studies on knowledge map have been done from the point of knowledge managers and/or system developers, rather than of knowledge users. Therefore, the research outcomes of knowledge map mainly focus on its abstract concept or programming issues. In other words, they are done at too macro level to be applied to practical use or at too micro level to provide practical guidelines for the developers or users knowledge map.

Eppler [6], Gaines [8], Gomez [10], and Gordon[11] suggested conceptual definition and structure of knowledge map. Eppler classified knowledge maps into five types that are knowledge source map, knowledge asset map, knowledge structure map, knowledge application map, and knowledge development map. Gaines proposed a concept mapping tool that facilitates visualizing the relationships among actor, activity, objective and material in dispersed learning communities. Gomez suggested a knowledge map that can be used at the stage of holistic testing in KMS implementation. He defined knowledge map as 'a type of mental diagram by means of which complex ideas can be easily and quickly set out in a logical order' and proposed framework to present knowledge map through relationships of artifacts of knowledge and related attributes. Gordon uses learning dependency and semantic network to organize knowledge map and drew it on the ideas of what knowledge is and on spatial representation structures. Using this knowledge map, he identified relationships between knowledge artifacts, such as subordination and inclusion.

In line with development and utilization of knowledge map Bargent[2], Godbout [9], Levine [13], and Speel[16] respectively suggested concept of a specific knowledge map and framework to design and develop it. Bargent suggested an approach following a typical life-cycle method to build general knowledge map, in which 11 steps such as identifying requirements, conduction an information audit, and etc. are involved. Godbout presented a model to filter knowledge based on such crucial factors as actor relevance factor, technical relevance factor, authority relevance factor, fidelity factor, and scientific acceptability factor. Levine proposed an approach based on contract (ABC) method, in which the contract between firm and its client is used as the base of business and knowledge modeling. These frameworks, albeit useful and meaningful, are limited to designing internal process of business and thus are not sufficient to integrate business modeling and knowledge management in a global context. Speel presented a set of practical techniques to develop knowledge map based on CommonKADS, which is composed of organization model and task model. The knowledge map proposed is for discovering knowledge strengths and weaknesses not for relationships between knowledge artifacts.

2.2 Text-mining

Text-mining which is also known as text data mining or knowledge discovery from textual databases refers to the process of extracting interesting and non-trivial patterns or knowledge from text documents. Using text-mining, we can perform semantic information retrieval or focus text exploration around a certain subject.

Atkinson[1] and Tan[17] introduced the concept of text-mining and its possibility of applications. Atkinson introduced many possibilities for automatic analysis of text based on

Data Mining or text collection. Tan suggested a text-mining framework consisting of two components: Text refining that transforms unstructured text documents into an intermediate form; and Knowledge distillation that deduces patterns or knowledge from the intermediate form. He surveyed the state-of-the-art text-mining products/applications and aligned them based on the text refining and knowledge distillation functions as well as the intermediate form that they adopt.

Frameworks to apply text-mining to knowledge management are suggested by researchers like Feldman[7], Mack[14] and etc. Feldman proposed an approach, called text-mining at the term level, in which knowledge discovery takes place on a collection of words and phrases. Based on this approach, Document Explorer that is a tool to implement text-mining at the term level is also put forward by him. Mack suggests methods to search and mine life-sciences documents. His methods provided an interpretive context for understanding the meaning of biological data. Zhua[18] proposed an approach to automatically extract and visualize information for technological forecasting using technology maps and text-mining method.

3. Framework of knowledge map

3.1 Overall procedure to construct knowledge map

Massive amount of obtainable knowledge gives necessity to classification and categorization of knowledge artifacts by proper dimension. However, it is practically impossible for a person or a group to arrange every knowledge item by itself, since the available knowledge items are too abundant to be managed at hand as an organization grows and evolves. Thus, we developed a framework to match or group knowledge artifacts automatically in accordance with their contents, which is presented by and large in figure 1.

