BBS-based Information Management System in Chinese Offshore Software Development Company

LI CAI, ZUOQI WANG, MASANORI AKIYOSHI, and NORIHISA KOMODA

***Abstract*—** This paper discusses BBS-based information management system, which is used in an offshore software development company. Employees encounter some questions during software development and exchange their knowledge by using the Bulletin Board System(BBS). Besides such BBS database stores reusable knowledge which are represented in both Chinese and Japanese BBS article threads, so employees can retrieve such knowledge via Chinese or Japanese query input as they want. These two functionalities are considered to be complementary from practical usage points of views, so evaluation by using questionnaire to users are necessary to clarify good/bad points. We confirmed information retrieval functionality should be improved to enhance practical usability.

***Keywords*—**BBS, Information Management, Knowledge Management, Offshore Development

I. Introduction

RECENT expansion of offshore software development in China as shown in Fig. 1 indicates some crucial problems on its productivity and quality level. One big problem is that they have faced to handle language problems, for instance, some design specification documents are described in Japanese and the Chinese engineers have to understand them completely.

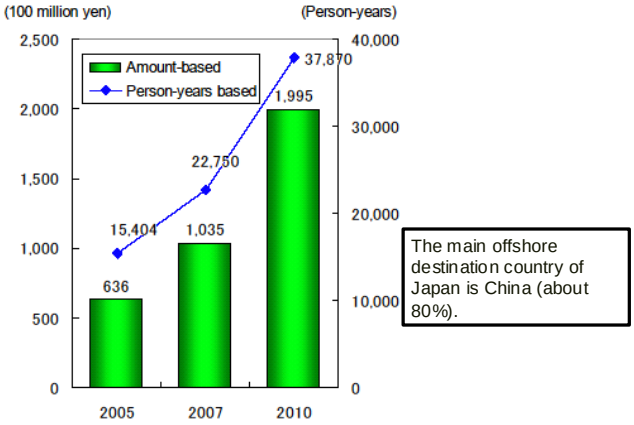

Fig. 1 Expansion of offshore software from Japan to China

Another big problem is that novice software engineers are increasing rapidly without well-organized training courses. Nevertheless offshore software development is expected to make great impact on changes in IT industry and is inevitable to promote under globalization. Therefore the 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)[1]. From knowledge management, it needs more intensive discussion on how to make a bridge of educational and cultural gaps among foreign companies. Of course CMM validation is necessary to accelerate expansion of their business, however, knowledge management is more crucial to execute daily development from practical points of views.

Based on these viewpoints, a Chinese software company, "Jinan Ryouka Science & Technology Co., Ltd.", which some of the authors manage, developed a knowledge management system named "Ryouka KM System (RKMS)"[6]. The major features of the system are a Bulletin Board System (BBS) style Question and Answer (Q&A) functionality, and information retrieval functionality for coping with Chinese and Japanese articles without regard to the language gaps. In this paper, firstly the outline and specific aspects of RKMS is explained. Then RKMS is discussed with users' questionnaire results.

II. Ryouka Knowledge Management System

A. Outline

An outline of the knowledge management system is shown in Fig.1. The system consists of BBS style Q&A functionality and information retrieval functionality. The BBS style Q&A functionality is a typical BBS, which involves bundles of a question article and answer articles as one thread. 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.

