Prototype of Knowledge Management System in Chinese Offshore Software Development Company

LI CAI*, ZUOQI WANG*, YUFENG JIAO*,
MASANORI AKIYOSHI**, NORIHISA KOMODA**
* Jinan Ryouka Science & Technology Co., Ltd.
Qilu Soft Park, No.1, Shunhua Road, Jinan Hi-tech Zone, Sandong, 250101
CHINA
** Graduate School of Information Science and Technology
Osaka University
2-1, Yamadaoka, Suita, Osaka, 565-0871
JAPAN
{cai, wangzq, jiaoyf}@ryouka.com.cn, {akiyoshi, komoda}@ist.osaka-u.ac.jp
http://www.ryouka.com.cn/, http://www-komo.ist.osaka-u.ac.jp/

Abstract: - This paper illustrates preliminary design and evaluation of a knowledge managing system in Chinese offshore software development company. The major features of the knowledge management system are a BBS style Q&A function, a management of Q&A results based on domain knowledge, an association retrieval function for coping with grammatical ambiguity including in the queries, and a display of Q&A texts on the plain for grasping the general distribution of knowledge and problems. The prototype system has developed using JAVA and .Net Frameworks, and has used in-house. Through the trial use in-house, not only the advantage but also several improvement problems are clarified.

Key-Words: - knowledge management, question and answer system, association retrieval

1 Introduction

Although there is a negative effect of soaring Chinese personnel-expenses and the exchange appreciation of the Chinese Yuan, offshore software development from Japan to China is increasing rapidly. Software export of China in 2001 is only 0.75 Billion dollars. It increased to 4.5 billion dollars in 2006. Namely, it rises at average annual rate of more than 30%. More than half is said for Japan. Major reasons are;

(1) China and Japan use almost same Chinese characters. Both people can roughly understand the meaning by reading characters.

(2) The distance is near. It takes only 3 hours flight from Tokyo or Osaka, Japan to Beijing or Shanghai, China.

(3) The reservation of inexpensive SE is relatively easy.

Offshore software development has not been well discussed as research target so far, because it includes know-how such as "best practice" which seems to belong to corporate secret. However an academic report on it[1] pointed that offshore software development would make great impact on changes in IT industry and is inevitable to promote under globalization, then management aspect comes to be under argument from both software engineering and knowledge management points of views. From software engineering, it is closely related to CMM (Capability Maturity Model)[2]. From knowledge management, it needs more intensive discussion on how to make a bridge of educational and cultural gaps among foreign companies. Especially, in most offshore software development companies in China, majority of employees have little experience in software development. In addition, employees have to understand the documents written in foreign language such as Japanese, and also have to write the documents in foreign language. Nevertheless, the foreign language level of the most employees is relatively low. Furthermore, because of the rapid improvement of the information technologies, employees have to continue learning.

To smoothly and effectively carry out their tasks in the above-mentioned situation, a support system for the novice software engineers is desired, which is different from conventional way of knowledge management systems[3].

A Chinese software company, "Jinan Ryouka Science & Technology Co., Ltd.", which the authors manage, planed the development a knowledge management system named "Ryouka KM System (RKMS)". The major features are a Bulletin Board System (BBS) style Q&A function, an association retrieval function for coping with grammatical ambiguity including in the queries, and a knowledge management function such as a display of Q&A texts on the plain for grasping the general distribution of knowledge and problems. A prototype system of the part of RKMS has been developed and used in-house on trial.

In this paper, firstly, the brief profile of Jinan Ryouka Science & Technology Co., Ltd. is introduced. Then, the outline of RKMS is explained. After that, two core subsystems are explained in detail. Finally, the prototype system and the result of trial use are shown.

2 Jinan Ryouka Science & Technology

Jinan Ryouka Science & Technology Co., Ltd. (RYOKA) was established for offshore software development to Japan in July 2002 in Jinan, the capital of Shandong Province, China. Currently, RYOKA has 3 subsidiaries in Beijing, Dailan, and Tokyo. As of the end of 2007, the company has a total 124 software engineers. RYOKA is planning to employ 120 novice engineers in 2008 fiscal year and 180 novice engineers in 2009 fiscal year. Fig. 1 shows the distribution of the employee's work experience in our company. From Figure 1, it is found that half of the employees are less than 2 years work experience.

