Fora for Distance Education: Another Way for Analysis of Discussions

Kiriakos Patriarcheas and Michalis Xenos
School of Sciences and Technology, Computer Science
Hellenic Open University
Patras, Greece
{k.patriac, xenos}@eap.gr

Abstract— **This paper refers to fora for distance education. The present research focuses on content analysis, a technique frequently used to approach issues concerning asynchronous computer mediated discussion groups. There is a variety of approaches, varying on the level of detail and on the type of categories of analysis they use. The content category is presented as a unit of analysis for the decodification of messages in the asynchronous distance education fora in which the modelling is incorporated in a formal language and the development of a respective system created by Hellenic Open University for this purpose. The creation of this system makes an important contribution to the decoding of discussions in fora, and aims at summary identification of discussions which do not develop in the desired way.**

Keywords-E-Learning; Asynchronous distance education fora; Content analysis; Formal language; Modelling; AdaBoost; Naive Bayes; 1-Nearest Neighnor; WINNOW.

I. INTRODUCTION

During the last decade, an increasing number of educational institutions, as well as companies, apply asynchronous educational services via internet [1][2][3]. One of the means utilised in distance education during the last decade is the electronic fora (fora hereinafter). Research efforts on distance education fora, at an international level, began during the ’90s [4][5][6]. However, this is a field, which requires constant updating and redefinition. Given, also, the fact that the practice of distance education during the last decade has acquired new features, both in its methodology and in the tools which are utilised, the further exploration of this field becomes necessary.

The structure of this article is as follows: the theoretical framework section is a short description of the respective assignment on the content analysis technique of asynchronous discussions at distance education fora. The role of the fora of Hellenic Open University (HOU hereinafter) concerning the educational procedure is described. The unit of analysis, which was used, is then presented, followed by the integration of the message content category as a unit of analysis in formal language. A presentation of the system of automatic text classification and the association with the message content category follows, including the description of the results of the experiments performed to control system operation. The paper closes with a discussion on the necessity of this system and the conclusions of this article.

II. THEORETICAL FRAMEWORK

Although researchers seem to agree that collaboration may encourage the learning procedure [1], there is no clear theory available to guide research on computer mediated interaction [2], empirical markers which shall be the base of a codification tool as a standard against which to evaluate whether or not effective learning is occurring through the online discussions [3]. In the last few years, numerous efforts to approach this issue were made, stemming from different theoretical backgrounds. Indicatively, Henri [4] uses the point of Cognitive and metacognitive knowledge, while others [5][6] the point of Critical thinking.

As is shown from all the above, an important issue arising is the unit of analysis which shall be used for the content analysis. Fahy et al. [7] consider each single sentence as one unit of analysis, and Pena-Shaff and Nicholls [8] uses the sentence as unit of analysis, trying to approach it at a paragraph level. Others choose the definition thematic unit (or otherwise of a “theme” or an “idea”) to be their unit of analysis [4][5][9]. Another approach [3][6][7][10] is to consider the whole message that a student enters at a specific moment in the conversation as the unit of analysis. Jarvela and Hakkinen [11] choose a Complete discussion, while during the last years there has been an approach of multiple point both at a micro and at a macro level [12]. Further down, a comprehensive review is presented in a table form (Table I), referring to the unit of analysis used by this field’s researchers.

TABLE I. OVERVIEW OF THE CONTENT ANALYSIS SCHEMES

Instrument	Theoretical background	Unit of analysis
Henri (1992)	Cognitive and metacognitive knowledge	Thematic unit
Newman et al. (1995)	Critical thinking	Thematic unit
Zhu (1996)	Theories of cognitive and constructive learning – knowledge construction	Message
Gunawardena et al. (1997)	Social constructivism – knowledge construction	Message
Bullen (1997)	Critical thinking	Message
Fahy et al. (2000)	Social network theory – Interactional exchange patterns	Sentence
Veerman & Veldhuis-Diermanse (2001)	Social constructivism – knowledge construction	Message
Jarvela & Hakkinen (2002)	Social constructivism – perspective taking	Complete discussion
Lockhorst et al. (2003)	Social constructivism – learning strategies	Thematic unit
Pena-Shaff & Nicholls (2004)	Social constructivism – knowledge construction	Paragraph
Weinberger & Fischer (2006)	Social constructivism –knowledge construction	Micro and macro-level

 ISBN: 978-1-61208-228-8

III. THE CASE OF HELLENIC OPEN UNIVERSITY

HOU is the eminent educational institution offering distance education in Greece. Today, HOU has 30,557 students (17,889 undergraduate, 12,600 postgraduate and 68 PhD candidates); it is staffed by 1642 professors (only 42 of which are permanent and the rest are associate professors-counsellors).

