Content Analysis to Detect The Role Behaviors of Student in Online Discussion

Erlin
Department of Information Technology
STMIK-AMIK Riau
Pekanbaru, Indonesia
erlin@stmik-amik-riau.ac.id

Rahmiati
Department of Informatics Management
STMIK-AMIK Riau
Pekanbaru, Indonesia
rahmiati@stmik-amik-riau.ac.id

Unang Rio
Department of Information Technology
STMIK-AMIK Riau
Pekanbaru, Indonesia
unang@stmik-amik-riau.ac.id

***Abstract*—Online discussion is a powerful way to conduct online conversation and a significant component of online learning. Online discussion can provide a platform for online learners to communicate with one another easily, without the constraint of place and time. In an online discussion, the students communicate a common interest, exchange information, share ideas, and assist each other in text/transcript form. So far, content analysis is a popular method for analyzing transcripts. However, using content analysis in computer supported collaborative learning (CSCL) or computer mediated communication (CMC) research focused on the surface of the transcripts. Usually, content analysis is employed to categorize news article, product reviews and web pages. Therefore, this study proposed content analysis to a deeper level is to detect the role behavior of students in an online discussion based on a conversation in text form. The findings showed that this method provides more meaningful students' interaction analysis in term of information on communication transcripts in online discussion. Educators can assess the contribution of students and can detect the role behavior of the student based on their conversation in transcript form; whether the role behavior as a mediator, motivator, informer, facilitator, or as a questioner.**

Keywords—content analysis; role behavior; student; online discussion

I. Introduction

Online discussion is a significant component of e-learning and a powerful way to conduct online conversation. These online discussions are now commonly utilized as the means of promoting interactions between distance learning course members in tertiary learning [1]. Numerous web-based courses depend on online discussions as the computer mediated communication (CMC) tool to increase learning due to the discussions' capability to sustain high levels of thinking and to offer a convenient and flexible communication forum to engage students actively [2].

The online discussions may comprise replies to the educator and to the other students, questions, arguments, discussions, debates, or to deliberate ideas and thoughts. As a consequence, there are several role behaviors of student in an online discussion; a student who asks questions or ask for help/advice in solving the problem, a student who tries to get agreement between student who disagree with each other and able to mediate disputes among them, a student who tries to make something clearer or easier to understand by giving a reason and example; a student who gives information or providing advice to the other student, etc.

However, the role behavior of students in an online discussion has not got serious attention from the educators. In many instances, only statistical information, such as frequency of postings is encompassed, but this is not very useful measurement of the quality of participation among students [3]. It is certainly not fair to the students who send or posted a lot of information, but only being considered as one contribution of participation. Students that posted or replied many ideas and information with replied many questions should not be treated equivalent level of participation.

One of the most popular methods to analyze discussion forum is content analysis [4], [5], [6]. The purpose of content analysis is to reveal information that is not situated on the surface of transcript [7]. Many researchers in the field argue for using the content analysis as a vehicle for classifying, analyzing and determining communication transcripts [7]. Unfortunately, analysis in CSCL or CMC research focused on the surface level characteristic of the communication [8].

Therefore, this study employs content analysis to detect the role behavior of students in an online discussion. Using content analysis offer a solution to analyze the text that have been posted by students. Thus, educators can assess the contribution of students and can detect the role behavior of the student based on their conversation in transcript form; whether the role behavior as a mediator, motivator, informer, facilitator, or as the questioner.

II. Description of the Method

A Content analysis is a method of a designed rule because it breaks data into meaningful units. It is frequently used for categorizing, analyzing, and testing message transcripts to different facet of communication and knowledge. De Wever et al., in [7] have argued that it's a technique commonly used in a formal computer mediated communication of educational setting where it's needed in situations of analyzing the transcripts of asynchronous discussion. It aims to separate data's that were not sited from the transcripts and allows exploration on the issues of "Is there evidence of knowledge construction among learners in an asynchronous online environment as revealed by their discussion?"

Content analysis is a well known method for measuring quality in conversation on forum postings [1]. A central idea in content analysis that the many words of the text are classified into much fewer content categories. Each category may consist of one, several, or many words. Words, phrases or other unit of text classified in the same category are presumed to have a similar meaning.