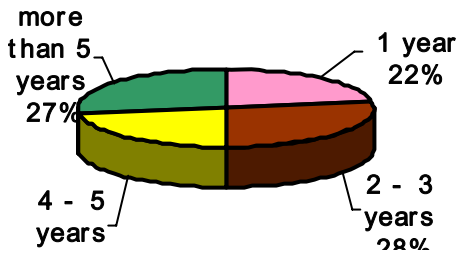

Fig. 1 Distribution of employees

Moreover, two thirds of the employee's Japanese-language level is JLPT (Japanese-Language Proficiency Test) level 2 or less. For effectively involving such novice engineers, a knowledge management system is necessary.

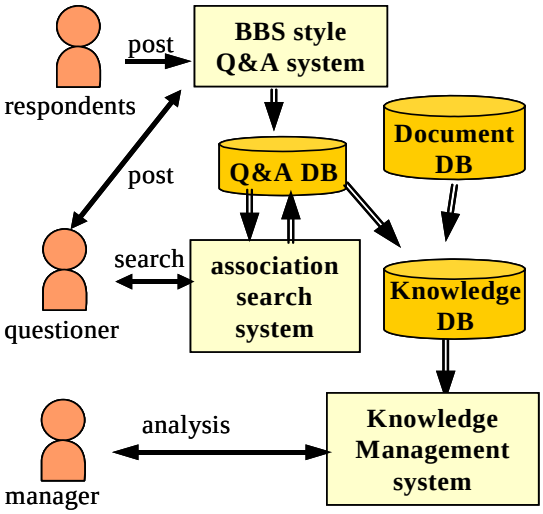

Fig. 2 Outline of RKMS

3 Outline of RKMS

An outline of the planning knowledge management system is shown in Fig.2. This system is consisted in three sub-systems, a BBS style Q&A system, an association retrieval system, and a knowledge management system.

A BBS style Q&A system is a typical BBS system. Though writing Q&A sentences in Japanese is encouraged to employees, Chinese sentences are also allowable. To preserve the leak of customer confidential information, the access control is carefully designed. For example, the information can be limited the range of disclosure.

An association retrieval system aims to reuse the existing Q&A database for self-education. The almost of all sentences in Q&A data stored in the offshore software development company are ungrammatical and are used wrong words because of the low ability of foreign language. Therefore, typical text search methods[4][5] are useless. To solve this problem, the association retrieval method of similar Q&A articles developed in Osaka University[6] is introduced.

A knowledge management system is developed for the managers to grasp the hot topics and weak points in the company. The threads of Q&A and the documents in the company such as daily reports are inputted. This subsystem has two types of user interfaces. The first one is a graphical trend display to understand the change of employees' typical interest. Another display is the user interface for analyzing atypical Q&A threads. Q&A threads are displayed on

the 2-demensional plain in accordance with their attributes.

Fig. 3 Distribution of employees

4 BBS style Q&A system

The prototype Q&A system is now tested in-house since the end of July. Since all employee login the management system everyday, this installation is convenient for the use of the Q&A system without additional operation. The stored Q&A contents can be searched by keyword. Fig.3 shows a retrieval screen. It is written in Japanese. The users can not only input plural key words in the upper input area but also select target fields. The lower part of the screen shows the retrieval result. About 80 employees participated the experience, and more than 500 Q&A sets are exchanged. Through the trial test, we get the good evaluation on the basic functions as a Q&A system. However, it is necessary to improve the administrative functions.

5 Association retrieval method

Web bulletin board in the BBS style Q&A system includes useful information for novice engineers. Users who want to obtain information generally retrieve by keyword, judging whether an article includes necessary information by reading its title or first sentence. However, only by the keyword matching, users cannot find out the suitable articles effectively, because the sentences in the articles are often written in erroneous usage of characters and expressions. Furthermore, in some cases, Chinese sentences are mixed. Therefore, a method for retrieving similar question articles to a query by natural sentence input to improve retrieval accuracy in an environment including the erroneous in characters and grammar is necessary.

To solve this problem, an association retrieval method[6] developed in Osaka University is introduced. Fig. 4 shows an outline of the association retrieval by natural sentence input. The flow of the retrieval procedure is shown below.

Step 1: A user inputs the retrieval query sentence.

Step 2: The input question sentence is analyzed in the morpheme, and the nouns are extracted.