The HOU's structural educational unit is the course module; presently, 203 course modules are offered by HOU. An important supportive mean of the educational procedure is the fora of HOU, which contribute both to the organization of the studies during the course module as well as to the elaboration and development of what the student have already studied.

The HOU's fora offer important help during the educational procedure. They may also contribute to the following:

a) as for the organization of the studies during the course module:

- to the communication between the tutor and the students (regularity of contacts, subject, resolution of "technical" problems etc.).
- to the organization of homework (method of use of the teaching material and the preparation of the activities, exploitation of the literature and the other sources, timetables, encountering problems related to it etc.)
- to the supply of information about the advisory meetings (their number, their duration, the timetables, the goals, their content and methodology applied, problems' encountering as for the ability to attend them etc.).
- to supply clarifications about the procedure of preparation and evaluation of the written assignments (form, method of preparation, evaluation criteria, ways to be supported by the tutor etc.).
- to inform about the procedure of final exams.

b) as for the elaboration and development of what the students have already studied, the HOU's fora may be exploited for:

- presentation of consolidation exercises, short suggestions, presentation of examples, methodologies, literature etc.,
- resolution of questions and the supply of clarifications about the teaching material.
- interconnection between what is already studied and the next chapters and the following written assignment.

In the discussion threads of each course module, the tutor and all the students of the course module have the chance to participate. As for the students of informatics, for the 16 course modules of informatics (for undergraduate level) offered by HOU, by the time this research was conducted, there were 753 discussion threads created with 6,663 messages. Concerning the evolution of the HOU forum's use, indicatively, at the course module "Introduction to Informatics" (INF10), during the last three academic years there is a great increase in the number of messages: 1808(2009-10), 1942 (2010-11) noted and 2913 (2011-12).

Given the big flow of information transferred through fora of HOU, simulated the development of a formal language to interpret messages in the fora of HOU, a system based on modelling with the use of a formal language was created, entering threads from discussion fora and exporting the respective strings in an automatic way.

IV. THE UNIT OF ANALYSIS

Given that the choice of a unit of analysis is dependent on the context and should be well-considered, because changes to the size of this unit will affect coding decisions and comparability of outcome between different models [14], as well as given the fact that Schrire [15] refers to a dynamic approach in which data is coded more than once and the grain size of the unit of analysis is set, depending on the purpose and the research question, it was decided not to take into consideration the discussion thread, not even the message as unit of analysis, nor the paragraph or the single sentence.

It was decided to use as unit of analysis, the category of the message's content, as for the observation of the discussion threads, it was noticed that there are cases of messages which may comprise two (or/and more) content categories, e.g., a question about the next advisory meeting and a reply to a question concerning the study of the educational material.

Thus, in that case, the analysis at a message's level used by some researchers [3][6][10][13] is insufficient for the exploitation of information that shall arise aiming to reach educational conclusions, as it is obvious that in a message more content categories may coexist.

Furthermore, the analysis at a level of a single sentence used by some researchers [7] could not be taken as a single unit of analysis since a content category may extent to two or/ and more sentences. For the same reason the unit of analysis at a paragraph level, used by Pena-Shaff and Nicholls [8], was not chosen. In addition, in our case, not even the Complete discussion used by Jarvela and Hakkinen [11] as unit of analysis cannot be exploited, since it is noted that in a discussion thread there may exist many more than one content categories. Otherwise, the title of a discussion thread may not be representative of this and the discussion may extent to more than one subjects.