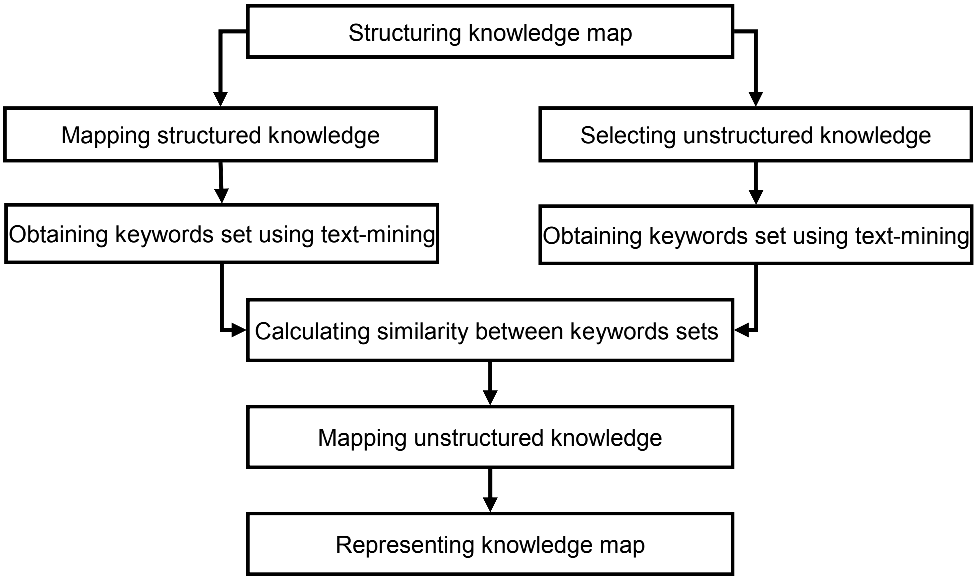

Figure 1. Overall framework to construct knowledge map

As depicted in figure 1, initially the structure of knowledge map is organized along with the needs of an organization, and newly registered knowledge artifact is mapped manually; we call this manually designated knowledge as *structured knowledge* in this research. And then, we obtain keywords set of each structured knowledge artifact using text-mining tool. In general, there may be copious already registered but not assigned knowledge items in an established knowledge repository of an organization, which are called *unstructured knowledge* in this paper. We also get keywords set of each unstructured knowledge artifact by means of text-mining tool. Afterward, similarity between the keywords set of structured knowledge which is already mapped and that of unstructured knowledge is calculated. With computed similarity, we can map unstructured knowledge to an appropriate location in which related structured knowledge is positioned. Finally, complete knowledge map, which covers both structured and unstructured knowledge items, is represented to knowledge user.

3.2 Structuring knowledge map

Structure of knowledge map can be organized in various ways to satisfy the needs of a firm, such as *technology-based knowledge map* concerning the technologies to design value, *product-based knowledge map* to capture value itself, *workflow-based knowledge map* centered upon the tasks to produce value, and *business model (BM) based knowledge map* focused on value transaction at market suggested by Kang et al [12]. As a matter of convenience, we are going to take up technology-based knowledge map for the knowledge map structure in this research. Technology-based knowledge map put forward in this paper originates at technology tree shaped in hierarchical form depicted in figure 2.

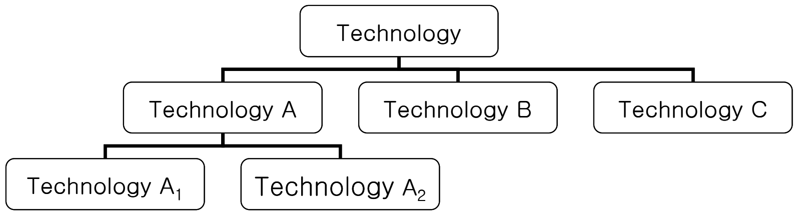

Figure 2. Technology tree

Once the structure of knowledge map is confirmed, it is required to allocate each knowledge artifact into a proper location. However, it is almost impossible to assign every single knowledge artifact to a certain position, since usually knowledge artifacts in an organization are too many in its number and scattered into various knowledge sources. In this regard, we consider only a newly registered knowledge artifact in this initiating mapping step and deal with existing knowledge artifacts generated before hand and stored in knowledge repository in next steps. As depicted in table 1 with its DB schema, a structured knowledge artifact is registered with its location, related technology, in the map as well as its general features like register, registered date, name and related document.