In recent years, research has shown that there have been extensive studies and actively explored various machine learning of content analysis. Among these are Bayesian network classifier [9], *k*-nearest neighbor classifier [10], decision tree [11], Neural Network [12] and Support Vector Machine [13]. Although there are many approaches for text document categorization, support vector machine (SVM) and Neural Network (NN) are two popular approaches considered so far. Based on experiments that have been conducted on these two approaches [14], the result shows that for text message categorization in online discussion, the performance of SVM outperform NN in term of error rate and precision; and falls behind NN in term of recall and F-measure. Therefore, this study employs SVM approach to categorize text as a basic for detecting role behavior of students.

SVM is one of the relatively new methods compared with other methods, but has given a better performance in various application fields such as image processing, handwriting and text classification. Joachims has successfully applied SVM to text categorization and achieved an outstanding improvement over another method [15]. He argued that SVM is an appropriate method for text categorization because SVM handles high dimensional feature spaces and few relevant features, which are the main properties of text categorization.

The simple concept of SVM can be explained as an attempt to find the best hyperplane (*h*) which serves as the dividing two classes in the input space. It tries to separate the two sets of training data by hyperplane that maximizes the margin (distance between the hyperplane and the closest point). Figure 1 shows some pattern of linearly separable data that is a member of two classes: +1 and -1. Pattern that joined in class -1 symbolized by the green color (boxes), whereas pattern in the class +1, symbolized by the yellow color (circle).

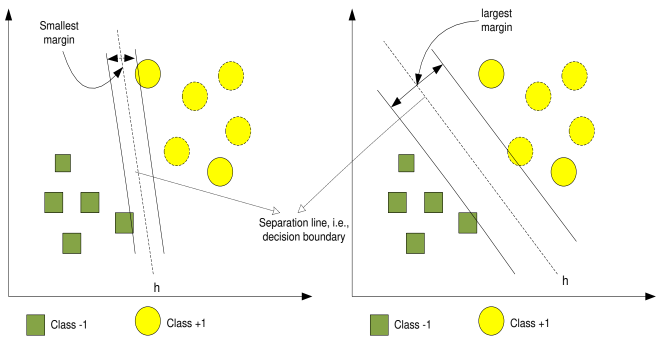

Fig. 1. Pattern of Linearly Separable Data of Support Vector Machine

III. Methodology

In order to pilot test the efficacy of content analysis for detecting the role behavior of students in an online discussion forum, the recorded transcripts of students' online discussion were coded using the Soller's model [16]. The selected transcripts of subject SCJ2013-01 2008/2009: Data Structure and Algorithm held on Moodle as a learning management system (LMS) in e-learning was examined for one thread. This thread was chosen because it has highest replies than other threads. There were 12 students completed the thread discussion.

A. Data Collection

Data for this study were transcripts of students' discussion using the threaded discussion tool on Moodle. There were three phases was occurred in the discussion. In this research, each phase was treated as phase: day 1 to day 6 was beginning phase; day 7 to day 12 was the middle phase; and day 13 to day 18 was the end phase. In the whole discussion threaded, the students posted 137 messages and break it into 394 sentences. The transcripts were important to analyze the dynamics of online discussion and what kind of feedback from one another. The transcripts were generated into text files. Transcripts were used to gather information about the threads of interactions. Table 1 shows the number of messages in each phase.

TABLE I. Distribution of Posted Messages and Sentences in Term of Phase

	Beginning Phase	Middle Phase	End Phase	Total
# of posted message	63	28	46	137
# of posted sentences	143	95	156	394

B. Data Analysis

Content analysis was used to analyze the data. Multiple sources of information and analytical technique are commonly employed in case studies to triangulate for strengthening the internal volatility and reliability of the results.

The qualitative data were analyzed using content analysis and choose sentence as a single unit of meaning and would be validated by two coders to calculate the reliability of coding categories. The log files of online discussion were segmented to extract the segments in which the students shared their knowledge. All of the messages in the sentence form of this threaded were analyzed according to the collaborative learning view [16] and each of them was classified into one of eight subcategories of collaborative learning skills.

C. Experiments

Students communicate each other in an online class through the discussion forum. Topic of discussion is usually provided by the lecturers or instructors to encourage the students to be more active in interaction for solving the problems. Students can exchange information, share ideas and assist each other. All of conversations among students will be stored in a database in the form of collection transcripts or the more popular called the corpus data. These transcripts are often used by educators in assessing the level of participation and quality of communication that occurred during the discussion.

Generally, corpus data is in a message form. Since this research chooses the sentence as a single unit analysis, hence message segmentation is needed to split the message into many sentences. Further, this sentence will be analyzed by human coder to categorize it into eight categories. The results of the human content analysis will be examined using multiple reliability coefficients. The reliability of human coder is needed to determine how well the human coded the list of corpus data based on coder training.