According to the study of the messages of INF10 for academic years 2009-2012 the messages as for their content may concern (in brackets you see the respective symbols used in formal Language): study of educational material (*M*), questions/answers for exercises – assignments (*X),* presentation of sample assignments by tutors (*P*), instructions (*I*), assignment comments, corrections (F), student comments on assignments (*D*), sending – receiving assignments (*J*), sending - receiving grade marks (*G*), notification of advisory meeting (*V*), and pointless message (*L*).

 ISBN: 978-1-61208-228-8

V. INTEGRATION OF THE MESSAGE CONTEXT CATEGORY AS A UNIT OF ANALYSIS IN FORMAL LANGUAGE

According to the study of the messages on HOU's fora during the academic years, a Language was developed, which is defined by mathematic terms and represents the messages using as unit of analysis the category of the message content. More specifically:

There are two categories of communication's carriers: a) Tutors, b) Students

For brevity reasons, tutors shall be symbolized with *T* and students with *E*

As for the type of message, they are discerned to questions and replies (symbols *q* and *a* respectively).

As for their content category, we have the symbols aforementioned in the previous section: *M, X, P, I, F, D, J, G, V*.

Finally, the order in which the above symbols appear is: a) the message carrier, b) the type of message and c) the content category to which the message belongs.

Thus, the Language contains:

a) Terminal symbols alphabet V_T, where V_T = {T, E, q, a, n, M, X, P, I, F, D, J, G, V, L }

b) Non terminals alphabet V_N, where V_N = {u, r, y, c}, more specifically :

r: represents the message carrier (where *T* for tutors and *E* for students)

u: represents a pair *yc* i.e., a message type *y* (whether it is a question *q* or an answer *a*) followed by its content category.

c) The grammar P

A set of rules of the form α → β, where α and β sequences containing terminal and non-terminal symbols and α is not an empty sequence, as follows:

1. S → ruS	8. y → q	15. c → F
2. S → ε	9. y → a	16. c → D
3. u → uyc	10. y → ε	17. c → J
4. u → ε	11. c → M	18. c → G
5. r → T	12. c → X	19. c → V
6. r → E	13. c → P	20. c → L
7. r → ε	14. c → I	21. c → ε

where ε stands for an empty symbol

d) Symbol S where every sentence generated starts with this symbol.

According to the above, when a message should be represented concerning a student's message, addressing a question about the study of the educational material, followed by another student's question about the following assignment and at the end of the thread there is the reply of the tutor both for the study of the material and for the following assignment, it shall be represented as follows: *EqMEqXTaMX* (*E* for the student's capacity, *q* for the question, *M* as it concerns the study of the educational material, *X* for the fact that the next message concerned an assignment, *T* for the tutor's capacity, *a* for the fact that it is an answer, *M* for the fact that this reply concerns the study of educational material and *X* for the fact that the second part of the message concerns an assignment.). According to the above, the sequence *EqMEqXTaMX* constitutes a sentence of the *Language* because:

Rule: (1) (1) (1) (3)
S —> ruS —> ruruS —> rururuS —> ruycruycruycS

(4)(6)(8)(11) (4)(6)(8)(11)
—————> EqMruycruycS —————> EqMEqXruycS

(3) (2)(4)(5)(9)(10)(12)
—>EqMEqXruycycS ——————————> EqMEqXTaMX

As it is obvious from the example, while to the first two messages corresponds one content category *M* and *X* respectively, at the 3rd message there are two content categories *MX*.

VI. THE SYSTEM - TIME ASSOCIATION

According to this approach, a system of automatic classification was developed, which comprised the following:

a) Data filtering: where some web pages are considered as input accommodating the discussion threads of a distance education forum of HOU (which include a great deal of irrelevant data containing essential information concerning the educational procedure, e.g., titles, images etc.) and creates a temporary file with the "useful" part (User name, date, message's content), which may become a source of information for educational conclusions.

b) Storage of root files: which is a dynamic way according to which word or phrases or symbols roots are stored, as well as the respective terminal symbols q if it is a question or a if it is an answer. The same was also done for the storage of information necessary in the determination of content category of a message, i.e., if it is about study, assignment, comment etc. or combination of them (e.g., a message concerning both the study and an assignment). To wit, it takes as input couples of information of the type root of a word or phrase and terminal symbol of the content category (M, X, P, I, F, D, J, G, V, L). As it is obvious, the system provides the ability to add further content categories if necessary.

c) Strings' production: receiving as input the temporary file with the "useful" information (User name, date, message's content) and the files with the couples of roots words/ phrases/ symbols and terminal symbols and presents the respective strings with the relative extensible file, so as the results to be kept for further exploitation (Figure 1).