Step 3: The articles, i.e. candidate articles, are retrieved from the question article database by using a set of extracted nouns.

Step 4: It is judged by using the modified cosine similarity index whether the question sentence in the candidate article is similar to the retrieval query sentence.

Step 5: As a result of the similarity judgment, users receive the article question sentence and question articles in order of their similarity to the question, with the most similar at the top. The order of articles is

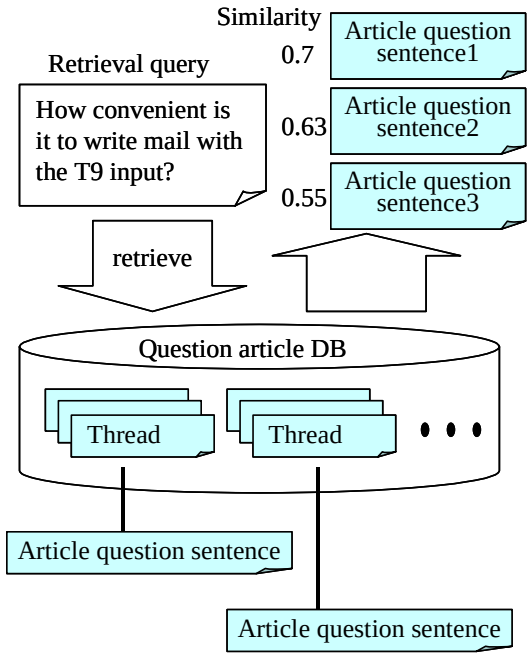

Fig. 4 Retrieval from Q&A DB by natural sentence

adjusted using the past users' evaluation.

The cosine similarity index is the one of the typical methods for judging the similarity between two sentences[7]. The modification of the cosine similarity index is done for coping with the erroneous usage of characters and expressions. The elements of the word vectors used in the cosine similarity index are derived from a sentence structure from the viewpoints of common words and non-common words between a user's query sentence and article question sentences.

6 Knowledge management system

6.1 Outline of the KM system

The Q&A thread of the Q&A system or the information on other documents in the company is stored as "card" in the knowledge DB. A card is a set of sentences such as a pair of question and answers, the description of problem, and so on. The outline of the support system is shown in Fig. 5.

The main purpose of this subsystem is that the manager easily grasps the trends of interest areas and finds new problems or topics. The cards containing typical topics (called typical cards) need to be grouped to count the number of the cards. Therefore, the key function of this subsystem is to extract atypical cards, which contain atypical topics (called atypical cards), from all cards and to group typical topics by contents. For categorizing typical cards and extracting the atypical cards, a typical word database reflecting analyst's background knowledge is used. As the result of the extraction, cards are divided into categorized typical cards and atypical cards. Different user interfaces are necessary, because the direction for use differs in atypical cards from typical cards. Typical cards are graphed to show time series behavior of the numbers of cards that are grouped by the content categories. On the other hand, to grasp the trend of atypical cards, another type of display is needed. Thereupon, the result of the atypical card classified by the contents is placed on the screen by the attributes of cards. Based on selected coordinate, the cards are displayed on the screen.

The prototype system has been developed in JAVA and .Net frameworks 2.0. The management functions such as maintenance of typical word DB, setting the attribute names of the card and so on are equipped. Fig.6 shows the screen for the typical cards display. In the center area, the time-series graph of the number of cards included in selected categories. The categories