Armed with a number of human-code sentence that was saved in knowledge based, the support vector machine is ready to be trained. The content analysis using SVM contains three sub-components, i.e. pre-processing (tokenization, stop word removal, stemming and feature weighting); dimensionality reduction (feature selection and text representation); and text classifier (support vector machine as text classifier) and training data.

Classifier or algorithm cannot be directly interpreted the text. Texts should first be transformed into a representation suitable for the classification algorithms to be applied. In order to transform a text into a feature vector, pre-processing is needed. Moreover, feature selection also called term space reduction (TSR) was employed to reduce the dimension of the input. This research reduced the size of dimension by computing the document frequency (*DF*). All potential features are ranked in each category based on the term occurs in the sentence. The top features for each category are chosen as its feature set.

Next step, text representation to transform the text into a representation suitable for categorization algorithms to applied. In this research, sentences are represented by the widely used vector-space model. In this model, create a space in which both texts and terms are represented by binary vector, based on term frequency, and indicate the presence or the absence of a particular term in the texts. The data in vector space model are ready to be trained by support vector machine for categorizing the text. As a result a corpus data will be categorized into a certain category for analyzing the transcripts of students. From that results, can be seen which category is most widely used by students. In addition, it can also be used to determine the role of each student by analyzing the highest category that was involved during an online discussion.

IV. RESULTS AND DISCUSSIONS

Altogether we analyzed 137 messages containing 394 sentences, which were posted by 12 students. The discussion was started by student 1.

A. *Evaluation of Collaborative Learning Category*

Content analysis was performed on one threaded of online asynchronous discussion transcripts contributed by twelve students. Statistical comparison was restricted to the number of sentences and frequency of each main category and sub category on the level of the member groups in online discussion. All the transcripts that were produced on the online discussion were categorized automatically by the content analysis. To demonstrate the capability of the content analysis, the evaluation of collaborative learning category would be conducted. In the whole of the discussion threaded, this research evaluated 137 messages containing 394 sentences.

Table 2 below shows the 394 sentences that were posted by twelve students have been categorized automatically by content analysis into eight categories. From this table, it can be seen that the highest category that was played by students were informed which consisted of 88 sentences (22.34%). It indicates that most of the students share their ideas and information to other students in completing an assignment that has been given previously by the instructor. On the other hand, the lowest category that was played by students were mediating consisted of 3 (0.76%), meaning that few students asking the instructor or another student in solving the different opinions among them. In addition, the student also preferred to ask their peers if they need assistance in completing a task or question that has given by facilitators, educators or other students in their group.

Furthermore, the most frequently involved in main category was “active learning” (46.20%) and the most frequently used in sub-category was “inform” (22.34%). Only 20.30% of the ideas revealed the creative conflict skill (positively or negatively) about a comment or suggestion. Next, 33.50% of the ideas revealed conversation skill (i.e. acknowledge, maintenance, task) that shows coordinate the group process on the task, appreciation of the other peer comment and support the group cohesion.

TABLE II. CATEGORY STATISTIC OF COLLABORATIVE LEARNING SKILL

Category Statistic

Student No.	Student Name	Creative Conflict				Active Learning						Conversation						Total
		Mediate	%	Discuss	%	Motivate	%	Inform	%	Request	%	Acknowledge	%	Maintenance	%	Task	%	
1	Student 1	0	0.00	4	1.02	5	1.27	5	1.27	6	1.52	14	3.55	3	0.76	15	3.81	52
2	Student 2	0	0.00	3	0.76	3	0.76	6	1.52	3	0.76	3	0.76	1	0.25	3	0.76	22
3	Student 3	0	0.00	12	3.05	2	0.51	16	4.06	9	2.28	7	1.78	6	1.52	2	0.51	54
4	Student 4	0	0.00	3	0.76	4	1.02	9	2.28	11	2.79	3	0.76	4	1.02	2	0.51	36
5	Student 5	1	0.25	4	1.02	2	0.51	4	1.02	0	0.00	1	0.25	1	0.25	1	0.25	14
6	Student 6	1	0.25	7	1.78	1	0.25	6	1.52	6	1.52	5	1.27	7	1.78	4	1.02	37
7	Student 7	0	0.00	8	2.03	2	0.51	1	0.25	5	1.27	6	1.52	2	0.51	2	0.51	26
8	Student 8	0	0.00	3	0.76	1	0.25	5	1.27	9	2.28	4	1.02	4	1.02	2	0.51	28
9	Student 9	1	0.25	15	3.81	1	0.25	11	2.79	5	1.27	0	0.00	3	0.76	1	0.25	37
10	Student 10	0	0.00	7	1.78	1	0.25	5	1.27	4	1.02	1	0.25	5	1.27	1	0.25	24
11	Student 11	0	0.00	6	1.52	2	0.51	11	2.79	6	1.52	4	1.02	5	1.27	2	0.51	36
12	Student 12	0	0.00	5	1.27	2	0.51	9	2.28	4	1.02	3	0.76	4	1.02	1	0.25	28
	Total	3		77		26		88		68		51		45		36		394
	Percentage	0.76		19.54		6.60		22.34		17.26		12.94		11.42		9.14		100