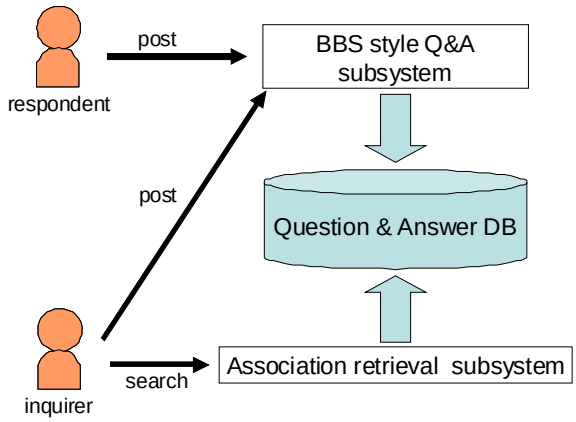

Fig. 1 Outline of RKMS

For example, the information can be limited the range of disclosure. Information retrieval functionality aims to reuse the existing Q&A database for self-education. The most Japanese sentences in Q&A data stored in the offshore software development company are ungrammatical and are used incorrect words because of the employees' low ability of foreign language. Therefore, typical text search methods[2][3] are useless. To solve this problem, the association retrieval method of similar Q&A articles [4] developed in Osaka University is introduced.

Users who want to obtain information generally retrieve by keywords, 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 Japanese articles are often written in erroneous usage of characters and expressions. Furthermore, in some cases, Chinese sentences are mixed. Therefore, a new method is necessary that can retrieve Japanese similar question articles to a query by natural sentence input under the situation including the erroneous in characters and grammar. The flow of the association retrieval procedure is shown below.

Step1: A user inputs a Japanese retrieval query sentence.
Step2: The input question sentence is analyzed in the morpheme, and the nouns are extracted.
Step3: The articles, i.e. candidate articles, are retrieved from the question article database by using a set of extracted nouns.
Step4: 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.
Step5: 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.

B. *Association Retrieval Mechanism*

The cosine similarity index is the one of the typical methods for judging the similarity between two sentences [5]. Fig. 2 shows the calculation definition of the conventional cosine similarity index. The cosine similarity index considers the retrieval query sentence and the article question sentence to be a set of words, and puts them into the word vector of n dimension. The cosine similarity index tends to be high when there are a lot of words common to both the retrieval query sentence and the article question sentence. However, as denoted before, common words are in fact few because the query sentence may have incorrect words and the article question sentence may have as well.

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 modified from the viewpoints of structural similarity related to common words and non-common words between a user's query sentence and article question sentences. Fig 3 shows an example of modifying the value of the vector in consideration of the sentence structure. Phrases with similar structure often have the same meaning, so in this study we use the partial structure of sentences composed of pairs comprising common words and non-common words of the retrieval query sentence and the article question sentence. The similarity index is improved by modifying the vector element of non-common words included in the structure of the retrieval query sentence to α.

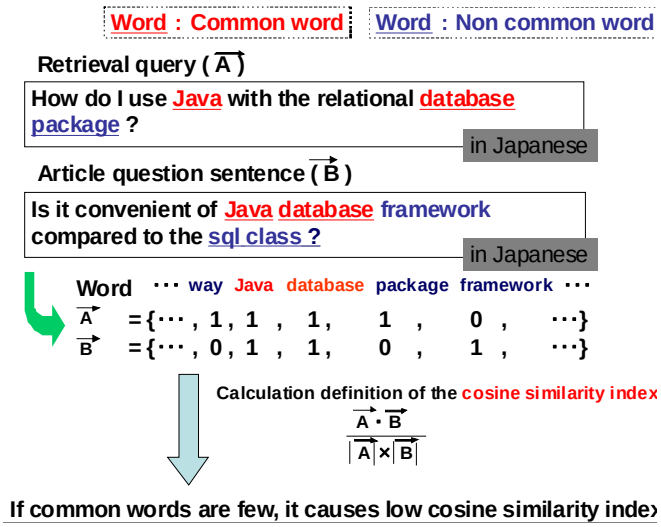

Fig. 2 Similarity judgment using cosine similarity index

Conversely, there are words that cannot apply to word matches or structural similarities of sentences. These words are considered to suppress the similarity of sentences. Non common words in the article question sentence are assumed not to be related to words in the retrieval query sentence. In this case, the similarity index is reduced by modifying the word vector element of the retrieval query sentence to β.

Thus even if a user input incorrect Japanese sentences with erroneous words, this association retrieval mechanism may achieve robust and adequate results.

