Regular Paper

DBSJ Journal, Vol.13, No.1
March 2015

An Automatic Scoring System for E-Reports based on Student Peer Evaluation using Groupware

Yuanyuan WANG♥ Yukiko KAWAI♥ Setsuko MIYAMOTO♦ Kazutoshi SUMIYA♠

Nowadays, many universities utilize groupware support for students to post and share their e-reports, and the students can browse and vote other students' reports in e-learning. Teachers then need to evaluate and grade all students' reports, but this will require a great deal of time and effort for a fair evaluation of the reports. Therefore, we develop an automatic scoring system for e-reports based on student peer evaluation by considering the relationship between voting and posting time of the e-reports, to promote the quality of the votes and prevent unfair votes. Then, the system provides a score ranking list of the reports based on a voting graph by analyzing the students who vote the reports, it is a grading tool to support teachers acquire the scores of the reports efficiently. Moreover, the system also enables students detect best reports easily. In this paper, we perform a student peer evaluation through groupware based on voting with a "Like" button in a course practice, and discuss an evaluation of our automatic scoring system's effectiveness compared to teachers' scoring.

1. Introduction

Currently, Web-based report systems, such as Bulletin Board Systems (BBSs) and groupware, are now one of the most frequently used tools for e-learning at universities and other educational institutions. Students then post and share their reports (essays: impressions of lectures) at anytime and from anywhere in a given period, i.e., after a lecture and before the next lecture; and the students can easily browse and vote other students' reports through these online systems. However, teachers need to review and grade all students' reports of each lecture, but this will require a great deal of time and effort for a fair approach to evaluate and grade the students' reports.

♥ Member Faculty of Computer Science and Engineering, Kyoto Sangyo University
{yuanw, kawai}@cc.kyoto-su.ac.jp
♦ Non Member School of Human Science and Environment, University of Hyogo
miyamoto@shse.u-hyogo.ac.jp
♠ Member School of Human Science and Environment, University of Hyogo
sumiya@shse.u-hyogo.ac.jp

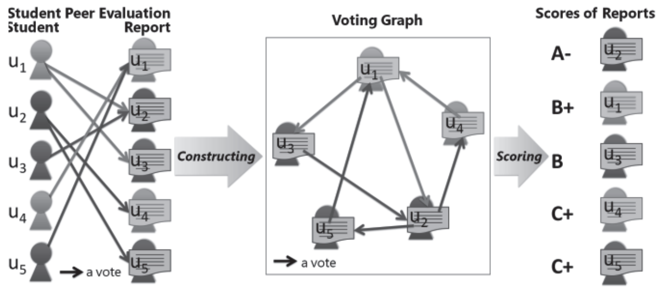

Fig. 1 Automatic scoring system for students' reports

Although most automatic scoring systems for document operational test question [1, 2] and essays [3, 4] have been developed, the mentioned problem has been never solved. These studies have focused on a specific, well-defined evaluation criteria to determine the answers are correct or not, or a standard format to measure the essay formats are appropriate or not. On the other hand, BBSs, groupware, or other online report systems are spread widely in e-learning. These online systems can be considered as a student peer evaluation platform, thus the student peer evaluation must be utilized for grading scores of their reports.

As depicted in Fig. 1, we develop an automatic report scoring system to enable students instead of teachers to grade scores of their reports by voting with each other based on student peer evaluation for their reports. It automatically provides a score ranking list of reports by analyzing the relationship between voting and posting time of the reports based on a voting graph of the reports, to promote the quality of the votes and prevent unfair votes. In this manner, more valuable reports are likely to receive more votes from other high-performing students and unfair vote-getting would be prevented by analyzing the relationship between voting based on the concept of PageRank [5], it can prevent "spam" on voting among students when they are friends. In addition, suppose that before the students post their reports, they often refer to previous others' reports. Our method then reduces the scores of the last posting reports by considering the posting time of the reports; it can redress the unfair scores of the reports caused by the posted order of the reports. For this, the voting graph is constructed by the votes between a student and his or her voted reports based on the student peer evaluation (center part of Fig. 1). In this paper, we first perform a student peer evaluation using groupware based on voting with a "Like" button in a course practice (left part of Fig. 1). Students can post their reports and browse others' reports. In addition, they can vote on the others' reports by pressing a "Like" button, when they think the reports are good [6]. The "Like" button can reduce the students' burden of evaluating others' reports without specific points. In our automatic report scoring system, teachers can efficiently acquire a score ranking list of students' reports based on student peer evaluation through groupware. Moreover, students can easily detect best reports from the score ranking list of their reports based on the students' viewpoints.