B. *Evaluation of Collaborative Learning Role Behavior*

In the evaluation of communication transcripts, it was decided to have an investigation to what extent the role of students in collaborative learning based to their posting. In order to analyze the student according to the role that was played during discussion, it is important to analyze the posting of student belonging into one category (i.e., inform, discuss, maintenance).

Eight categories of roles in collaborative learning can be distinguished: mediator, clarifier, motivator, advisor, questioner, recognizer, keeper and facilitator. Each of which is comprised of several different roles. However, these are based on pertain to roles that students can perform during collaboration. Table 3 offers brief descriptions of the role behavior of students during discussion.

TABLE III. THE ROLE BEHAVIOR OF STUDENT IN COMMUNICATION TRANSCRIPTS

Highest in Category	Role Behavior	Description	Example
Mediate	Mediator	a student who tries to get agreement between student who disagree with each other	Let's ask our instructor
Discuss	Clarifier	a student who tries to make something clearer or easier to understand by giving a reason and example	I disagree because the system is so efficient and that's why people attracted to it
Motivate	Motivator	the student who causes the emergence of positive motivation to others to carry out something or driving	Do not give up, you have time to do your assignment
Inform	Advisor/ Informer	a student who gives information or providing advice to the other student	Do a triple check before submitting the final report at the end of this week
Request	Questioner	a student who asks questions or ask for help/advice in solving the problem	Can you tell me more about how to increase the image resolution?
Acknowledge	Recognizer	the student who informs his/her peer that they read or give appreciation for peer's comment/advice	My thanks to all of you for your help
Maintenance	Keeper	the student who maintains group cohesion and peer involvement	I apologize for the delay in response
Task	Facilitator	a student who acts as a facilitator that helps a process take place	Let's discuss and share opinion about this case

The data from twelve students who naturally played different roles of creative conflict skill during online discussion can be illustrated in figure 2. Qualitative analysis of the transcripts shows that the student 5, student 6 and student 9 played the role of mediators; they tried to get agreement between student who disagree among them and interposed between students at variance for the purpose of reconciling them by asking the instructor. Student 9 also the most widely played a discuss category, hence, he played the role behavior of a clarifier; make something clearer or easier and (more) comprehensible to understand by giving reasons about comments and suggestion to other students.

The students who played the role of mediators usually tried to end a disagreement between two or more students by talking to them and trying to find things that everyone can agree. They asked the instructor or other students to mediate in the dispute and succeed in finding a solution to a disagreement between students. Furthermore, the students who played the role of classifiers are the students who like to argue or discuss by giving the reason for what they have proposed. Besides that, they also offered another alternative if their suggestions were not approved. In simple words, this student describes something by giving a reason.

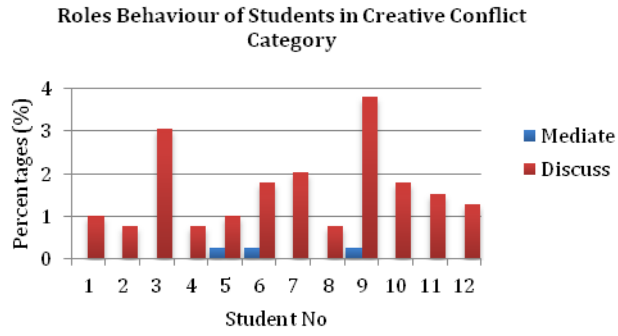

Fig. 2. Roles Behavior of Students in Creative Conflict Category

The role of active learning skill that was played by twelve students can be seen in figure 3. This figure shows that student 1 played the role of a motivator of collaborative learning. She was active in giving and providing positive feedback and reinforcement to others in the group. She was also making specific recommendations to other students. While student 3 leading in giving and sharing information to others, that made her to have a role of advisor or informer, student 4 leading in asking several clarification questions that made her to have a role of a questioner.