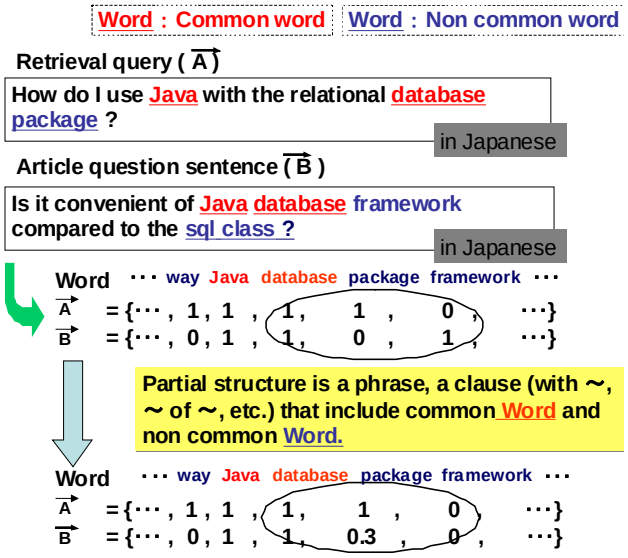

Fig. 3 Modified similarity judgment

The structure of a sentence is systematically decided by using the dependency analysis tool " Cabocha". The definition of the structure in the sentence, including common and non-common words is as follows when the dependency analysis is used.

- A clause including common words qualifies a clause with non-common words.
- A clause including non-common words qualifies a clause with common words.

If a word related to non-common words of the retrieval query sentence is included in the article question sentence, it is assumed that the structure of the retrieval query sentence is included in the article question sentence. If so, the structural similarity of the sentences is compared using the relation between words with the co-occurrences. The co-occurrences of a word is when the two words exist simultaneously in the same sentence. The co-occurrence dictionary for the word is automatically produced beforehand from all the article question sentences. On the contrary, if non-common words of the article question sentence did not co-occur with all non-common words of the retrieval query sentence in the comparison of the sentence structure, we consider the word that suppresses the similarity of the retrieval query sentence and the article question sentence.

Thus modified cosine similarity index is calculated and retrieval, so called "association retrieval", is executed by using this index.

C. *User Interface*

Fig. 4 shows the interface of BBS-style question input, which consists of necessary items and optional items. The most significant necessary item is inquiry theme, because BBS database has 7 categories of "development technology", "basic knowledge study", "daily business work", "in-house system usage", "business general knowledge", "project management" and the others.

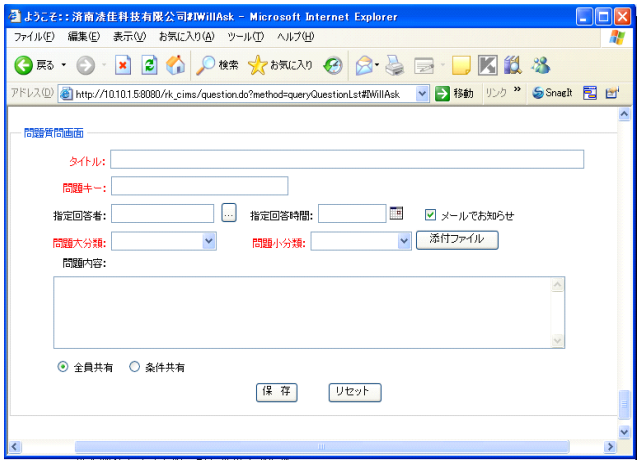

Fig. 4 BBS-style question input interface

In addition to such a query input, RKMS provides advanced support to get reply as soon as possible. Fig. 5 shows the setting menu of expected respondents related to the inquiry. Based on this setting data, RKMS sends the request message e-mail automatically, and notifies the inquirer if some of respondents or other members post the message to BBS.

