Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Engineering 32 (2012) 1169 – 1177

Procedia Engineering

www.elsevier.com/locate/procedia

I-SEEC2011

Optimizing Knowledge Management using Knowledge Map

J. Watthananon[a*] and A. Mingkhwan[b**]

[a]Department of Information Technology, Faculty of Science and Technology, Suan Dusit Rajabhat University, Bangkok, 10300, Thailand

[b]Faculty of Industrial and Technology Management, King Mongkut's University of Technology North Bangkok, Thailand

Elsevier use only: Received 30 September 2011; Revised 10 November 2011; Accepted 25 November 2011.

Abstract

A knowledge map is a technique that increasing efficiency of an explaining associations of knowledge bodies with the purpose of managing knowledge in an organization by creating similar content associations. The most commonly-used algorithm is the prediction using association rules. However, this technique sometimes produced unsatisfactory results because the retrieved information hardly met the demand and could not explain associations of information contents to users deriving in less accessibility to knowledge. This research presents methods of explaining knowledge associations using a knowledge map, and estimating an appropriate association value for each piece of information. The value was then computed to find an assembly point of multiple-relation knowledge. A comparison, in which 1,000 academic articles represented knowledge stored in an organization, revealed that the method of explaining associations hidden in the contents using the knowledge map yielded similar and more detailed prediction values compared to the method of defining categories from knowledge topics. This enabled users to see associations of knowledge in an organization and gave them access to knowledge stored within a knowledge base, consequently stimulating a knowledge spiral and integrated knowledge in a simple and effective manner.

Keywords: Multiple Relation; Knowledge Map; Knowledge Direction; Integrated Knowledge

1. Introduction

In recent time, many organizations focus on knowledge management to provide users with access and retrieval of information. It is widely recognized that good knowledge management (KM) leads to higher efficiency and enterprise performance. NONAKA [1] classified knowledge into two types:

* Corresponding author. Tel.: +66 81 531 1236.
E-mail address: watthananon@hotmail.com, anirach@ieee.org.

1877-7058 © 2012 Published by Elsevier Ltd. Open access under CC BY-NC-ND license.
doi:10.1016/j.proeng.2012.02.073

tacit knowledge and explicit knowledge, and through what he called “a knowledge spiral” [2], knowledge is adjusted and developed. This concept is accepted by leading public and private organizations around the world. As a result, information technology has become more important and been applied to several procedures to improve management and store knowledge and data in a more systematic way. According to some researches [3] [4], state-of-art methods of data management and storage leads to rapidly increasing amounts of messages and documents. Due to improper storage of data and limited access to knowledge base, relations and connections of the stored data are hardly recognized. Data is gathered in an aggregate manner for separate searches for knowledge, keywords and query words. Relations of data contents are not analyzed. There are limitations to data searching and storage techniques and explanation of relations between different datasets. One way to manage knowledge within an organization is to create a knowledge map that gives access to desired knowledge and demonstrates directions of contents in a timely manner [1].

This research aims to present ways of increasing efficiency of knowledge management with the use of knowledge maps. An in-depth analysis of hierarchical structure of contents was conducted to identify relations between similar data and help organizations see knowledge directions in order to improve existing sciences and manage vanished sciences.

2. Related Works

Many researches try to present a guideline of creating knowledge maps to develop and increase competitiveness of organizations and institutions. For example, Vega-Riveros and colleagues (1998) [5] employed concept maps to demonstrate learning maps of living things to undergraduate students. Concept maps can be used to create pictures explaining concepts that students want to present. The research by Richardson (2001) [6] applied a knowledge map called Skill Matrix to improve personnel’s knowledge and capacity. There are two development phases in this research. The first phase determines competency in terms of knowledge, capacity and skills. The second phase examines and tests personnel in organizations for their skills and capacity to find out their fields of knowledge and capacity. The research by Eppler (2001) [7] introduced a knowledge map as a knowledge management instrument. The research classified knowledge maps into five groups: 1) a knowledge resource map identifying which organizations, sub-units or persons keep the knowledge; 2) a knowledge asset map of organizations; 3) a knowledge structure map identifying knowledge structure; 4) a knowledge application map identifying which fields of knowledge are applied by organizations; and 5) a knowledge development map. It can be seen that this research relied on relations of knowledge to clarify efficiency of the maps. The research by Lin and Hsueh (2002) [8] presented a method of knowledge map creation and data maintenance applying data retrieval and data mining techniques to explain data synthesis. This method was tested with dissertations from the National Library of Taiwan with cooperation from 16 experts. It was found that the knowledge map including maintenance functions helped users increase knowledge and improve storage structure. The research by Kim and colleagues (2003) [9] presented a method of creating an industrial knowledge map by capturing and representing organizational knowledge. It was found that a knowledge map was a tool representing knowledge and suggested effective concrete procedures. The research by Judith and colleagues (2004) [10] applied a network plan to create a graphical knowledge map identical to a concept map, in which nodes were nouns representing knowledge and lines between nodes were verbs representing relations of several kinds. It was found from the research that the knowledge map showed relations among procedures, concepts and competency giving easy and effective access to knowledge sources. In 2006, the research by Eppler [11] further developed a knowledge map using principles of a thesaurus for the purpose of data cleaning by screening words with similar meanings in order to classify knowledge maps into different types. The results of the research clearly indicated differences between words and types of diagrams enhancing understanding and memories.