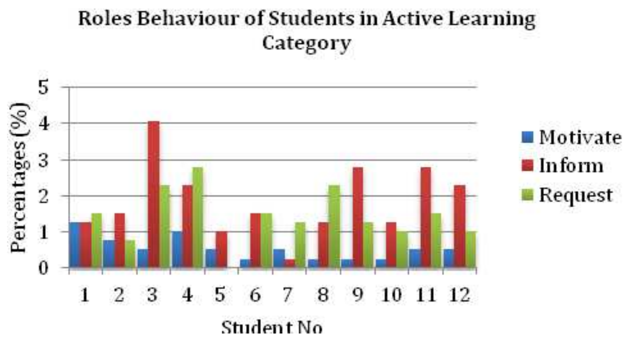

Fig. 3. Roles Behavior of Students in Active Learning Category

The students who played the role of a motivator have a desire to give the positive motivational influence to others in carrying out something. Another possible category in active learning skills is the role of an advisor. A student that can be categorized as an advisor is the student who gives advice, especially somebody who knows a lot about a particular subject or task. The last role behavior of active learning category is questioner. The student that plays the role of a questioner is a student who asks questions, and shows that he or she needs information or that he or she has doubts to clear. From this figure, it can also be seen that all students played three roles in active learning skill, however, the portion of each of them different from each other.

Furthermore, figure 4 illustrates the twelve students who naturally played different roles in conversation skill. Student 1 played a role of a recognizer; accept and approve of comments officially, admitted that something was true or false, and often gave appreciation for peer's comment. She also played a role of a facilitator, focus on the group to the new task or shift the current topic. There is one student, namely student 6 that played the role of a keeper; she maintained the harmony of the discussion by starting a conversation with the words that tend to be polite and thoughtful such as sorry, excuse, pardon, etc.

In the conversation skill, almost all students played all categories except student 9. He did not play the acknowledge category. However, student 1 dominated the discussion by taking two of three categories of conversation skill namely, acknowledge and task. It indicated that this student was more active in conversation by confirming and appreciating other students. In fact, acknowledge category was the most favourite category that was played by students compared to other categories in conversation skill. This indicates that in online discussion, there is exchange of mutual appreciation by way of giving comments or opinions among students.

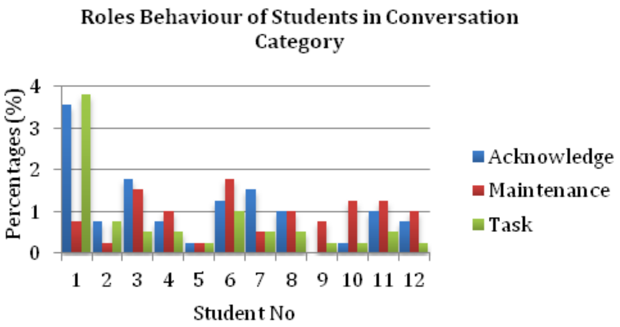

Fig. 4. Roles Behavior of Students in Conversation Category

From table 4, it can be seen that student 1 had three roles in collaborative learning; motivator, recognizer and facilitator. Student 1 was the highest in these three categories that associated with the roles compared to other students in the same group. Role in collaborative learning is gained from comparing individual student with the entire students in the group. For example, student 1 had a role as a motivator, meaning that she played the motivate category more widely than any other student in her group.

Further, the dominant role of an individual student derived from comparing the highest category that was played by the student to other categories. The type of this role depends to the individual itself. For instance, Student 1 had a dominant role as a facilitator, meaning that she played task category more widely than others category.

In the role of collaborative learning, it is possible that a particular student may not have any role while other student may have more than one. All depends on how much the student played that category in relation to the particular role compares to another student. The different situation can be seen as the dominant role of individual student, where every student, definitely, has one role. It is based on which are most category that is played by student compares to other category with the requirement that a student has posted at least once. If the student has only one role in collaborative learning, then this role will automatically become the dominant role of the individual student. Furthermore, if a student has more than one role in collaborative learning, then one of these roles will become the dominant role of individual students. In simple word, the first one refers to the whole of the students and the other one refers to the individual itself.