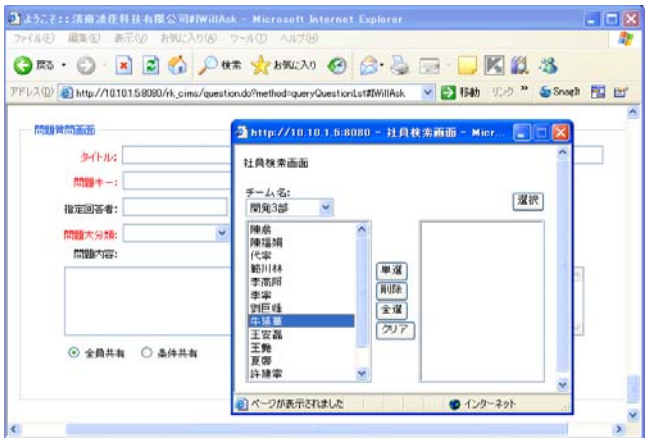

Fig. 5 Setting menu related to respondents

Fig. 6 shows the interface of BBS-style answer input that automatically records the name of respondent and time stamp. Auxiliary documents can be attached when making the inquiry.

Fig. 7 shows the interface of information retrieval, which provides narrowing retrieval by using option items such as team identifier of the inquirer and level of categories. In addition to the retrieval, valuable information is displayed such as the list of recent valuable or solved Q&A threads when a user starts to search as shown in red-colored oval area of Fig. 8. Other statistical information is also displayed, for instance, frequency of replies as to each person, respondent ranking and so on. Fig. 9 shows the screenshot of recent valuable or solved Q&A threads. Fig. 10 shows statistical information as

mentioned.

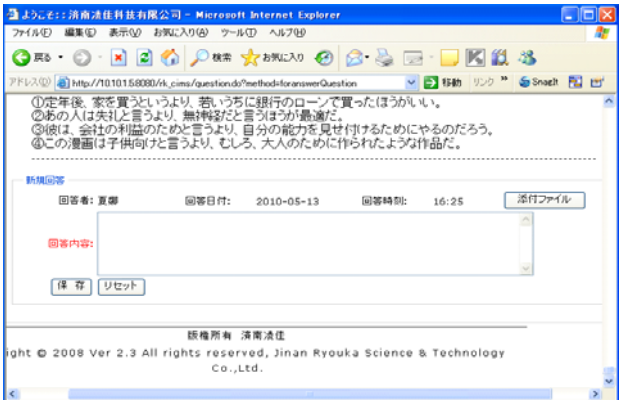

Fig. 6 BBS-style answer input interface

Fig. 7 Information retrieval interface

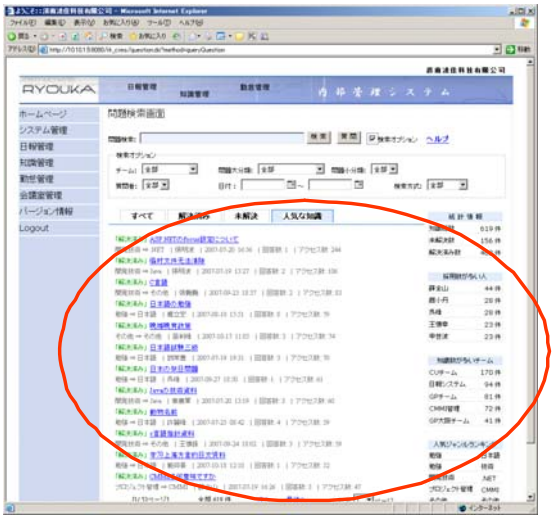

Fig. 8 Advanced information retrieval interface

Fig. 9 Screenshot of recent valuable or solved Q&A threads

Fig. 10 Screenshot of statistical information

As denoted the above, RKMS provides several options based on BBS-style Q&A and association retrieval functionalities, which are expected to make a user useful and convenient for daily use.

III. EVALUATION

We collected questionnaires on each functionality from 158 engineers which includes project managers (PM), project leaders (PL), project sub leaders (SL) and programmers (PG). The number of stored article threads is more than 10,000.