The next section describes an overview of our system. Section 3 explains our proposed report scoring method

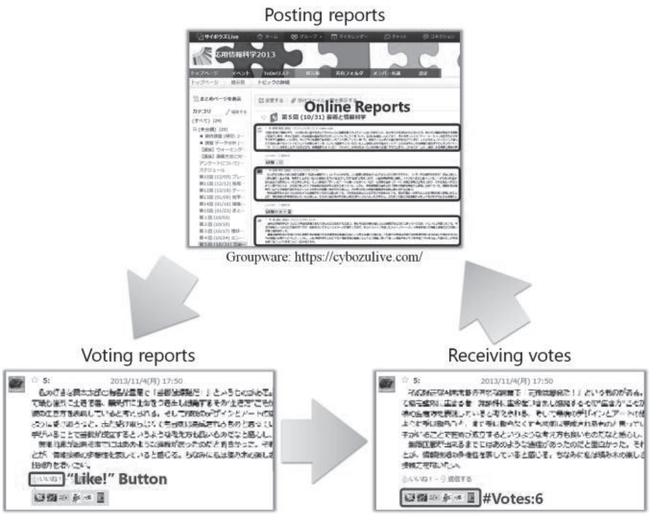

Fig. 2 Procedures for a course using groupware system

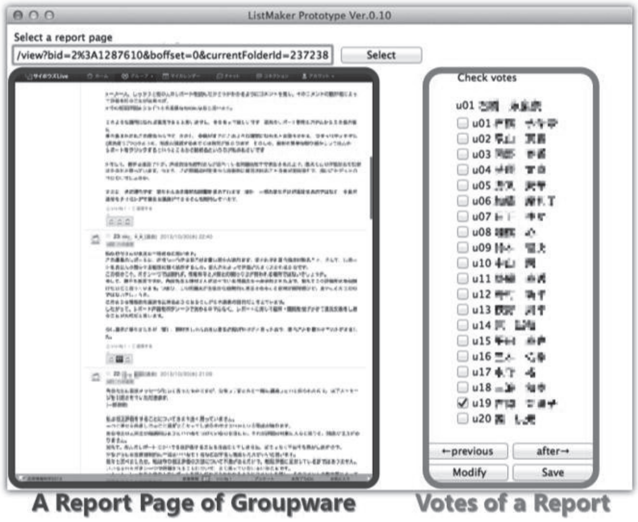

Fig. 3 System snapshot: reviewing and checking

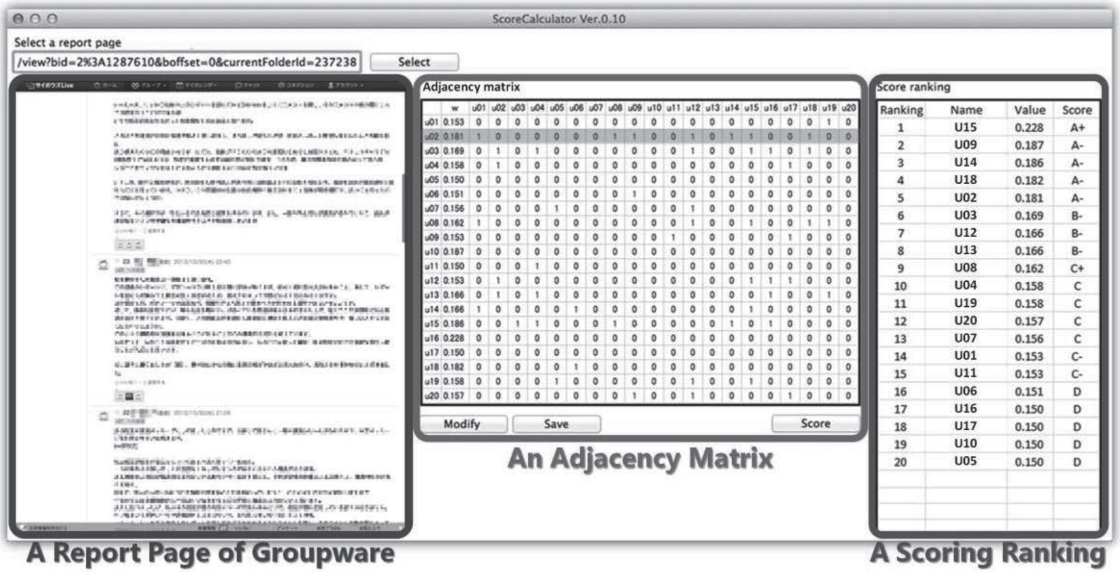

Fig. 4 System snapshot: an adjacency matrix and a scoring ranking

based on student peer evaluation. Section 4, we summarize the experimental results of our developed prototype system in a course practice. Finally, in Section 5, we conclude this paper with suggestions for future work.