Table 1. DB schema of structured knowledge

Column Code	Column Name	Data Typpe	PK	Comment
Knowledge ID	s_id	varchar(5)	✓	ID composed of 1~5 characters
Knowledge Name	s_name	varchar(30)		Name of knowledge
Document	s_document	varchar(60)		Text file attached to knowledge artifact
Technology	s_technology	varchar(5)		ID of technology in which a knowledge artifact is located
Register	s_register	varchar(30)		A person in charge of a registered knowledge artifact
Date	s_date	varchar(20)		Date when a knowledge artifact is registered
Keyword Set	u_keywords	text		Keywords set obtained by text-mining method

3.3 Obtaining keywords set using text-mining

If a knowledge artifact is an electronic document based on text data, we can obtain keywords set of each knowledge artifact using text-mining tool. To this end, *TextAnalyst*, one of commercialized text-mining tools, is applied to this research, which is developed by *Megaputer Intelligence Inc.* (http://www.megaputer.com) TextAnalyst allows users to analyze large volumes of textual information, summarize, efficiently navigate and cluster documents in a text base. It can provide the ability to perform semantic information retrieval or focus the text exploration around a certain subject. It is based on an integration of a unique linguistic and a neural network that ensures high speed and accuracy in the analysis of unstructured texts. Using TextAnalyst, we obtains summary and keywords set from an electronic document, and insert these into a relevant record of DB that stores the knowledge artifact. Figure 3 shows an exemplary keywords set of a technological report concerning on-line music forecast. The numeric value left to each keyword in figure 3 is its semantic weight that indicates importance of a word. Using these weights of keywords, we can filter keywords set to exclude trivial ones of which weights are not significant. Thus we excluded obtained keywords of which weights are under 80 from keywords set to be stored, and due to this filtering manner each keywords set may have different number of keywords.