The majority of the above researches introduced creation of knowledge maps to classify increasing knowledge contents into different categories, but did not identify relations between fields of knowledge and in-depth connections of contents. This research has initiatives to manage relevant knowledge contents, explain implicit relations of the contents with knowledge maps by appropriately weighting relations of knowledge contents, and, finally, make a conclusion by determining similarity values.

3. Research methods

Explaining relations of knowledge contents by appropriately weighting relations of knowledge contents can better explain similar relations of each article. With this method, multiple relations of contents were analyzed in order to explain in-depth relations of articles using the following related categories and contents. Therefore, the research methodological follows five steps, which are: (1) Data selection; (2) Data cleaning; (3) Appropriate association value estimating; (4) Document classification and (5) Similarity computation.

3.1. Data selection

Data used in the experiment were derived from 1,000 academic articles achieved in two electronic databases: the ACM Digital Library [12], a computer and information technology database storing journals, magazines, academic conference documents, newsletters and news articles published by the Association for Computing Machinery (ACM); and EBSCO Academic Search Premier [13], a website collecting academic articles covering interdisciplinary studies such as education science, history, social science, humanities, liberal arts, jurisprudence, general sciences, health science, business administration, etc. Over 8,500 indexes and abstracts and 4,640 full articles were given [14].

3.2. Data cleaning

Data cleaning assures reliability and integrity of articles employed in the research. Incomplete articles with symbols and special features were deleted because they had no significance. This procedure reduced the number of articles to 983.

3.3. Selection of representative data

Selection of representative data drew out only related and required data. Most academic articles comprise of ten elements: title, abstract, keywords, date, introduction, related works, method, experimental result, conclusion, and references. Only four elements were selected in this study, namely title, abstract, keywords and date, to represent academic articles employed in the research. The research by Fuller and colleagues (2008) [15] found that some, not all; part of data could be selected as representative data and referred to required documents. In the research by Peery and colleagues (2008) [16], data were retrieved using only three elements of representative data: content, metadata and structure. It was found that precision of data retrieval was increased by 50%. Hence, in this study, only four elements of representative data were selected and, then, turned into calculable Cartesian products as seen in Equation (1) below.

$$Doc = D_1 \times D_2 \times D_3 \times ... \times D_d \qquad (1)$$

where d is feature of articles and the feature of each article has properties are text only, for example.

$$Doc = Title \times Abstract \times Keywords \times Date \tag{2}$$

where Title Feature means Term used to describe the subject of the article.
Abstract Feature means Term used to describe the abstract of the article.
Keyword Feature means the keyword term used to describe in the article.
Date Feature means the number used to describe day/ month/ year of the article.

After that, association values of the first three elements were calculated. In the research by Watthananon (2010) [17], it was found that the best resulting proportion of representative articles was 4: 3: 3 for title: abstract: keywords. Results of a multiple relation content analysis were consistent with analysis results of librarians and analysis results of the OhioLink database. Dates of articles were stored for tracking and classifying purposes so they should not be used to calculate weight of representative data, represented as Cartesian products as follows.

$$Doc = Title_{w1} \times Abstract_{w2} \times Keywords_{w3} \tag{3}$$

where $w1$ means appropriate proportions to calculate the weight of the title.
$w2$ means appropriate proportions to calculate the weight of the abstract.
$w3$ means appropriate proportions to calculate the weight of the keyword.