2. System Overview

In the preprocessing, our system first accesses an existing groupware service (Cybozulive[1]). During a course using the groupware service is shown in Fig. 2, all students are required to complete the follow steps: 1) posting reports after lectures in a certain period; 2) browsing the reports of other students and voting with a "Like" button; and 3) receiving votes for their own reports.

To use our system for automatically grading scores of students' posted reports based on student peer evaluation by voting. The flow of the system is described as follows:

1. A teacher selects a lecture to review students' reports of this lecture, the system then presents the Web page of the reports through the groupware service, in which the teacher need to login to the groupware service (left part of Fig. 3). The teacher then checks and records the reports in a posted order.
2. When the teacher checks each report received whose votes in the check boxes and clicks a "Save" button (right part of Fig. 3), the system constructs a voting graph and returns its corresponding adjacency matrix (center table of Fig. 4) and the weight of the posting time of each report (column w in the center table of Fig. 4). For example, the weight of the posting time of student u02's report is 0.181, and the u02's report received 7 votes from students, u01, u08, u09, u12, u14, u15, and u18.
3. When the teacher clicks a "Score" button, the system presents a score ranking list of the reports, including

[1] https://cybozulive.com/

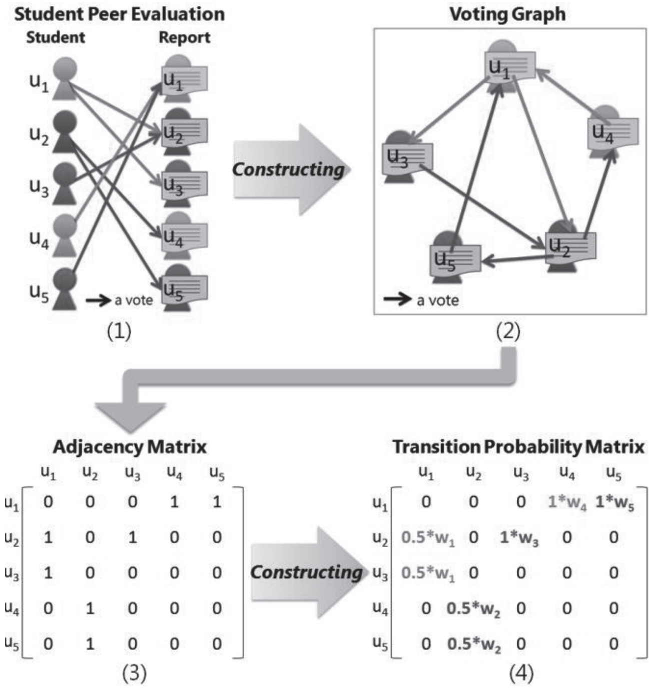

Fig. 5 Voting graph and transition probability matrix

ranking numbers, students' names of the reports, proper values of the reports, and the scores of the reports (right part of Fig. 4).

Therefore, the teachers can acquire the scores of the reports easily and efficiently. In addition, the students can browse the score ranking list of the reports and find the best reports easily and efficiently.

3. Report Scoring Method based on Student Peer Evaluation

After a user selects a lecture, the system first returns a Web page of this lecture's reports through groupware service. When the user checks and records the votes of each report, a voting graph, its adjacency matrix, and transition probability matrix are automatically constructed. Finally, scores of reports are calculated based on the transition probability matrix with the weight of the posting time of the reports.

3.1 Construction of Voting Graph and Adjacency Matrix

A voting graph is first constructed. The nodes of the directed graph consist of students' reports, and the links can be considered as votes from students for the others' reports. Therefore, the reports are voted from other students (back links), e.g., student u_1's report are voted by students, u_4 and u_5 (Fig. 5 (1)), and the students of the reports vote to other students' reports (forward links), e.g., u_1 votes the reports of students, u_2 and u_3 (Fig. 5 (1)). In our previous work, we developed a system that evaluates users who browse the Web pages based on their links between a user and his or her browsing pages [7, 8]. In this work, in order to evaluate the reports based on student peer evaluation; we focused on the students who vote on the others' reports.

