No Need to Read Messages Right Now: Helping Mediators to Steer Educational Forums Using Statistical and Visual Information

Marco Aurélio Gerosa, Mariano Gomes Pimentel, Hugo Fuks & Carlos J.P. Lucena

Computer Science Department, Catholic University of Rio de Janeiro (PUC-Rio)
R. M. S. Vicente, 225, Rio de Janeiro, Brazil - 22453-900
{gerosa, mariano, hugo, lucena}@inf.puc-rio.br

Abstract. In an education environment, a forum provides a valuable tool that can be used to foster reflection and a deeper analysis of subjects being discussed. However, as an asynchronous communication tool, participation can occur at any time, demanding a constant attention of the teacher to satisfactorily mediate the group and the discussion. In addition, a reasonable number of messages can be posted in a short period of time, making it hard to follow up and coordinate the discussion. This paper proposes an approach based on statistical and visual analysis of messages characteristics to alert the teacher about potential problems.

Keywords: Forum, Coordination, Message Inspection, Collaboration

INTRODUCTION

Working in groups brings some characteristics, such as synergy, ability to consider more information, objective evaluation, cognitive stimulation, contribution of different understanding and exposure to alternative points of view, that can enhance learning [Benbunan-Fich & Hiltz, 1999][Hiltz, 1994]. Group members can monitor individual thinking and group structure provides social support and encouragement for individual effort. By formulating ideas in their words and receiving evaluation from peers, students' knowledge, thinking skills and meanings are socially constructed [Harasim et al, 1997]. Moreover, working with peers tends to reduce anxiety as learners share solutions to complex tasks, increasing satisfaction with the process and results.

In asynchronous collaboration events, learners can participate at a time and a place convenient to themselves and appropriate to the task, having more time to reflect before composing their contributions. In addition, besides the fact that extrovert personalities continue to contribute more than quieter members do, they cannot dominate completely as in synchronous or face-to-face situations. Quieter members still have the opportunity to contribute [Straus, 1996]. Others learners that may feel that they do not have enough time to prepare their thoughts in a fast-paced discussion also have in the asynchronous events a better opportunity to expose their ideas and to contribute to the discussion in a slower-paced, less time-limited way [Funaro & Montell, 1999]. The asynchronous format also allows students to work through difficult texts and concepts more slowly, and to help each other to understand the contents [Bull et al, 2001].

A forum is an asynchronous textual communication tool that can be used to create threaded discussions, where the relations between a message and the one that it is responding to is visually characterized. As students' responses are posted publicly and become persistent in the environment, they tend to contribute more thoughtfully, working their arguments and backing up their ideas with evidence before turning them into essays [LaGrandeur, 1996]. They can also share their thinking with each other, comment on each other's ideas and find partners that share their interests in order to get in a deeper discussion [Funaro & Montel, 1999]. In this networked learning, students should be aware that they are much more responsible for their success, looking for their own sources of information and learning how to deal with information overload and how to convert collaboratively information into knowledge, turning them into learners who actively generate knowledge rather than being passive receivers [Harasim et al, 1997].

However, in order to be successful, a forum demands a close attention of the mediator, mainly in the initial sessions when participants are not used to the adopted dynamic yet. The mediator must mediate the forum in order to guarantee that learners participate properly, that the discussion does not drift to a nonproductive direction and that the flow of information is neither too monotonous nor unmanageable [Salmon, 2000]. In addition, by reducing the pressure to respond, it is easier for a student to drop out of the group [Graham et al, 1999]. The mediator has to demand regular contributions in an appropriate timeframe to avoid dispersion.

Most environments do not provide a desirable specific computational support, leaving to the mediator all the effort to collect and analyze the information necessary to mediate group discussion. As the contributions may occur at any time and at any rate, it gets really hard to the mediator to constantly monitor, read and follow up the discussion in order to intervene before noise propagation.

This paper proposes the use of a statistical and visual analysis that can be offered by the environment, allowing the mediator to timely intervene in the discussion, by saving him from the immediate need of inspecting each message. In order to mediate the discussion, the mediator must communicate, coordinate and cooperate [Fuks, Gerosa & Lucena, 2002]. This paper focuses on the coordination aspect of the collaboration. Coordination is the additional effort needed to organize the group in a manner that channels communication and cooperation towards the group's objective [Raposo & Fuks, 2002]. When coordinating a forum group discussion, the mediator must ensure that each learner participates, that the contributions add value to the discussion and that the conversation does not drift to nonproductive direction.