The questionnaire on this functionality is as follows.

- Frequency of use

Project members this RKMS in the RYOUKA enterprise

system for daily work. RYOUKA enterprise system consists of 6 types of systems such as “system management”, “daily reporting management”, working time management”, “human resource management”, and “RKMS”. Of course the usage depends on job titles, but RKMS is promoted to launch when a member comes to office.

- Valuable categories

As mentioned before, BBS database in the RKMS has 7 categories based on our preliminary investigation. Such categories are also used in the option items of “search”. Therefore it is crucial to know if they are sufficient or not to improve the RKMS.

- Accuracy of retrieval

The RKMS specific functionality is to support incorrect query input caused by novice engineers’ ability to Japanese language. This association retrieval functionality is originally designed compensation for short expressions in the public BBS. This time we introduce it to compensate incorrect expressions in the in-house BBS. Therefore it is also crucial to know if they are effective or not to improve the functionality.

- Usability of overall system

Usability is not measured quantitatively. We allowed engineers to denote positive or negative opinions in addition to 5-level scores.

A. Frequency of use

Table 1 shows the result of evaluation on "Frequency of use". Programmers use this system in daily work, especially low-level engineers are likely to ask others support when solving the problem. However, approximately 50% project members do not use everyday even if they launch the RKMS during the working time. Of course they are sometimes occupied by meetings, document writing and so on. Further investigation is necessary when they use and what duration they spend.

Table 1 Evaluation on “Frequency of use”

	PM	PL	SL	PG
2 or 3 times per a day	1.9	2.0	11.3	41.0
4 or 5 times per a week	0.6	1.9	5.0	20.6
2 or 3 times per a week	0.6	2.5	3.8	7.5
Not often	0.0	0.0	0.0	1.3

The unit is %.

B. Valuable categories

Table 2 shows the result of evaluation on "Valuable categories". In this case, multiple choices are allowed. "Basic knowledge study" marked high percentage regardless to engineer positions. This result indicates the system is expected to contribute self-education. The above results show that knowledge exchange is probably done among programmers by posting questions and answers.

As to programmers, they also marked both “daily business work” and “in-house system usage”. These categories indicate that re-engineering of workflow and improvement of the ROUKA enterprise system may be necessary to raise productivity. And “others” is marked as 12.5% as to programmers. Based on post investigation of BBS threads on this category, we concluded it is significant how to detect increasing thread topics and introduce new categories, which has not been supported in the current RKMS.

Table 2 Evaluation on “Valuable categories”

	PM	PL	SL	PG
development technology	3.1	9.4	18.0	56.3
basic knowledge study	3.1	9.4	15.6	65.8
daily business work	3.1	7.5	12.5	50.0
in-house system usage	1.3	3.1	9.4	62.5
business general knowledge	3.1	9.4	11.3	6.3
project management	3.1	9.4	15.0	9.4
others	0.6	1.9	3.8	12.5

The unit is %

C. Accuracy of retrieval

Table 3 shows the result of evaluation on "Accuracy of retrieval". Compared to the simple keyword retrieval, the association retrieval functionality is not so powerful than we expected. This is caused by rapid IT terminology changes, which involves abbreviation and compound terminologies. Therefore automatically collected terminology dictionary from Web will be inevitable to improve the accuracy.

Table 3 Evaluation on “Accuracy of retrieval”

	PM	PL	SL	PG
Extremely fine	0.6	3.1	6.2	30.0
Fine	1.3	5.0	7.5	9.4
Fair	1.3	0.6	2.5	12.5
Poor	0.0	0.6	1.9	15.6
Extremely poor	0.0	0.0	0.6	1.3

The unit is %

D. *Usability of overall system*

Table 4 shows the result of evaluation on "Usability of overall system". Practical usability is considered to achieve from the questionnaire.