If one student u_i is voting another student u_j's report, then, a link from u_i's report to u_j's report (arrows in Fig. 5 (2)), and the element of its corresponding adjacency matrix (u_j, u_i) is set to 1. As an example shown in Fig. 5 (3), the elements (u_1, u_4), (u_1, u_5), (u_2, u_1), (u_2, u_3), (u_3, u_1), (u_4, u_2), (u_5, u_2) become 1.

3.2 Construction of Transition Probability Matrix

We next describe the construction of the transition probability matrix from the adjacency matrix. For example, it can be transformed to a transition probability matrix as shown in Fig. 5 (4). In this work, we suppose that before the students post their reports, they often refer to previous others' reports, then, we should reduce the scores of the last posting reports. Then, w_i is the weight assigned to each student's report by considering its posting time, e.g., if the posting number of the report is in the last of the posted order, the weight of the report becomes low. Therefore, more valuable reports are likely to post on the front and receive more votes from other high-performing students.

3.3 Score Calculation

We next calculate the scores of students' reports by the following formula.

$$S(r) = (1-d) + d * \left(\frac{S(v_1)}{T(v_1)} * w_1 + \cdots + \frac{S(v_i)}{T(v_i)} * w_i \right) \quad (1)$$

- r: a student's report, e.g., the report of u_1, u_2, u_3, u_4, or u_5
- v_1, ..., v_i: the set of votes of r, e.g., the votes of u_1's report are from u_4 and u_5
- $S(v_i)$: the numerical weight of each vote contained in the set of the votes of a student's report, e.g., supposed $S(v_1)$=1 that is the numerical weight for a vote of u_2's report from u_1 based on the set initial value 1
- $T(v_i)$: the number of votes from a student, e.g., $T(v_1)$=2 that indicates there are two votes from u_1 to the reports of u_2 and u_3
- w_i: the weight of posting time of a report, e.g., w_1 is the weight of u_1's report. There are two methods:

1. $$w_i = \frac{1}{n_i} \quad (2)$$

   n_i: the posting number of a report, e.g., n_1 is the posting number of u_1's report. Eq.(2) returns the reciprocal of the posting number. If n_i is large, it denotes the report is posted in the last, w_i may become low.

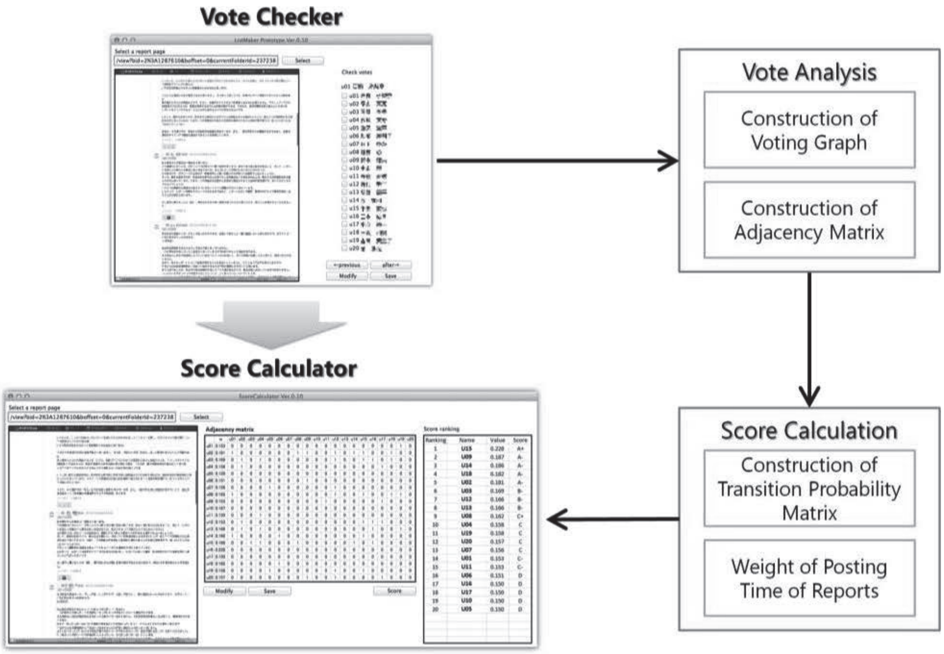

Fig. 6 System configuration diagram

2. $$w_i = \frac{m - n_i + 1}{m} \quad (3)$$

   m: the total number of students' reports in a course, e.g., reports of 20 students who participated in a course, then, m=20. Eq.(3) returns the ratio of the posting number of the report in all reports. The highest value of it is 1, since the report is the first posted report.