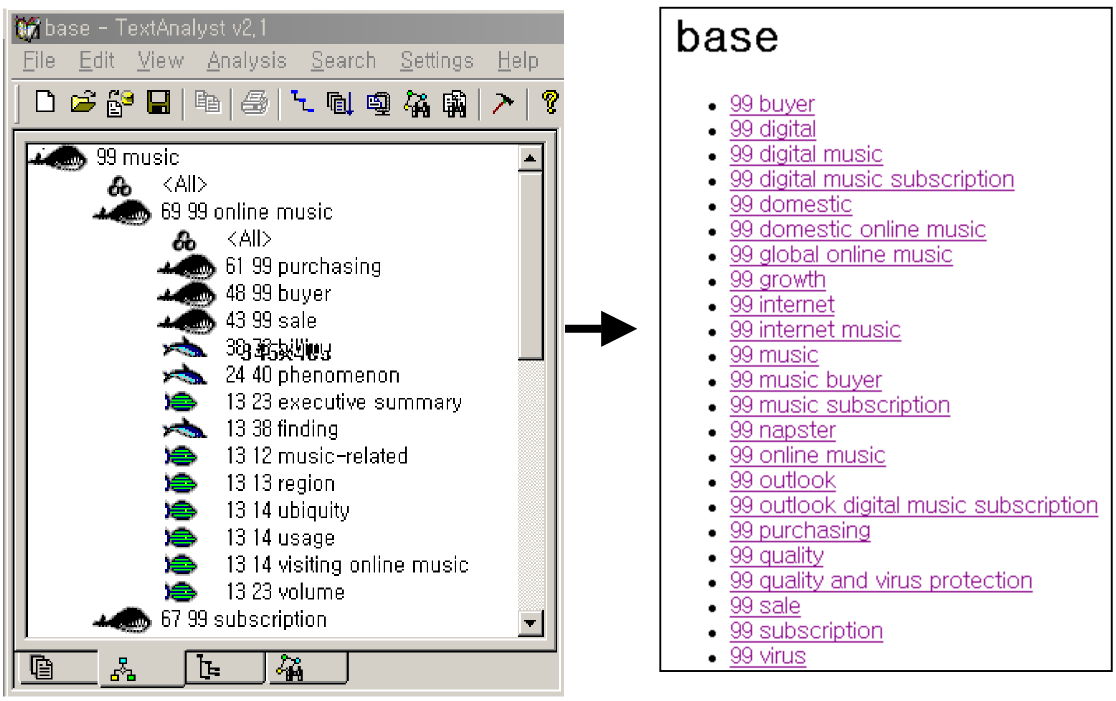

Figure 3. Obtained keywords set using TextAnalyst

To incorporate obtained keywords set into the DB schema of unstructured knowledge, we need additional column named *u_keywords*. Thus, unstructured knowledge in this research has the DB schema described in table 2. In the same way, DB schema of structured knowledge also needs to be altered to store keywords set obtained with text-mining tool, TextAnalyst.

Table 2. DB schema of unstructured knowledge

Column Code	Column Name	Data Type	PK	Comment
Knowledge ID	u_id	varchar(5)	✓	ID composed of 1~5 characters
Knowledge Name	u_name	varchar(30)		Name of knowledge
Document	u_document	varchar(60)		Text file which is attached to knowledge artifact
Register	u_register	varchar(30)		A person who is in charge of a registered knowledge artifact
Date	u_date	varchar(20)		Date when a knowledge artifact is registered
Keyword Set	u_keywords	text		Keywords set obtained by text-mining method

3.4 Calculating similarity between knowledge artifacts and mapping unstructured knowledge

Once we obtain keywords set of each knowledge artifact, we need to calculate the *similarity* between two knowledge artifacts. To this end, similarity between knowledge artifacts has to be defined quantitatively and acquired automatically. As previously pointed out, text-mining tool can keywords set from a knowledge artifact and the similarity is to be calculated based on the derived keywords set in this research. In this way, using text-mining method, we can develop knowledge map based on the contextual similarity between knowledge artifacts.

Although similarity between knowledge artifacts can be defined in various ways in accordance with purposes or methods of developing KMS, we identified the definition of similarity *as the number of same keywords* in two keywords sets in this research. For example, let us assume that we obtain following 3 keywords sets, *A*, *B*, and *C* from 3 knowledge artifacts A, B and C respectively based on text-mining method.

A = {Extranet, Internet, Intranet, Linux, operation system, Unix, virtual machine, Windows}

B = {data, digital contents, entertainment, game, information, Internet, music, movie, TV}
C = {Extranet, Internet, Java, JSP, operation system, platform, Sun, Windows}

In this case, the similarity between A and B is 1 while that between A and C is 4, then we can say that knowledge artifact C is more similar to A than B. Figure 4 shows the pseudo code to calculate the similarity between two knowledge artifacts.

```
S_keyNum = the number of keywords of a structured knowledge artifact
U_keyNum = the number of keywords of an unstructured knowledge artifact

For (i = 1 to S_keyNum) {
        S_key[i] = i'th keywords of the structured knowledge artifact
        For (j = 1 to U_keyNum) {
                U_key[j] = j'th keywords of the unstructured knowledge artifact
                If (S_key[i] = u_key[j]) {
                        similarity = similarity + 1
                }
                j = j + 1
        }
        i = i + 1
}
```

Figure 4. Pseudo code to calculate the similarity between two knowledge artifacts

3.5 Mapping knowledge artifacts into knowledge map

Using the previously mentioned similarity, we can map related unstructured knowledge artifacts that have similarity above a specific threshold into a proper location. That is, we can find the unstructured knowledge artifacts associated with a certain structured knowledge artifact, and map those ones into a location in which that structured knowledge artifact is placed. Therefore, we need a filtering rule to differentiate more related knowledge from less related one.