3.4. Appropriate Association Value Estimation

After data were transformed into calculable figures, appropriate association values of each data were estimated. This process is divided into two steps. The first step is extracting words from contents. Then, insignificant words that did not affect the contents were left out such as a, an, the, and, of, etc. The system will not consider these words when classifying the contents. Remaining words were, then, processed into two-word phrases and three-word phrases, respectively. The N-grams method separated contents into terms using N values to cut words, and terms were replaced by N letters lining up one after the other. N-grams is a technique of determining probability of terms combined into sentences: $w_1, w_2, w_3, \ldots, w_n$. Association values can be estimated as seen in Equation (4) below.

$$P(w_1 w_2 K w_r = \prod_{i=1}^{T} P(w_i | w_i K w_{i-1}) \tag{4}$$

where w is Term, n is the next numeral, P is Probability from database, T is number of term, i is sequence of term by beginning from level 1 and ($w_1, w_2, w_3, \ldots, w_n$) is a set of words contain more than 3 word.

From the equation, as you can see $P(w_1|w_1,\ldots w_{i-1})$ is probability of word w_i after the occurrence of the word $w_1, w_2, \ldots, w_{i-1}$ precedent. Therefore, the probability of sentence by 2-grams method is $P(w_1, w_2, w_3,\ldots, w_n) = P(w_1), P(w_2|w_1), P(w_3|w_2),\ldots,(P(w_n|w_{n-1})$ and the probability of sentence by 3-grams method is $P(w_1, w_2, w_3,\ldots, w_n) = P(w_1), P(w_2|w_1), P(w_3|w_1w_2), \ldots, (P(w_n|w_{n-2}w_{n-1})$

The second step is combining relation proportions of every frequency to find relations of contents in articles. Frequencies of extracted words were counted and compared with words stored in the database to calculate DDC categories in the order of 4, 3, 2 and 1, respectively. When all figures are received, relations of all categories are gathered to identify the number of each category. If numbers of categories are redundant, their association values will be combined because they are in the same category. After that, all frequencies of categories are adjusted into percentage as seen in Equation (5) below.

$$Percentage = \frac{f_i \times 100}{\sum_{i=1}^{n} f_i} \tag{5}$$

where f is value of terms in each class and i is level of class

3.5. Similarity Computation

Data input into this system collect association values of n articles comprising of m categories. The data were stored in the matrix form. There are two methods of calculating similarity of two articles [18], [19]: correlation-based calculation and cosine-based calculation. However, the researches by Herlocker and colleagues (1999) [20] and by Lertmahakiat and colleagues (2009) [21] showed that Pearson Correlation Coefficient (PCC) was the most effective method of similarity computation. Therefore, this research chose PCC to calculate similarity between two articles. The first step of calculation includes set of articles that had association values derived from a content analysis of each article. The PCC calculation between document a (the article requiring comparison) and document b (the article stored in the system) is shown in Equation (6) below.

$$C_{a,b} = \frac{\operatorname{cov}(r_a, r_b)}{\sigma_{r_a} \sigma_{r_b}} \tag{6}$$

where r_a and r_b are relation proportions from *document a* and *document b*

Covariance:

$$\operatorname{cov}(r_a, r_b) = \frac{\sum_{i=1}^{m} (r_{a,i} - \overline{r_a})(r_{b,i} - \overline{r_b})}{m} \tag{7}$$

$$\overline{r_b} = \frac{\sum_{i=1}^{m} r_{b,i}}{m} \tag{8}$$

where $r_{a,i}$ and $r_{b,i}$ are relation proportions of class i related with *document a* and *document b*
$\overline{r_b}$ is the average relation proportions of *document b*
m is number of *co-rated items*

Standard Deviation:

$$\sigma_{r_b} = \sqrt{\frac{\sum_{i=1}^{m} (r_{u,i} - \overline{r_u})^2}{m}} \tag{9}$$

The research by Herlocker and colleagues (1999) [20] stated that using only PCC was not enough to calculate similarity because it is not known whether the resulting score represented articles with contents similar to articles stored in the system. Thus, Herlocker and colleagues (1999) employed the $S_{a,u}$

(Significance Weight) equation to calculate weight values of articles with contents similar to target articles as seen in Equation (10) below.

$$S_{a,b} = \begin{Bmatrix} 1 & if & m > 50 \\ \dfrac{m}{50} & if & m \le 50 \end{Bmatrix} \tag{10}$$

where m is number of *co-rated items*

In this way, we can calculate the similarity between the articles by Pearson Correlation Coefficient (PCC) and Significance weight is shown in Equation (11) below.

$$sim(a,b) = S_{a,b} C_{a,b} \tag{11}$$

where $S_{a,b}$ is Significance weight
$C_{a,b}$ is Pearson's correlation between *document a* และ *document b*

4. Experimental results

4.1. Analysis content

After appropriate association values of each data were estimated and knowledge contents were classified using the Dewey Decimal Classification (DDC) to store data into the system as the database of knowledge maps. Table 1 shows relations of articles after contents were analyzed by the above methods.