- d: a damping factor adjusts the derived value downward. Various studies have tested different damping factors, but it is generally assumed that the damping factor is set at approximately 0.85

Initially, the weight of each vote is 1, if a student votes multiple report, the weight distributes through each vote evenly by the function $T(v_i)$, e.g., u_1 votes reports of u_2 and u_3, $S(v_1)$=1, $T(v_1)$=2, then, the weight of each vote from u_1 to the report of u_2 or u_3, becomes 0.5*w_1 (Fig. 5 (4)). Finally, we normalize the score of each report into a grading scale range from A+ to D-, and F (a failing grade).

Although the calculation formula is similar to PageRank [5] and ObjectRank [9], the sense of our proposed report scoring method is different from in that scores of reports are calculated based on a voting graph of weighted links for the reports, instead of only links to Web pages [5] or only database entries [9]. By providing a score ranking of the reports, the system can support teachers review the scores of the reports efficiently, and it can also help students discover the best reports easily.

4. Evaluation

4.1 Implementation of Prototype System

Based on the method described above, we have built a prototype system to support report scoring (see Fig. 6), using Python 2.7.8. The interface is programmed using Tkinter (GUI: graphical user interface). The prototype system has two stages: analysis and calculation. Firstly, in the analysis stage by using our developed Vote Checker, we first construct a directed graph of votes, which consists of students' reports as the nodes by analyzing how many votes are received of each report and who voted. Then, its corresponding adjacency matrix is constructed. In this stage, the teachers can review the reports through groupware and check the votes of each report.

Secondly, in the calculation stage by using our developed Score Calculator, the adjacency matrix is transformed to a transition probability matrix with the weight of the posting time of the reports, then, the scores of the reports are calculated based on the transition probability matrix, and the scores are ranked in an order from high to low. At this stage, the teachers can acquire a score ranking of the report efficiently. Moreover, the students can discover the best report from the score ranking easily.

4.2 Experiment and Results

In this section, we present our findings from the results of our proposed report scoring method based on student peer evaluation in a course practice. This is a course of Applied Informatics, which consists of 10 lectures on different topics, and 20 students who participated in this course need to post their reports about impressions of each lecture after the lecture and before the next lecture. Using the "Like" button as a vote through an online groupware based on the student peer evaluation, (1) the students must to browse any other's report (need not to browse all others' reports) and vote on it, when they think it is good; (2) each student votes at least one report and up to five reports. We then calculated the scores of the reports by the following methods:

I. Baseline: counting the sum of the number of "Like" from students

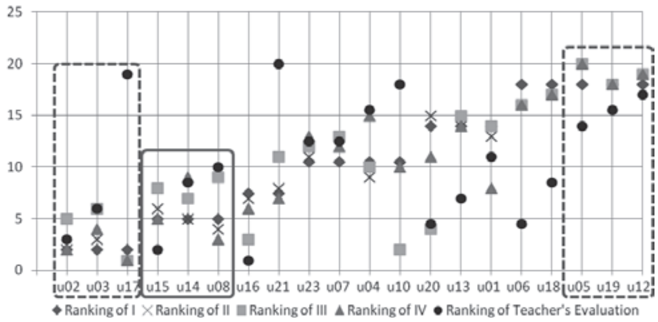

Fig. 7 Correlation diagram

II. Previous: using Eq. (1) without the weight w_i by considering the quality of the votes only [10];

III. Proposed 1: using Eqs. (1) and (2) by considering both the quality of the votes and the posting time

IV. Proposed 2: using Eqs. (1) and (3) by considering both the quality of the votes and the posting time

As a correlation diagram of the scoring rankings of all reports by those above methods and the teacher's evaluation of Lecture #3 is shown in Fig. 7, the horizontal axis denotes student numbers of the reports in an order of the ranking based on baseline I, and the vertical axis denotes the ranking number. Here, the teacher's evaluation as a correct evaluation criterion of our motivation that reduces the teacher's burden of grading the students' reports, and the teacher emphasized on the content of the report to evaluate them. The results and our findings were summarized as follows:

- Although, some reports gained the same number of votes by baseline I; their scores were different by our previous method II, and our proposed methods, III and IV. For example, the scores for the reports of students, u15, u14 and u08, were identical by I (frame in Fig. 7). However, they were different based on II, III, and IV.
- The top ranked reports have high scores by all above methods, and they were correlated with the teacher's evaluation. For example, the scores of the top ranked three reports of students, u02, u03, and u17, were all high by I, II, III, or IV (left dashed frame in Fig. 7). Except u17's report, the scores of the reports of u02 and u03 were all high by the teacher's evaluation.
- The lowest ranked reports have low scores by all above methods, and they were correlated with the teacher's evaluation. For example, the scores of the lowest ranked three reports of students, u05, u19, and u12, were all low by I, II, III, or IV (right dashed frame in Fig. 7), and they were also low by the teacher's evaluation.

From this experimental result, we considered that the student peer evaluation becomes more meaningful when using a transition probability matrix of the methods, II, III, and IV; since more valuable reports are likely to receive more votes from other high-performing students and unfair vote-getting would be prevented by the student peer evaluation when the students are friends.

For evaluating each report scoring method, we also calculated the Spearman's rank correlation coefficient ρ [11] between the score rankings by the teacher and those above methods of Lecture #3. The correlation value is calculated as follows:

Table 1 Correlation results

Report Scoring Method	Correlation Value
I. Baseline	0.464
II. Previous	0.456
III. Proposed 1	**0.475**
IV. Proposed 2	**0.507**

$$\rho = 1 - \frac{6 \times \sum D^2}{N^3 - N}$$

Function D that calculates the difference of the rank number of the same student's report between two rankings. N denotes the number of the students' reports in a ranking (in this experiment, N = 20). The correlation value ranges from -1 to 1, where -1 indicates that two rankings are completely reversed, whereas 1 indicates that the rankings are exactly the same.

Furthermore, we summarized the correlation result of each scoring method (I, II, III, or IV) and the teacher's evaluation that is listed in Table 1, and the results can be explained as follows:

- The correlation values of all methods and the teacher's evaluation were not very close to 1.
- The correlation value of our previous method II and the teacher's evaluation was a little lower than that of baseline I.
- The correlation values of our proposed methods (III and IV) and the teacher's evaluation were higher than those of baseline I and our previous method II.

Although our proposed methods (III and IV) did not reach a very high correlation value, this experiment indicated that our proposed report scoring calculation methods have the potential to support teachers easily and efficiently evaluate students' reports based on the student peer evaluation using groupware in Japan. Since our proposed methods by considering both the quality of the votes and the posting time of the reports (strategy aspect of utilizing the cultural psychology of Japanese), achieved a good performance compared with the conventional scoring calculation method (I) by counting the total number of the votes, or our previous scoring calculation method (II) by considering the quality of the votes only.

Future work will deeply analyze the correlation between our proposed methods based on the student peer evaluation and the teachers' evaluation with large datasets in different courses with a variety of topics. In the teachers' evaluation, we need to adopt different methods to grade the score of the reports by considering the posting time of reports or not. In order to verify the reliability of our proposed methods, we should do an interview of the students who participated in the student peer evaluation of their reports.

5. Conclusions

In this paper, we proposed an automatic scoring system for students' reports based on student peer evaluation using groupware. In a course practice, students performed a peer evaluation for their reports by voting for valuable reports using a "Like" button. Therefore, it is not only a total number of votes for evaluating the reports, but also considering both the relationship between voting and the posting time of the reports. It can lead to a new method rooted in the indigenous culture of review by the student peer evaluation.

In the future, we need to measure inter-rater reliability of our proposed report scoring methods by combining student peer evaluation and teachers' evaluation with content analysis (e.g., TF-IDF values). Furthermore, in order to promote the collaboration between student peer voting, we also plan to attach a communication function of our previous developed system [12].

[Acknowledgments]

This work was supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE), and JSPS KAKENHI Grant Number 26280042.

[References]

[1] L. Hao and Q. Guan, "The Design of Automatic Scoring System for Document Operational Test Question of MS-Office", in *Proc. of 2014 International Conference on Education, Management and Computing Technology* (*ICEMCT 2014*), pp. 393-396, 2014.