Since each knowledge artifact has different number of keywords, it is not proper to apply absolute number to filtering less related knowledge artifacts. Thus we adopt *relative number of keywords*, which is a proportion of total number of a knowledge artifact's keywords computed with following expression.

$$(\textit{ relative number of keywords }) = (\textit{ total number of keywords }) * (\textit{ user defined ratio })$$

For example, if knowledge artifact A has 50 keywords and designated ratio is 30%, then relative number of keywords is 15. Then, we map unstructured knowledge artifacts that have more than 15 similar keywords into a proper location. The pseudo code that select related knowledge artifacts from knowledge repository is depicted in figure 5.

```
S_keyNum = the number of keywords of a structured knowledge artifact
T_ItemNum = the number of total unstructured knowledge artifacts
R_Num = relative number of keywords
ratio = ratio designated by user
R_Num = S_keyNum * ratio

For (k = 1 to T_ItemNum) {
        Similarity[k] = similarity of k'th knowledge artifact
        If (Similarity[k] > R_Num {
                Map k'th knowledge artifact
        }
        k = k + 1
}
```

Figure 5. Pseudo code to select related knowledge artifacts from knowledge repository

4. Development of system

To develop web-based knowledge map system using suggested framework, we assume the case of IT Company engaging in system integration. Knowledge artifacts incorporated into developed system are technological reports from external and internal source, and they are in the form of electronic document. Since applied text-mining tool, TextAnalyst, can only analyze the electronic document in text form (*.txt) and rich text format (*.rtf), we converted all documents into rich text format and stored in our knowledge repository.

4.1 Registration of structured knowledge

To construct knowledge map and retrieve a specific knowledge artifact, we first register structured knowledge into knowledge repository. Values of the records in DB schema of structured knowledge as described in table 1 are inserted into structured knowledge object manually or automatically. Process adding a certain knowledge artifact into knowledge repository is described in figure 6 with *sequence diagram* of UML (unified modeling language) [5] [15], which shows object interactions arranged in time sequence.

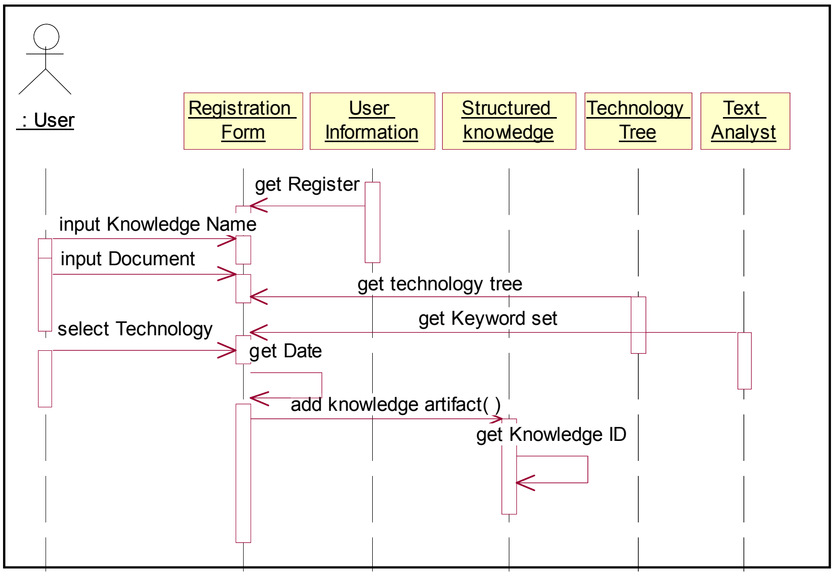

Figure 6. Sequence diagram to depict knowledge registration process

Registration of a structured knowledge artifact is fulfilled with following steps; *Register* is acquired from *User Information* system using login information; user manually inputs *Knowledge Name and Document*; *Technology* is manually selected by user from technology tree; *Date* and *Knowledge ID* are generated automatically by system. Thus all required information about a structured knowledge artifact is stored in knowledge repository. In relation with registration of structured knowledge, some captured screens of developed system are illustrated in figure 7.