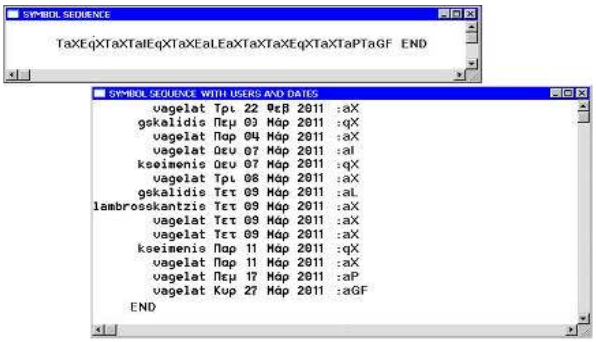

Figure 1. Representation of a discussion thread both in simple string and also after the addition of User names and dates (Days: Κυρ=Sun, Δευ=Mon, Τρι=Tue, Τετ=Wed, Πεμ=Thu, Παρ=Fri, Σαβ=Sat, Months:Φεβ=Feb, Μάρ=Mar, Απρ=Apr).

 ISBN: 978-1-61208-228-8

It is worthy to note here that this specific system incorporates the sense of time along with its association with each of the nine (9) categories of message content (except from the pointless message category) chosen as unit of analysis. More specifically and given that within a message (as it is deduced both from literary review and from the observation of the fora of HOU) more than one contents may exist, the dates are recorded for each such case and not simply in each message.

In fact, after each couple yc, there is a date's record. Certainly, so as to effectuate the above procedure nine (9) stacks were used – as many at the message's content categories, each one having as many figures as the number of appearance of terminal symbols resulting from the non-terminal symbol c. Consequently, time differences may automatically exist (in days, if from each current date, by content category, it is deduced the previous one) and thus there may arise another nine (9) respective stacks with the above date references. Of course, the length of these stacks is equal to the length of dates minus one (-1), i.e., apart from the initial message, which is considered to be the point zero (0), where the numbering of the time differences begins. The contents of the stacks of time differences may constitute an important criterion, which may participate as such (in combination with other criteria) in case of evaluation of a forum's consequences to the educational process.

Also, the final form that will have the 9 stacks with the dates they will be as follows (Table II):

TABLE II. DATE'S STACKS

null										
3/20/12	4/1/12	4/2/12	4/5/12	4/6/12	4/7/12	4/7/12	4/7/12	4/9/12	4/9/12	null
4/15/12	null									
4/5/12	null									
4/25/12	null									
null										
null										
4/25/12	4/28/12	null								
null										

while the counter table (Table III) has the following contents:

TABLE III. COUNTER TABLE CONTENTS

1	2	3	4	5	6	7	8	9
0	10	1	1	1	0	0	2	0

As for the time differences stacks (Table IV) has the following form:

TABLE IV. TIME DIFFERENCES STACKS

Content	Appearances	1	2	3	4	5	6	7	8	9
M	*0*									
X	*10*	9	1	3	1	1	0	0	2	0
P	*1*	0								
I	*1*	0								
F	*1*	0								
D	*0*									
J	*0*									
G	*2*	3								
V	*0*									

VII. EXPERIMENTS

A. Phase A

At first, experiments were carried out by using 80 discussion threads of the INF10 module of the academic year 2010-11. Given that 219 threads with 1,942 messages had been created throughout the year, there was the ratio of approximately 9 messages (in particular 8.87) per discussion thread. Therefore, out of the 80 selected threads, an effort was made to use those containing 8 or 9 messages for the purpose of experimental control. Thus, we finally chose 80 discussion threads with 712 messages in total (average 8.90 messages/thread).