Table 4 Evaluation on “Usability of overall system”

	PM	PL	SL	PG
Extremely fine	1.3	5.0	9.3	40.0
Fine	1.9	1.9	5.6	16.3
Fair	0.0	2.5	3.1	9.4
Poor	0.0	0.0	0.6	2.5
Extremely poor	0.0	0.0	0.0	0.6

The unit is %

As indicated in Table 4, programmers are mostly satisfied with usability through current various types of interfaces. However, system functionalities such as usage of search history and personal database for storing useful threads are expected development for further support.

IV. FUTURE WORK

RKMS discussed so far is initially intended to support novice software engineers' daily work from knowledge management points of views. However, through the evaluation process, project managers or project leaders consider this RKMS could be used as managers’ task enhancement. Posted inquiries, searched inquiries and replied answers indicate that projects needs some specific knowledge or some members are necessary to be educated in a certain category of IT, etc. To achieve it, automatic categorization of Q&A database shown in Fig. 1 is necessary without investigating whole threads.

Text categorization technology is applicable to some extent [7][8]. As indicated in [7], classification performances depend on training sets. Unfortunately our training sets are not trust-worthy due to incorrect word usage. Against such types of data, a new classification method with data mining techniques is proposed [9]. This research work handles classification of world wide web documents, which usually have more words and sentences than Q&A thread targeted in this paper. Therefore one of the future research issues is how to classify less words and sentences in data mining. Even if such classification is done, “class separability” is significant from managers points of views, because it affects proper decision-making. As for the “class separability”, new evaluation index is discussed as to treatise filtering system’s performance[10]. Although our target domain is different from treatise filtering, we think such index to measure the “class separability” may be designed based on [10].

Fig. 11 shows the above-mentioned extension of RKMS, which also supports managers’ task for planning their training programs and detecting possible troubles caused by lack of programmers’ knowledge.

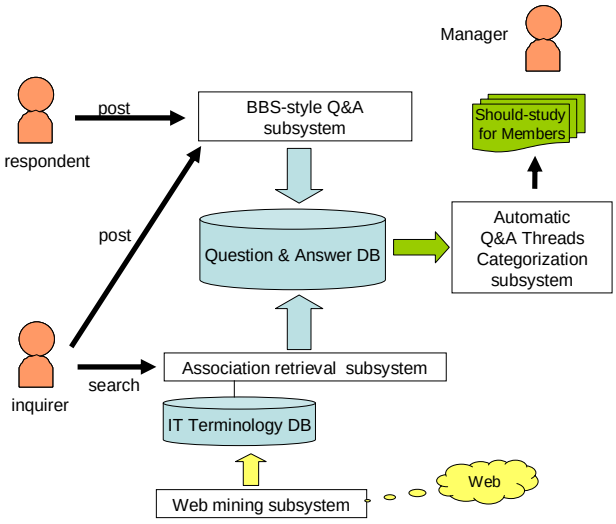

Fig. 11 Extended system configuration of RKMS

- Automatic Q&A threads categorization
  A Question and Answer thread consists of one question and a series of answers from respondents. Basic categorization technique is done by clustering techniques with similarity measurement. In this case, similarity is defined by using question-similarity and answers-similarity. Generally a series of answers are trust-worthy, because posted questions are incorrect words or misunderstood terminology. From this viewpoint, it is necessary to define similarity between various lengths of a series of answers. Since specific words or phrases are considered to characterize such answers, text mining may be used as extracting specific words or phrases automatically and similarity measurement can be achieved. As for the log of searched inquiries, we think option data such as a selected item from 7 categories is a key to grasp inquirers’ activity.

- Web mining
  Web mining is used for extracting trends on some keywords, finding important connected links, generating word corpus and so on. As indicated before, our “accuracy of retrieval” is expected to be improved with IT terminology dictionary. IT terminology varies from hardware to software in the whole Web. Our project members under offshore software development mostly need programming-related terminology. Therefore the most significant key technique is how to identify the Web page is related to such target domain.