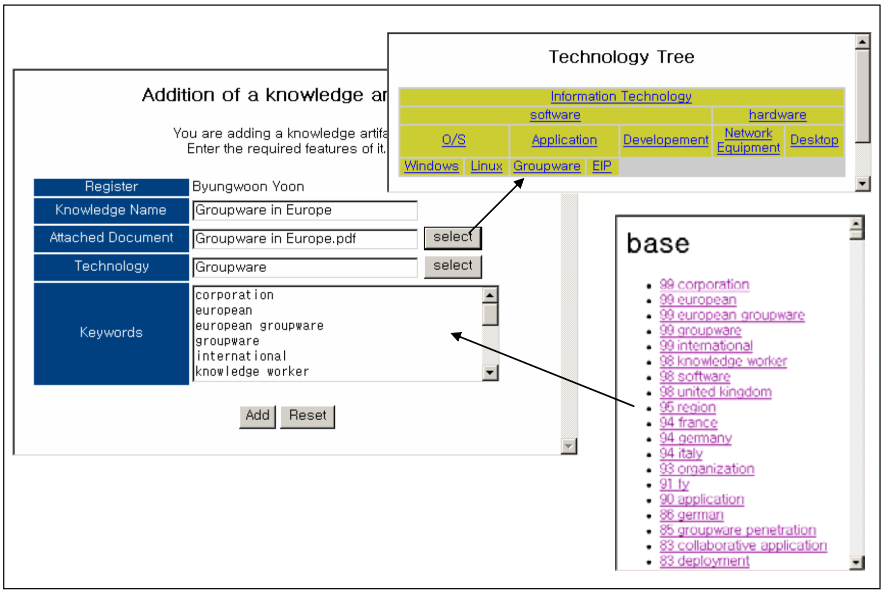

Figure 7. Captured screens to illustrate registration of knowledge artifact

4.2 Retrieval of a knowledge artifact

User's retrieval of a knowledge artifact follows the procedure described in figure 8 with sequence diagram. A user first selects a specific technology from technology tree presented by our system, and then the structured knowledge artifacts previously designated to a certain technology in registration process are represented to the user, which are called *related knowledge* in this research. And he or she can pick up one knowledge artifact from the listed knowledge artifacts. Structured knowledge can be retrieved in this way, since it is associated with a proper technology by user in its registration process.

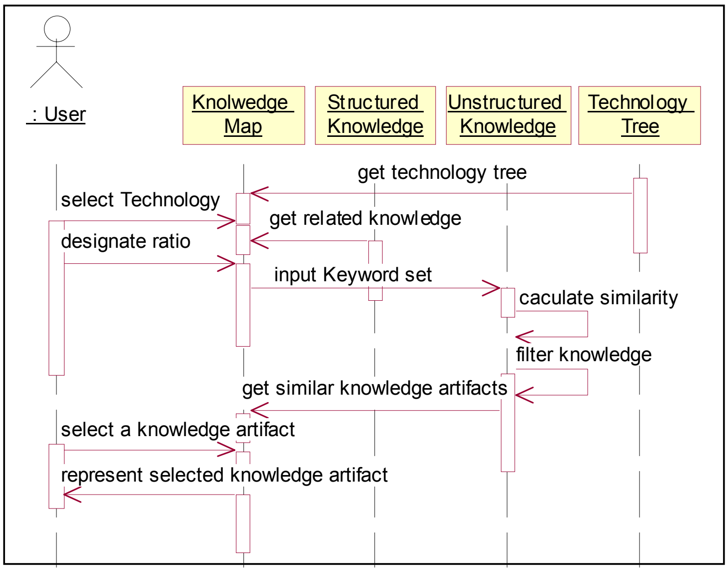

Figure 8. Sequence diagram to depict knowledge retrieval process

Usually, the user needs to retrieve unstructured knowledge artifacts similar to a specific structured knowledge artifact listed, which we call *similar knowledge* in this research. If then, he or she can do this with choosing certain knowledge artifact and designating ratio of similarity. Once a knowledge artifact is chosen and ratio of similarity is designated, our system calculates each unstructured knowledge artifact's similarity to the selected knowledge artifact, and picks the knowledge artifacts that have similarities over a certain value out of knowledge repository. Figure 9 shows retrieval practice with some screen captures that are represented to system user.