Table 1. Shows sample relations of articles after contents were analyzed

Class	Proportion (%)
000 Computer science, information & general works	2.03
000 Computer science, knowledge & systems	0.92
000 Computer science & general works	0.46
001 Knowledge	0.03
002 The book	-
003 Systems	0.09
004 Data processing	0.36
005 Computer programming	1.75
006 Special computer methods	0.09
010 Bibliographies	-
010 Bibliography	-
011 Bibliographies	-
…	…
…	…
…	…
990 General history of other areas	0.18
993 General history of other areas; New Zealand	-

Table 1. Shows sample relations of articles after contents were analyzed (Cont.)

Class	Proportion (%)
994 General history of other areas; Australia	-
995 General history of other areas; Melanesia	0.55
996 General history of other areas; Other parts of Pacific	0.18
997 General history of other areas; Atlantic Ocean Islands	-
998 General history of other areas; Arctic islands & Antarctica	-
999 Extraterrestrial worlds	-

Table 1 shows sample articles, of which association relations were classified using the weight of 1,000 categories. It can be explained that articles used in the test, when decoded, can clearly classify in-depth relations of knowledge contents as calculated by association values. Classification by category weight is more detailed than classification by title of knowledge. When relations of contents of 983 articles were analyzed and calculated using the DDC, it was found that the analyzed contents had relations divided into categories of weight of the real contents.

Contents from the knowledge are focused on three main groups, i.e. class social science, class pure science and mathematics, and class technology, with association values of 19.11%, 15.63%, and 18.48%, respectively. It can be explained that the analyzed contents of articles of social sciences have association values similar to those of technology. This means that current social science articles have improved in many ways and in the direction more related to everyday life, for example, management, education, commerce, communications, transportation, etc. These disciplines require knowledge of technology for continuous development while technology requires scientific knowledge to assist with cognition, analysis, calculation, proving with mathematic equations.

4.2. Similarity Analysis

This research selected the PCC to analyze similarity of articles. Criteria of content comparison were set at 0.50. It was found that 634 articles or 64.57% of all articles had correlation coefficients of over 0.50%, meaning that more than 50% of the articles in the knowledge base have similar contents. Similarity levels of articles are shown in Table 2.

Table 2. Shows sample similarity levels of articles

Rank	Doc No.	Pearson's
1	273	0.8637057838**
2	584	0.8560661351**
3	92	0.7676458653**
4	701	0.7646068655**
…	..	…
983	412	0.1208162422**

**Correlation is significant at the 0.01 level (2-tailed).

5. Discussion

5.1. Relation Structure

Analysis of appropriate weight for each data and usage of coordinating points of organizational graphs, or knowledge maps, clearly demonstrate multi-relations connecting to concrete knowledge. This helps humans see overall structure of knowledge assets of complex organizations. These connections are relation structure of nodes or relation categories. Relational nodes around central nodes are sub classes and super classes. These nodes are related to development of linked lines and knowledge than can be classified. For example, the word 'network' appears in an abstract of an article. Then, that content is analyzed and relations are identified using the DDC. Many words are then combined into knowledge groups. When graphs are created, DATA are encoded, analyzed and classified. A multi-relation network was, then, created. That data may be a sub class of a super class or a sup class or that data may be linked to other related knowledge nodes. Therefore, when lines between central nodes and relational nodes were drawn, directions and patterns of knowledge maps are clearly seen. In addition, organizations can use basic structure of knowledge as a foundation to systematically increase knowledge.

5.2. Knowledge Direction

Creating knowledge maps by estimating weight of this research contributes to added value of knowledge in organizations because it gives access to knowledge or unseen information. Knowledge maps become contents presented to users in the form of relations increasing knowledge spirals in organizations and ability to create progression towards mutual targets of organizations. Moreover, knowledge map application is an instrument similar to the art of managing organizations of executives because organizations have clear directions of knowledge management. Personnel's clear understanding of directions of organizations and awareness of managed knowledge lead to high performance organizations (HPO) with back-up plans resulted from knowledge analysis. It is important to select appropriate knowledge maps and factors driving towards HPO's.