At the first experimental operation, the word root files in relation to both the type (question/answer) and (mainly) the content category of message contained 18 and 92 entries respectively. Under these conditions (Table VI), we ended up having 58 discussion threads with no errors and 16 threads with only one wrong symbol (compared to what was expected). Namely, out of (approximately) 9 messages (of each of the 16 threads), 8 of them were correct and one message was wrong because it did not contain not even one of the 92 provided word roots. Respectively, there were 5 threads with two errors and 1 thread with more errors (this thread was created before Christmas holidays and its messages contained mainly wishes). We should note here that there has been no error regarding the type of messages (question/answer), only in terms of determining the content category.

Following the observation/study of messages in the 21 threads that contained 1 or 2 errors, 49 additional word roots (concerning the content category) were recorded and it was decided to enter them in the root file. The experimental operation performed in the same 80 threads had, clearly, better results, with total success in 70 threads, one wrong symbol in 8 threads, two errors in 1 thread, and 1 thread that did not actually refer to educational content (Table V).

At this point, it should be clarified that the control of the results produced by the system in this phase (A), was conducted with manual comparison of all the messages in the discussion threads that were used in order to control system reliability at the first degree.

TABLE V. EXPERIMENTAL OPERATION - PHASE A'

2010-11	**1st Exp. operation**	**2nd Exp. operation**
Threads	80	80
Messages	712	712
Messages/Thread	8.9	8.9
Full success (threads with no errors)	58	70
Threads with one error	16	8
Threads with two errors	5	1
Threads with more errors	1	1
Correct messages interpretation	677	693
Wrong messages interpretation	35	19

B. Phase B

Given that the 8 discussion threads with one error were found not to have any common word root feature that would adequately correspond, we decided to initiate the second

 ISBN: 978-1-61208-228-8

experimental phase (B'). Classification was performed according to international literature [16-21], using the algorithms indicated for this purpose: Naive Bayes (NB), 1-Nearest Neighnor (1-NN), WINNOW and discrete AdaBoost (in the form generalized by Nock and Nielsen [19] based on Freund and Schapire [20]).

During this phase every algorithm was formed using the data collected from the academic year 2010-11. Subsequently, a group of data for two other academic years (2009-10 and 2011-12) was also collected. The results show that the discrete AdaBoost algorithm produced the greatest accuracy. This result complies with Bloehdorn and Hotho [21] who used the discrete AdaBoost algorithm in a similar experiment. The accuracy is denoted in the Table VI.

TABLE VI. ACCURACY OF ALGORITHMS FOR THE ACADEMIC YEARS 2009-12

	2009-10	2010-11	2011-12	Average Accuracy (1)	Average accuracy (2)
In thread level					
AdaBoost	75.11	80.08	87.21	80.64	80.80
Naive Bayes	72.47 *	77.83	86.18	78.66	78.82
1-Nearest Neighnor	73.45	76.66	83.65*	77.77	77.92
WINNOW	70.13 *	73.24*	83.10*	75.34	75.49
In message level					
AdaBoost	92.36	95.19	97.89	94.96	95.15
Naive Bayes	89.11*	92.51	96.73	92.59	92.78
1-Nearest Neighnor	90.31	91.13	93.89*	91.60	91.78
WINNOW	86.23*	87.06*	93.27*	88.67	88.85

The star (*) indicates that the algorithm which were used ,performed statistically better than the specific classifier according to t-test with $p<0.05$. In all the other cases, there is no significant statistical difference between the results (Draws).

The average accuracy (1) corresponds to the total number of threads and messages, while in (2) the years have an equal participation (1/3) in the total average.

It is worth noting that in all cases, the type of message has been correctly identified (i.e., whether it is a question or answer), and therefore any errors concerned the content category (as noticed in the initial experimental operation). Results are shown in Table VII.

TABLE VII. RESULTS FOR THE ACADEMIC YEARS 2009-12

Year	2009-10	2010-11	2011-12	Total
Threads	356	219	178	753
Messages	1808	1942	2913	6663
Messages/Thread	5.08	8.87	16.36	8.85
Threads with no	269	191	159	707
Percentage	75.56%	87.21%	89.33%	80.64%
Correct messages	1678	1901	2864	6443
Percentage	92.81%	97.89%	98.32%	96.70%

The above results, which followed a calibration process of repeated readjustment, were deemed satisfactory (98.32% correct message interpretation for 2011-12) and in the end, the development of this system gives a clear affirmative answer to the question "is there an automated method to interpret messages in a distance education forum?" Therefore, by using this system, it is now possible to read study and classify, within a few minutes, a large number of messages (6,663 messages) which took 12 months to be completed within the framework of this paper.