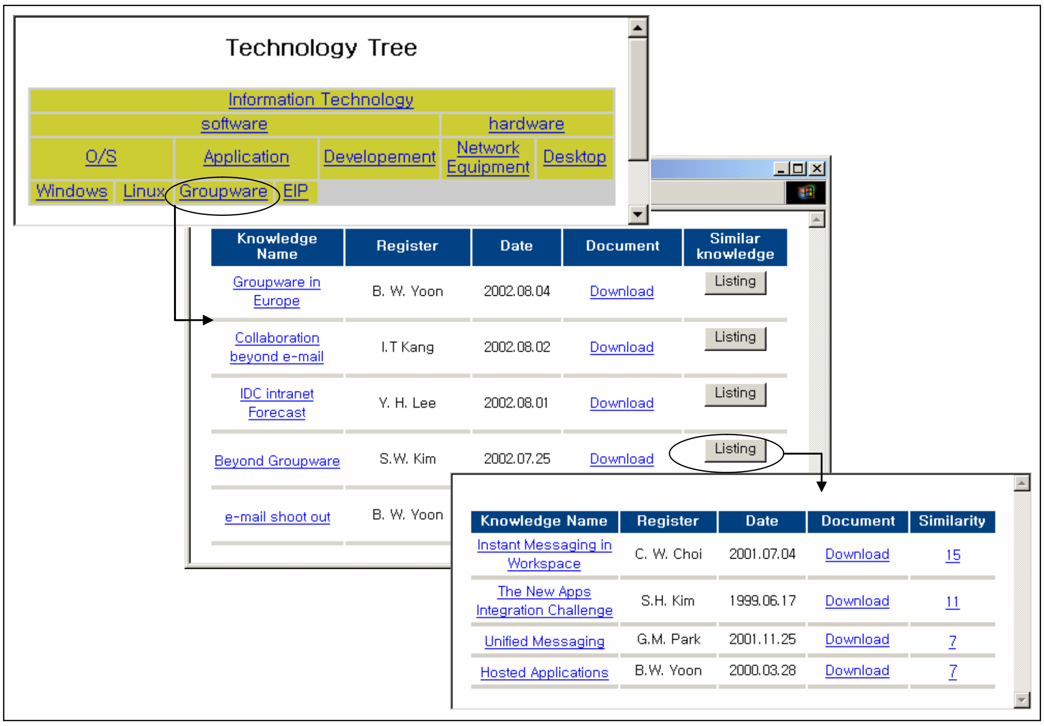

Figure 9. Captured screens to illustrate retrieval of knowledge artifact

5. Conclusion

To fully utilize knowledge source available, KMS needs to deal with existing knowledge repository as well as contemporary one newly registered into KMS. In that regard, the incorporation of existing knowledge artifacts into a knowledge map in KMS is as crucial as proper structure of knowledge map. In this research, we proposed a framework to incorporate these unstructured knowledge artifacts stored in existing knowledge repository into knowledge map, and materialized it with developing a web-based pilot system based on text-mining method. Using text-mining tool, we could obtain keywords set from a knowledge artifact and, comparing obtained keywords set with another, calculate similarity between knowledge artifacts. Based on this similarity we can get knowledge artifacts related with specific knowledge from existing knowledge repository. This framework can be considered as an application of text-mining method and an extension of conventional knowledge mapping approaches, to provide an approach to incorporate unstructured knowledge as well as structured one into knowledge map.