6. Conclusion

An increasing amount of knowledge limits access to knowledge of users who may be lost in space due to their lack of understanding of relations and connections of stored knowledge. One way to manage the ever-increasing knowledge is to create relations of knowledge by connecting and explaining related materials using association values to consider knowledge contents. After that, graphs of knowledge relations called knowledge maps were created. Knowledge maps are a method of knowledge management that every organization can create because they are not complicated. There are many benefits of knowledge maps. The bigger the organizations or institutions, the more benefits they will get from knowledge maps. On knowledge maps, there are papers, academic documents or opinions useful for work and solving problems in different situations. Knowledge maps can help solve problems when personnel have questions and executives can see big pictures, directions of organizations, and what types of knowledge organizations need that are in line with strategies and visions of organizations at present. Finally, executives can utilize knowledge maps to analyze directions of organizations, create knowledge relations and determine strategies of successful management of organizations.

References

[1] Nonaka, I. *The knowledge creating company*. Harvard Business Review. 1991; 96-104.

[2] Nonaka, I. and Takeuchi, H. *The knowledge creating company : How Japaness companies create the dynamics of innovation.* [n.p.];Oxford University Press; 1995.

[3] Oded, N. *Creativity, Knowledge and IS: A Critical View*. Proceedings of the 38th Hawaii International Conference on System Sciences. [n.p. : n.p.]; 2005.

[4] Sajjad, M. *Storage of Transferred Knowledge or Transfer of Stored Knowledge: Which Direction? If both, then how?* Proceedings of the 38th Hawaii International Conference on System Sciences. [n.p. : n.p.]; 2005.

[5] Vega-Riveros, J. F., Marciales-Vivas, G. P., and Martinez-Melo, M. *Concept Maps in Engineering Education: A Case Study.* Journal of Engineering Education, Vol. 2, No. 1, 1998.

[6] Richardson, Don. *The Practical Reality of Knowledge Management within Development Initiatives*. Materials for the 2nd Comprehensive Workshop, ENRAP, Singapore, February 6-9; 2001.

[7] Eppler, M.J. *Making knowledge visible through intranet knowledge maps: concepts, elements, cases.* Paper presented at the International Conference on System Science, Hawaii; 2001.

[8] Lin, F. and Hsueh, C. *Knowledge Map Creation and Maintenance for Virtual Communities of Practice*. Proceedings of the 36th Hawaii International Conference on System Sciences; 2002.

[9] Kim, S et al. *Building the knowledge map: an industrial case study*. Journal of knowledge management, vol. 7, no. 2, 2003, pp. 34-45.

[10] Judith, H. H.,Kampschroer, K., Powell, M. K., and Loftness, V. *Collaborative knowledge work environments, Building Research & Information*, Publisher: Routledge, part of the Taylor & Francis Group Issue: Volume 32, Number 6 / November-December 2004, p.510 – 528

[11] Eppler, M. J. *A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing*. Information Visualization, 5, 2006, p. 202–210.

[12] http://portal.acm.org

[13] http://search.ebscohost.com

[14] http://www.arit.dusit.ac.th

[15] Fuller, M. C., Biros, P. D. and Delen, D. *Exploration of Feature Selection and Advanced Classification Models for High-Stakes Deception Detection.* Proceedings of the 41st Hawaii International Conference on System Sciences, IEEE, 2008.

[16] Peery, C., Wei Wang, A. M. and Nguyen, Thu D. *Multi-Dimensional Search for Personal Information Management Systems.* EDBT'08. Nantes, France, 2008.

[17] Watthananon, J. *The Multiple Relation of Knowledge Using Dewey Decimal Classification*. Ph.D. Thesis, Department of Information Technology, King Mongkut's Universit of Technology North Bangkok, Thailand. 2010.

[18] Breese, J. S., Heckerman, D., and Kadie, C. *Empirical Analysis of Predictive Algorithms for Collaborative Filtering.* Technical Report MSR-TR-98-12. [n.p. : n.p.], 1998.

[19] Sarwar, B., et al. *Item-based Collaborative Filtering Recommendation Algorithms*. Proceedings of the WWW10 Conference. [n.p. : n.p.], 2001.

[20] Herlocker, J. L., et al. *An Algorithmic Framework for Performing Collaborative Filtering*. Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. [n.p. : n.p.], 1999 : 230-237.

[21] Lertmahakiat, W. and Mingkwan, A. *A Study of Retrieval Algorithm for DDC-Multiple Relation Classification in Content.* Proceedings PG Net 2009, the 10th Annual Postgraduate Symposium on the Convergence of Telecommunications, Networking and Broadcasting, Liverpool John Moores University. [n.p. : n.p.], 2009.