VIII. DISCUSSION

At this point, it is important to initially discuss the need for such a system using the content category as analysis unit. As it is deduced from the related literature review, it is concluded a gap as for the methodological approach which is to be based on modelling with the use of formal language and which will examine the content analysis from the message content category view. This research aims to contribute to the covering of this gap, consequently a need arose to create a system to interpret the discussion of a forum about distance education in a structured way, through the approach presented above and at the same time taking into consideration the concept of time and producing respective results, so as to help the participating students to improve their educational practices.

In fact, this system defines a "code" clarifying some issues which determine both the quality of the communication relations and the educational principles of teaching/ learning procedure. In other words, this system aims to encourage towards the direction of the use of "good" or "desirable" educational techniques, adding up to the distance education. At this point it should be clarified that the development of this system does not intend to disorientate from the basic principles of distance education, but aims to contribute to its further development and upgrading and to act auxiliary and not in excess. The results of its use shall constitute data for the creation of a database [22] aiming to investigate the effects of fora in educational procedure from the point of causal interpretation point of view. Given that the HOU is not a conventional university (with the features of a homogenous student community), but it addresses to adults with special educational needs and incongruity (both as far as their age, their professional and family obligations are concerned), the future research access to such issues becomes particularly important.

More specifically, the system in the future shall collect the students' particulars (e.g., marital status, age, sex, profession etc.), their performance at course modules of HOU (final mark, assignments' marks, effort of success of the course module) and the strings produces by the system so as to interpret the messages of HOU's fora, in order to reach educational conclusions in combination with the use of the a tool, weka type. In fact, this application refers to a wider field of interdisciplinary encounter, by the merge of cognitive theories and artificial intelligence.

IX. CONCLUSIONS AND FUTURE WORK

The practice of distance education during the last years has acquired new features, both in relation to methodology and in the tools it uses. It is also a fact that the subject of electronic fora in distance education is a dynamically formed field requiring constant updating and redefinition. A big part of the research presented in the international literature concerning distance education's fora, refer to the content analysis, which principally aims despite the fact that this research technique is frequently used, though there are

 ISBN: 978-1-61208-228-8

still no standards established. There is a variety of approaches, varying both at detail's level and at the type of categories of analysis they use. As it was deduced from the above presentation and study of the discussion threads of HOU, it was noticed that there are cases of messages, which may comprise two (or/and more) content categories, e.g., a question about the next advisory meeting and a reply to a question concerning the study of the educational material. For this reason, this paper uses the content category as unit of analysis for the messages' interpretation in Asynchronous distance education fora and for this purpose incorporates it in a modelling in a formal language. Furthermore, time indexes of participation were integrated in combination with the content categories of the message, in order to define the way these elements could improve the capacity of the tutor to evaluate the progress of a discussion thread in a distance education forum.

Among others, the prediction for future research actions are long-term studies concerning the main issue: what reinforces the participation at fora and how this contributes to the educational process effectiveness by investigating side questions, such as how much it affects the person who starts the thread (tutor or student), how it starts, the period when the thread starts, how important the time of response in threads, is the groups' size etc. and their association with the elements concerning the students' profiles and their performance in course modules of HOU, intending to reach educational conclusions.

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees for their careful review and useful suggestions.

REFERENCES

[1] A.W. Lazonder, P., Wilhelm, and S.A.W. Ootes, "Using sentence openers to foster student interaction in computer-mediated learning environments" Computers & Education, vol. 41, pp. 291–308, 2003.

[2] G. Stahl, "Building collaborative knowing: contributions to a social theory of CSCL" In What we know about CSCL in higher education (Strijbos, J.W., Kirschner, P., & Martens, R.L. Eds.), Kluwer, Amsterdam, 2004.