[2] H. Liu, "Design and Development of the College English Test System Based on Network", in *Journal of Applied Sciences*, 14(11), pp. 1168-1173, 2014.

[3] P. W. Foltz, L. A. Streeter, K. E. Lochbaum, and T. K. Landauer, "Implementation and Applications of the Intelligent Essay Assessor", in *A Handbook of Automated Essay Evaluation: Current Applications and New Directions*, pp. 68-88, 2013.

[4] Z. Rahimi, D. J. Litman, R. Correnti, L. C. Matsumura, E. Wang, and Z. Kisaa, "Automatic Scoring of an Analytical Response-To-Text Assessment", in *Intelligent Tutoring Systems*, 8474, pp. 601-610, 2014.

[5] S. Brin and L. Page, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", in *Computer Networks*, 30(1-7), pp. 107-117, 1998.

[6] J. Iturrioz, I. Azpeitia, and O. Diaz, "Generalizing the Like Button: Empowering Websites with Monitoring Capabilities", in *Proc. of the 29th Symposium On Applied Computing* (*SAC 2014*), pp. 743-750, 2014.

[7] Y. Shiraishi, J. Zhang, Y. Kawai, and T. Akiyama, "Simultaneous Realization of Page-centric Communication and Search", in *Proc. of the 21st ACM International Conference on Information and Knowledge Management* (*CIKM 2012*), pp. 2719-2721, 2012.

[8] Y. Shiraishi, J. Zhang, Y. Kawai, and T. Akiyama, "Proposal of Combination System of Page-centric Communication and Search", in *Proc. of the 1st International Workshop on Online Social Systems* (*WOSS 2012*) *in conjunction with VLDB 2012*, 2012.

[9] A. Balmin, V. Hristidis, and Y. Papakonstantinou, "Objectrank: Authority-based Keyword Search in Databases", in *Proc. of the 30th International Conference on Very Large Data Bases* (*VLDB 2004*), pp. 564-575, 2004.

[10] Y. Wang, Y. Kawai, S. Miyamoto, and K. Sumiya, "A Students' Mutual Evaluation Method for their Reports using Pagerank Algorithm", in *Proc. of the 22nd International Conference on Computers in Education* (*ICCE 2014*), pp. 113-115, 2014.

[11] J. L. Myers and A. D. Well, "Research Design and Statistical Analysis", in *Lawrence Erlbaum Associates*, 2003.

[12] Y. Shiraishi, J. Zhang, Y. Kawai, and T. Akiyama, "Web Page Centered Communication System based on a Physical Property", in *Proc. of the 19th International Conference on Database Systems for Advanced Applications* (*DASFAA 2014*), pp. 513-522, 2014.

Yuanyuan WANG

She is a researcher at Faculty of Computer Science and Engineering, Kyoto Sangyo University. She received her MS and PhD degrees in Human Science and Environment from University of Hyogo, in 2011 and 2014, respectively. Her research interests include e-learning systems, multimedia databases and human-computer interaction. She is a member of Information Processing Society of Japan, the Japanese Society for Artificial Intelligence, and the Database Society of Japan.

Yukiko KAWAI

She is an associate professor at Faculty of Computer Science and Engineering, Kyoto Sangyo University. She received the MS and PhD degrees in Information Science and Technology from Nara Institute of Science and Technology, in 1999 and 2001, respectively. Her research interests include data mining, information analyzing and Web information retrieval. She is a member of Information Processing Society of Japan, the Database Society of Japan, and the Institute of Electronics, Information and Communication Engineers.

Setsuko MIYAMOTO

She is a professor at School of Human Science and Environment, University of Hyogo. She received her PhD degree in International Communication from Nagoya University in 1998. Her research interests include development and evaluation of multimedia course materials. She is a member of Japan Society for Educational Technology, Japan Association for Educational Media Study, Communication Association of Japan, and Society for Intercultural Education Training and Research.

Kazutoshi SUMIYA

He is a professor at School of Human Science and Environment, University of Hyogo. He received his PhD in Engineering from Kobe University in 1998. His research interests include Web information systems, multimedia databases, and broadcasting computing. He is a member of IEEE Computer Society, ACM, the Institute of Image Information and Television Engineers, Information Processing Society of Japan, the Database Society of Japan and the Institute of Electronics, Information and Communication Engineers.