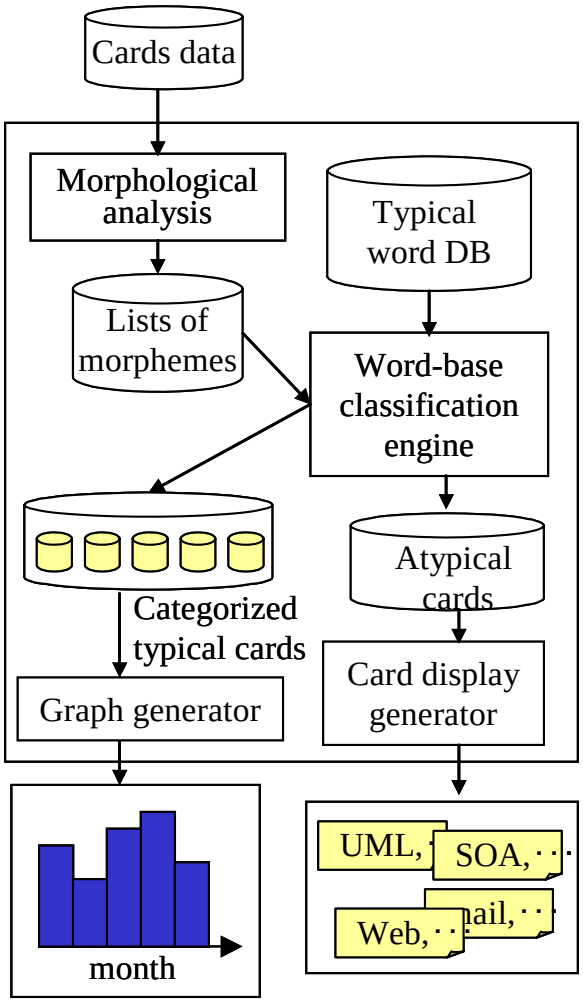

Fig. 5 Outline of Knowledge management system

to be displayed can be selected by using right side area of display. The graph can be rotated three dimensionally by mouse operation. By cricking a bar of graph, the card contents included in the designated

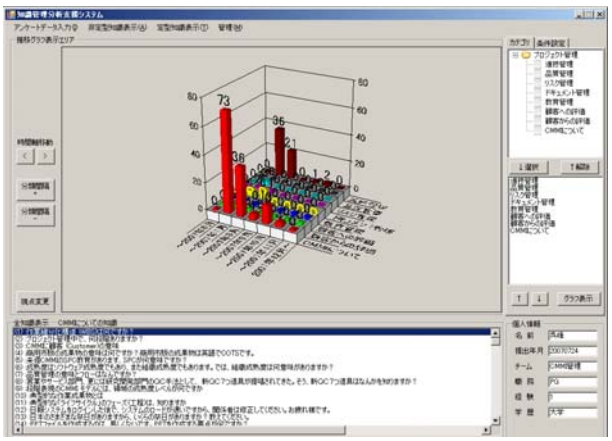

Fig. 6 Typical cards display

category and month are listed in the bottom left area. Also, by cricking the card content displayed in the bottom left area, the submitting employee's attributes, such as name, involved project team, years with ROUKA, and academic history are displayed in the bottom right area.

6.2 Atypical card Interface

Fig.7 shows the user interface for analyzing atypical cards. There are three window areas: Classified result area, Card area, and Text information area. In the classified result area, categories of classified atypical card are listed. In the end of the category name, the number of cards in the category is shown. When users select certain category by mouse operation, cards are displayed by contents similarity in the card area. In this area, the card is displayed largely as it is recent and only characteristic keywords are shown in the cards. Viewpoint is changed by mouse operation. When one card is clicked, raw text card and personal information of the card writer are shown in the text information area. To cope with vague changeable borderline between the typical and the atypical, users can add definition of typical cards by dragging and dropping the card that users judge as the typical.

6.3 Outline of atypical card extraction

The cards have to be classified into several typical categories and an atypical card category[8][9]. Fig. 8 shows the outline of the classification function. The meaning of cards with word units is judged by word based, because there are so many grammatical errors and incorrect letters in the cards.