- Association retrieval
  Current association retrieval functionality provides modified similarity judgment between a query sentence and article question sentences in BBS. By using IT

terminology dictionary, an inquirer's inputted query can be corrected by presenting some candidate words as correct ones. This correction leads to increase common words between a query sentence and article question sentences. Also this increase affects structural similarity detection and finally calculate modified cosine similarity index more precisely.

In addition to the above new featured extension, usage of search history and personal database for storing useful threads pointed in "Usability of overall system" is planned as future work.

V. Conclusion

This paper discusses BBS-based information management system, which handles engineers' practical knowledge in daily work of offshore software development. BBS is useful for simply exchanging opinions, however, added association retrieval functionality provides more powerful exchange and reuse of knowledge. From the evaluation, information retrieval functionality is crucial to enhance this in-house information management system. Also future work indicated managers' task support based on this system.

References

[1] Carnegie Mellon University, The Capability Maturity Model: Guidelines for Improving the Software Process, Addison Wesley Longman, Inc., 1994.
[2] K. Sparck-Jones and P. Willett(eds.), Readings in Information Retrieval, Morgan Kaufmann Publishers, 1997.
[3] 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.
[4] Y. Sakurai, et. al., A retrieval method of similar question articles from Web Bulletin Board, in Proc. of the 1st Int. Conf. on Software and Data Technologie (ICSOFT-2006), 2006, pp.238-243
[5] G. Salton, Automatic text processing: The transformation analysis and retrieval of information by computer, Addison-Wesley, 1988
[6] Li Cai, Zuoqi Wang, Yufeng Jiao, Masanori Akiyoshi, and Norihisa Komoda; Prototype of Knowledge Management System in Chinese Offshore Software Development Company, WSEAS Transactions on Information Science & Applications, Volume 5, Issue 3, 2008, pp.252-257.
[7] M. Ikonomakis, S. Kotslantis and V. Tampakas; Text Classification: A Recent Overview, in Proc. of the 5th WSEAS Int. Conf. on APPLIED INFORMATICS and COMMUNICATIONS, 2005
[8] Hana Kopackova; Text Categorization – Potential Tool for Managerial Decision-making, in Proc. of the 5th WSEAS Int. Conf. on APPLIED INFORMATICS and COMMUNICATIONS, 2005
[9] Hanan Ahmed Hosni Mahmoud Abd Alla; A Novel Efficient Classification Algorithm for Search Engines, in Proc. of the 8th WSEAS Int. Conf. on APPLIED INFORMATICS and COMMUNICATIONS, 2008
[10] Masao Yamamoto, Hiroshi Yajima and Hiroshi Kinukawa; A New Evaluation Index for Comparison of Treatise Filtering System's, WSEAS Transactions on Computers, Volume 2, Issue 3, 2003, pp.692-699
[11] Cabocha, http://chasen.org/ taku/software/cabocha/

Li Cai was born in China and received a bachelor of engineering degree from Harbin Institute of Technology, Harbin, China and a master of information science from Osaka University, Osaka, Japan in 1990 and 1999, respectively.
He worked for Hitachi Information Systems Ltd. as a engineer from 1999 to 2002. Then he established Jinan Ryouka in 2002, and is a president of Jinan Ryouka Science & Technology Co., Ltd. and a president of Beijing Ryouka Science & Technology Co., Ltd, China. Currently he is also a Ph.D candidate of the Department of Multimedia Engineering, Graduate School of Information Science and Technologies, Osaka University. His research area is software engineering.
He published papers such as 1)Prototype of Knowledge Management System in Chinese Offshore Software Development Company", WSEAS Transactions on Information Science & Applications, Volume 5, Issue 3, pp.252-257, 2008, 2) "A Knowledge Sharing and Managing System for Offshore Software Development Company", in Proc. of The 7th WSEAS International Conference on Applied Computer Science (ACS'07), pp.300-305, 2007, 3) "A Classification Method of Knowledge CARDS Represented in Japanese and Chinese at Offshore Software Development Company", in Proc. of the IADIS International Conference Applied Computing 2009, pp.63-67, 2009.
Mr. Cai is a member of IEEJ (The Institute of Electrical Engineers of Japan).