TABLE IV. ROLES OF STUDENTS IN COMMUNICATION TRANSCRIPTS

Roles of Student in Communication Transcripts

Student's Name	Roles in Collaborative Learning
Student 1	Motivator, Recognizer, Facilitator
Student 2	-
Student 3	Advisor
Student 4	Questioner
Student 5	Mediator
Student 6	Mediator, Keeper
Student 7	-
Student 8	-
Student 9	Mediator, Clarifier
Student 10	-
Student 11	-
Student 12	-

V. CONCLUSION AND FUTURE WORK

In an online discussion, each student may play any roles during online discussion. However, the dominant role that was played by a student would be the characteristic behavior of that student. To assess the student who has the dominant role, hence the role of each student would have to be compared. For instance, everyone may have a motivator role behavior in the online discussion, but who has the strongest motivator sense that indicated by the most played in the motivate category, thus they are a dominant motivator compared to the student who also played the motivate category. This rule applies to all roles that have linked to all students involved in the online discussion. One student may have a more dominant role than another in collaborative learning. It is possible because they can contribute as much as they post their ideas or information in each category. Who have the most postings in one category; hence they considered a dominant role in that category.

ACKNOWLEDGMENT

This work is supported by The Ministry of National Education of the Republic of Indonesia and STMIK-AMIK Riau.

REFERENCES

[1] A. Spatariu, K. Hartley, and L.D. Bendixen, "Defining and measuring quality in on-line discussion," Journal of Interactive Online Learning, Vol 2 (4), 2004.

[2] Z.L. Berge, "Teaching How to Learn Through Online Discussion. In E. Biech (Ed.) Pfeiffer HRD Annual, Volume 1: Training. San Francisco, CA: Jossey-Bass/Pfeiffer, pp. 307-313, 2002.

[3] Erlin, Y. Norazah, A.A. Azizah, "Students' Interactions in Online Asynchronous Discussion Forum : A Social Network Analysis, International Conference on Education Technology and Computer, pp. 25-29, 2009.

[4] K. A. Neuendorf, The content analysis guidebook. Thousand Oaks, CA: Sage Publications, 2002.

[5] T. Anderson, L. Rourke, D.R. Garrison, and W. Archer, "Assessing teaching presence in a computer conference context," Journal of Asynchronous Learning Networks, 2001, from http://www.sloan-org/publications/jaln/v5n2/pdf/v5n2_anderson.pdf

[6] M. De Laat, "Network and content in an online community discourse, 2002, from http://www.uu.nl/uupublish/content/2002%20Networked%20Learning%201.pd

[7] B. De Wever, T. Schellens, M. Valcke, and H. Van Keer, "Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A review," Computers & Education 46 (1), pp. 6–28, 2005.

[8] J.W. Strijbos, R. Martens, J. Prins, and W. Jochems, "Content analysis: What are they talking about?," Journal of Computer & Education, Elsevier. 46, pp. 29-48, 2006.

[9] W. Lam, and K.F. Low, "Automatic Document Classification based on Probabilistic Reasoning: Model and Performance Analysis. Proceedings of the *IEEE International Conference on Systems, Man and Cybernetics*, Vol.3, pp. 2719-2723, 1997.

[10] P. Soucy, and G.W. Mineau, "A Simple k-NN Program for Text Categorization. *The first IEEE International Conference on Data Mining* (ICDM_01), Vol. 28, pp. 647-648, 2001.

[11] L. Ma, J. Shepherd, and Y. Zhang, "Enhancing Text Clasification using Synopses Extraction," *The fourth International Conference on Web Information System Engineering, pp.* 115-124, 2003.

[12] Z. Chen, C. Ni, and Y.L. Murphey, "Neural Network Approaches for Text Document Categorization," International Joint Conference on Neural Networks, Sheraton Vancouver Wall Centre Hotel, pp. 16-21, 2006.

[13] T.Y. Wang, and H.M. Chiang, "Fuzzy Support Vector Machine for Multi-class Text Categorization'" Journal of Information Processing and Management, Elsevier, 43, pp. 914-929, 2007.

[14] Erlin, Rahmiati, and Unang, R., "Two Text Classifiers in Online Discussion: Support Vector Machine vs Back-Propagation Neural Network," Journal of Telkomnika, Vol. 12, No. 1, 2014.

[15] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *10th European Conference on Machine Learning*, pp. 137-142, 1998.

[16] A. Soller, "Computational modeling and analysis of knowledge sharing in collaborative distance learning. Journal of User Modeling and User-Adapted Interaction 14, pp. 351–381, 2004.