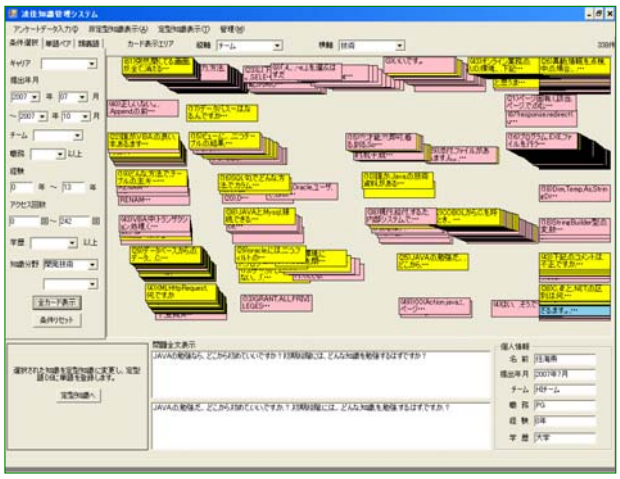

Fig. 7 Atypical cards display

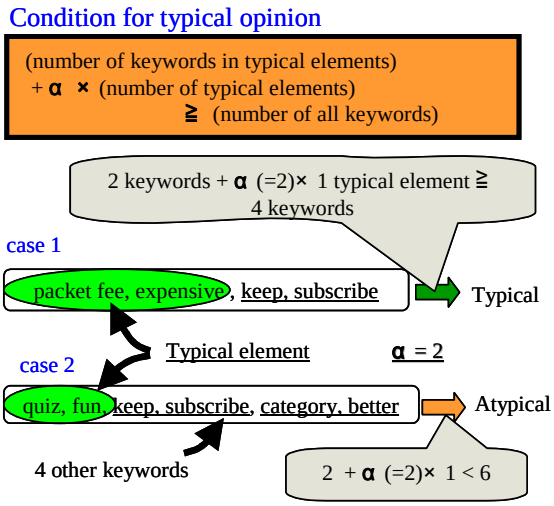

Fig. 8 atypical card extraction method

Firstly, the card data is decomposed into word lists by morphological analysis with "ChaSen" [10], Japanese software. Then, from the word lists, nouns, independent adjectives, and independent verbs as the minimum words for understanding a sentence are extracted. In Japanese, nouns are roughly divided into 14 kinds. There are also words that join keywords in context. Based on morpheme connection and extraction rules, morphemes are transformed to keywords. For example, "Web" and "service" are transformed to keyword "Web service" by the rules.

In the typical word database, the combinations of words are set up beforehand. System managers can maintains the typical word database. When they recognized a new category of the cards, they can add them to the typical word database. A matching engine to extract atypical cards compares keywords of cards with the typical word database.

The prototype knowledge management system was evaluated using about 200 man-made problem cards. Through the trail test of the basic method used in the system, it is cleared that a coordinate method of typical erroneous usage of characters is needed. Also, the improvement of the classification method is requested, because the length of sentences are long compared with the original application.

7 Conclusion

This paper described a knowledge managing system in the offshore software development company from organizational usage style including managers and

employees, which is expected to circulate various know-how in them. The development project on the proposed system has started. A prototype system of the BBS style Q&A system and the Knowledge management system has been developed and has been used in-house. These two are seamlessly integrated with transferring data between databases. Based on the evaluation results and user opinions, they will be improved for further effective use in the software development projects. Also, for improving the classification accuracy, and for easily generation of the typical word DB, a new classification engine is tested.

References:

[1] W. Aspray, F. Mayadas and M. Y. Vardi(eds.), *Globalization and Offshoring of Software*, Report of the ACM Job Migration Task Force, Association for Computing Machinery, 2006

[2] Carnegie Mellon University, *The Capability Maturity Model: Guidelines for Improving the Software Process*, Addison Wesley Longman, Inc., 1994

[3] R. Maier(eds.), *Knowledge Management Systems: Information And Communication Technologies for Knowledge*, Springer, 2007

[4] K. Sparck-Jones and P. Willett(eds.), *Readings in Information Retrieval*, Morgan Kaufmann Publishers, 1997

[5] A. Hickl, et. Al., Question Answering with LCC's CHAUCER at TREC 2006, in *Proc. of the fifteenth Text REtrieval Conference(TREC 2006)* , 2006, in http://trec.nist.gov/pubs/trec15/t15_proceedings.html

[6] Y. Sakurai, et. al., A retrieval method of similar question articles from Web Bulletin Board, in *Proc. of the 1st Int'l Conf. on Software and Data Technologie (ICSOFT-2006)*, 2006, pp.238-243

[7] G. Salton, *Automatic text processing: The transformation analysis and retrieval of information by computer*, Addison-Wesley, 1988

[8] A. Hiramatsu, et. al., Man-machine interface of a support system for analyzing open-ended questionnaires, in *Proc. of 18th Int. Conf. on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems (IEA/AIE2005)*, 2005, pp. 318-320

[9] K. Kimura, et. al., Analysis support system of open-ended questionnaires based on atypical and typical threads classification, in *Proc. of IEEE Int'l Conf. on Computational Cybernetics (IEEE-ICCC2006)*, 2006 pp.171-176

[10] Y.Matsumoto, et. Al., Japanese Morphological Analysis System ChaSen version 2.2.1, 2001, http://chasen.aist-nara.ac.jp/chasen/doc/chasen-2.2.1.pdf.