Zuoqi Wang was born in China and received a bachelor of engineering degree from Jilin University, Jilin, China in 1988.
He joineded Jinan Ryouka in 2003, and is a vice president of Jinan Ryouka Science & Technology Co., Ltd, China. Currently he is also a Ph.D candidate of the Department of Multimedia Engineering, Graduate School of Information Science and Technologies, Osaka University. His research area is software engineering.
He published papers such as 1)"Effectiveness evaluation of pre-review of software developed by novice engineers in China", IEEJ Transactions on Electrical and Electronic Engineering, Vol 4, No.6, pp.750-754, 2009, 2) "A Predictive Model of the Pointed Defects Rate in Software Pre-review for Novice Engineers in Chinese Offshore Company", in Proc. of 3rd Int. Conf. on Software and Data Technologies (ICSOFT2008), pp. 228-233, 2008, 3) "E-mail Log Analysis in Offshore Software Development", in Proc. of the 2nd Japan-China Joint Symposium on Information Systems, pp.29-32, 2009.
Mr. Wang is a member of IEEJ (The Institute of Electrical Engineers of Japan).

Masanori Akiyoshi was born in Japan and received a bachelor of engineering degree from Kyoto University, Kyoto, Japan and a master of engineering degree from Kyoto University, Kyoto, Japan and a doctor of engineering degree in 1985, 1987 and 1996, respectively.
He worked for Mitsubishi Electric Corp. as a researcher and a manager from 1987 to 2005. Then he joined Osaka University in 2005, and is an associate professor of the Department of Multimedia Engineering, Graduate School of Information Science and Technologies, Osaka University. His research area is business information systems, electronic commerce systems, and knowledge information processing.
He published papers such as 1) "Questionnaire-based Evaluation of Characteristics of a Community in SNS", WSEAS Transactions on Information Science & Applications, Volume 5, Issue 3, pp.237-245, 2008, 2) "Predicting Information Dissemination by Evaluating Propagation Distance of Information on Social Network", International Journal of Knowledge and Systems Sciences (ISKSS), Vol.5, No.1, pp.54-61, 2008, 3) "Analysis Support System of Open-ended Questionnaires Based on Atypical and Typical Opinions Classification", Studies in Informatics and Control, Vol. 18, No.3, pp.195-204, 2009.
Dr. Akiyoshi is a member of IEEE (The Institute of Electrical and Electronic Engineering) and IEEJ (The Institute of Electrical Engineers of Japan).

Norihisa Komoda was born in Japan and received a bachelor of engineering degree from Osaka University, Osaka, Japan and a master of engineering degree from Osaka University, Osaka, Japan and a doctor of engineering degree in 1972, 1974 and 1982, respectively.
He worked for Hitachi Ltd. as a researcher and a manager from 1974 to 1991. Then he joined Osaka University in 1991, and is a professor of the Department of Multimedia Engineering, Graduate School of Information Science and Technologies, Osaka University. His research area is business information systems, electronic commerce systems, and knowledge information processing.
He published papers such as 1) "A Model to Explain Quality Improvement Effect of Peer Reviews", IEICE (The Institute of Electronics, Information and Communication Engineers) tranzactions on Information and Systems, Vol.E93-D, No.1, pp.43-49, 2010, 2) "Selfish Constraint Satisfaction Genetic Algorithm for Planning a Long-distance Transportation Network", Journal of Computers, Vol.3, No.8, pp.77-85, 2008, 3) "Business scenario evaluation using Monte Carlo simulation on qualitative and quantitative hybrid model",

The Journal Electronics and Communications in Japan, Vol.170, Issue 3, pp.9-18 2010.
Dr. Komoda is a member of IEEE (The Institute of Electrical and Electronic Engineering) and IEEJ (The Institute of Electrical Engineers of Japan).