Although we believe this research can give a valuable contribution to developing knowledge map in both practical and academic setting, it is yet merely tentative and subject to some limitations. In particular, it is concerning only a part of whole knowledge management system or knowledge mapping approach. First, we considered only the text-based knowledge, however there exist huge amount of knowledge artifacts that don't take shape of electronic text. For example there may be many knowledge items in the shape of picture, off-line paper, and multi-media. Above all, tacit knowledge in employee's brain should be considered in constructing knowledge map, but we cannot take this into account in this study. Second, knowledge map needs to be personalized, since the amount of knowledge in a corporate is usually too large for an individual knowledge user to manage with ease. This personalization problem covers types of knowledge map, knowledge filtering, and security plans. We expect to solve this problem with developing various types of knowledge map like product based knowledge map, business model based one and workflow based one. In addition, to personalize knowledge map based on this relationships, we are going to develop a framework to relate user's department and role with technology, product, business model, and workflow.

References

1. Atkinson, J., "Text-mining: Principles and Applications", Revista Facultad De Ingenieria, U.T.A. (CHILE), vol.7, 2000
2. Bargent, J., "11 steps to building a knowledge map", [Online], Available: http://www.ibm.com, 2002
3. Hosapple, C. and Joshi, K., "Knowledge selection; Concept, Issues, and Technologies", in Jay Liebowitz, "Knowledge Management: Handbook", CRC Press LLC, 1999
4. Davenport, T. and Prusak, L. "Working Knowledge: How organization manage what they know", Harvard Business School Press, 1998
5. D'souza, D., "Objects, Components, and Frameworks with UML", Addison Wesley, 1999
6. Eppler, M., "Making Knowledge Visible through Intranet Knowledge Maps: Concepts, Elements, Cases", Proc. 34th Hawaii int'l Conf. System Sciences, IEEE, 2001
7. Feldman, R. et al., "Knowledge Management: A Text-mining Approach", Proc. the 2nd Int. Conf. on Practical Aspects of Knowledge Management Basel, Switzerland, 1998
8. Gaines and Shaw, M., "Collaboration through Concept Maps", Proc. Computer Supported Cooperative Learning, Mahwah, New Jersey: Lawrence Erlbaum, 1995
9. Godbout, "Filtering Knowledge: Changing Information into Knowledge Assets", Journal of Systemic Knowledge Management, [Online], Available: http://www.scoap.com/ki/articles/godbout/godbout03.htm, 1999
10. Gomez, A. Moreno, J. Pazos, and Sierra-Alonso, A., "Knowledge maps: An essential technique for conceptualization", Data and Knowledge Engineering, vol. 33, No 2, pp. 169-190, 2000

11. Gordon, J., “Creating knowledge maps by exploiting dependent relationships”, Knowledge-Based Systems, ELSEVIER, vol. 13, issue 2-3, pp. 71-79, 2000
12. Kang, I., Park, Y. and Kim, S., “Knowledge Maps for Innovation: Major Types and Integration”, Proc. Fourth International Conference on Operations and Quantitative Management, 2003
13. Levine, P., “From Business Modeling Based on the Semantics of Contracts to Knowledge Modeling and Management”, Proc. 34th Hawaii International Conference on System Sciences, IEEE, 2001
14. Mack, R. and Hehenberger, M., “Text-based knowledge discovery: search and mining of life-sciences documents”, DDT, vol. 7, No. 11, 2002
15. Quatrani, T., “Visual Modeling with rational rose and UML”, Addison Wesley Longman Inc., 1998
16. Speel, P., et al., “Knowledge mapping for industrial purposes”, KAW’99, Alberta, Canada, online-available: http://sern.ucalgary.ca/KSI/KAW/KAW99, 1998
17. Tan, A., “Predictive self-organizing networks for text categorization”, Proc. The Fifth Pacific-Asia Conference on Knowledge Discovery and Data Mining, 2001
18. Zhua, D. and Porter, A., “Automated extraction and visualization of information for technological intelligence and forecasting”, Technological Forecasting & Social Change, North Holland, 2002