[3] C.N. Gunawardena, K. Carabajal, and C.A. Lowe, "Critical analysis of models and methods used to evaluate online learning networks" In American Educational Research Association Annual Meeting, American Educational Research Association, Seattle, 2001.

[4] F. Henri, "Computer conferencing and content analysis". In *Collaborative learning through computer conferencing* (KAYE AR Ed), pp. 117–136, The Najadan Papers, Springer-Verlag, London, 1992.

[5] D.R. Newman,, B. Webb, and C. Cochrane, "A content analysis method to measure critical thinking in face-toface and computer supported group learning" Interpersonal Computing and Technology, vol. 3, pp. 56–77, 1995.

[6] M. Bullen, "A case study of participation and critical thinking in a university-level course delivered by computer conferencing", University of British Columbia, Vancouver, Canada, 1997.

[7] P. Fahy, G. Crawford, and M. Ally, "Patterns of interaction in a computer conference transcript" International Review of Research in Open and Distance Learning, vol. 2, pp. 1-24, 2001.

[8] J.B. Pena-Shaff,, and C. Nicholls, "Analyzing student interactions and meaning construction in computer bulletin board discussions", Computers & Education,vol. 42, pp. 243–265, 2004.

[9] D. Lockhorst, W. Admiraal,, A., Pilot, and W. Veen, "Analysis of electronic communication using 5 different perspectives" Paper presented in a symposium conducted at the 30th Onderwijs Research Dagen (ORD), Kerkrade, The Netherlands, 2003.

[10] E. Zhu, "Meaning negotiation, knowledge construction, and mentoring in a distance learning course" In proceedings of selected research and development presentations at the 1996 national convention of the association for educational communications and technology. Available from ERIC documents: ED 397 849, Indeanapolis, 1996.

[11] S. Jarvela, and P. Hakkinen, "Web-based cases in teaching and learning: The quality of discussions and a stage of perspective taking in asynchronous communication" Interactive Learning Environments, vol. 10, pp. 1–22, 2002.

[12] A., Weinberger, and F. Fischer, "A framework to analyze argumentative knowledge construction in computer-supported collaborative learning" Computers & Education, vol. 46, pp. 71-95, 2006.

[13] A. Veerman, and E. Veldhuis-Diermanse, Collaborative learning through computer-mediated communication in academic education. In Proceedings of the Euro CSCL 2001, pp. 625–632, McLuhan institute, Maastricht, 2001.

[14] D. Cook, and J. Ralston, "Sharpening the focus: methodological issues in analysing online conferences" Technology. Pedagogy and Education, vol. 12, pp. 361–376, 2003.

[15] S. Schrire, "Knowledge-building in asynchronous discussion groups: Going beyond quantitative analysis", Computers & Education, vol. 46, pp. 49–70 , 2006.

[16] M. Kongovi, J. C. Guzman, and V. Dasigi, "Text Categorization: An Experiment Using Phrases", Lecture Notes In Computer Science, Vol. 2291, pp. 213-228, Springer-Verlag, Berlin, 2002.

[17] F. Sebastiani, "Machine Learning in Automated Text Categorization", *ACM Computing Surveys*, vol. 34, pp. 1-47, 2002.

[18] F. Sebastiani, "*Classification of Text, Automatic. Encyclopedia of Language & Linguistics*", 2nd Edition, Section: Applications of natural language processing, Vol. 14. Elsevier Science Publishers, 2006.

[19] R. Nock and F. Nielsen, "A Real Generalization of discrete AdaBoost", *Artificial Intelligence*, vol. 171, pp. 25-41, 2007.

[20] Y. Freund and R. E. Schapire, " A Decision-Theoretic generalization of on-line learning and an application to Boosting", *Journal of Computerand System Sciences,* vol. 55, pp. 119–139, 1997.

[21] S. Bloehdorn and A. Hotho, "Boosting for Text Classification with Semantic Features", Advances in Web Mining and Web Usage Analysis. Lecture Notes in Computer Science, Vol. 3932, pp. 149-146, Springer-Verlag. 2006.

[22] S. Kotsiantis, K. Patriarcheas, and M. Xenos, "A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education" Knowledge Based Systems, vol. 23, pp. 529-535, 2010.

 ISBN: 978-1-61208-228-8