In this paper, this analysis is restricted to a set of message characteristics that can be extracted without the need of human inspection, namely message chaining, category, size and date. Measurements such as the average depth level of the discussion tree indicates the depth of the discussion that is taking place, and the percentage of leafs do the same for the level of interaction. Message categorization provides a way to focus the analysis on the specific message type in need of attention. Message size helps to identify messages that are not consistent with others of the same type. By analyzing message timestamp, it is possible to identify the time range between messages and their rate. Cross referencing this data also reveals other information, such as the type of message expected per level, how fast the tree grows, which types of messages are answered more quickly etc.

For the case study, data were extracted from a regular online course of the Computer Science Department of the Catholic University of Rio de Janeiro. This course, which currently is in its 14th edition, runs in the AulaNet learning management system. The course and the system are briefly presented in the next section.

THE COLLABORATION SCENARIO

In order to collaborate, members of a group should communicate, coordinate themselves and cooperate. While communicating, they negotiate and make decisions. While coordinating, they deal with conflicts and organize the group in a manner that avoids the loss of communication and cooperation efforts. While cooperating, they work together in a shared space, seeking to complete tasks, generating and manipulating cooperation objects. Renegotiation and decision making regarding unexpected situations turn up while cooperating, thus, demanding new rounds of communication, which on its turn, will require more coordination to reorganize the tasks to be further executed during the next round of cooperation. This cycle shows the iterative nature of collaboration. Participants interact, obtaining feedback from their actions and feedthrough from the actions of their companions by means of awareness elements. The diagram shown in Figure 1 summarizes this cycle. It is based on models found in the literature, such as the 3C model proposed by Ellis et al. [1991] and the Clover design model [Laurillau & Nigay, 2002]. This paper focus is on the coordination aspect of collaboration.

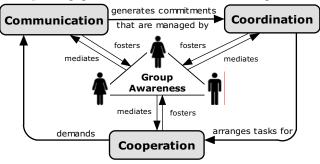


Figure 1. Overview of the 3C collaboration model.

The AulaNet learning management system provides an environment for teaching and learning via Web and was developed based on the 3C collaboration model. It has been under development since June 1997 by the Software Engineering Lab of the Catholic University of Rio de Janeiro (PUC-Rio). The AulaNet is a freeware and it is available in Portuguese, English and Spanish versions¹. During 2002, 2003 and 2004, more than 6000 university students and 20000 workers in their companies have been using AulaNet.

The Information Technology Applied to Education (ITAE) course has been taught entirely at a distance using the AulaNet environment since 1998. Its objective is to get learners to collaborate using information technology, becoming Web-based educators [Fuks, Gerosa & Lucena, 2002]. The course seeks to build a

¹ AulaNet: http://groupware.les.inf.puc-rio.br or http://www.eduweb.com.br

learning network where the group learns, mainly, through the interaction of its participants in collaborative learning activities.

Some points of the course's dynamics need to be characterized for the sake of the case study. The course is organized by topics, with one topic being discussed per week. Learners read selected content regarding the weekly topic, conduct research to enhance their understanding about it and, then, take part in an asynchronous discussion about three specific questions regarding that topic. This discussion is carried out over 50 hours using the AulaNet Conferences service, which implements a forum. To close the weekly discussion, learners revisit the same questions using a chat tool during one hour.

In the ITAE course, the role of transmitting information and leading the discussion, which generally is an attribute of course mediators, is shared with learners. A learner is selected in each conference to play the role of the *seminar leader*, being responsible for preparing an initial message followed by three aforementioned questions, referred by group members to develop their argumentation. During this phase, the seminar leader is also responsible for keeping the discussion going on and maintaining the conference's dynamics.

In AulaNet, the learner can select a category for the message from a set that have been previously defined by the course teacher [Gerosa, Fuks & Lucena, 2001]. The available categories in the ITAE course, used to identify the message type, are *Seminar*, *Question*, *Argumentation*, *Counter-Argumentation* and *Clarification*. In the ITAE course, the seminar leader posts a message from the *Seminar* category to serve as the root of the discussion, as well as three messages from the *Question* category. During the following 50 hours, all learners engage in the discussion.

Each Conference message is graded and commented upon individually by the mediators in order to provide guidance to learners on how to prepare their texts, avoiding the sending in of contributions that do not add value to the group. Contributions are commented upon in the message itself, generally in a form that is visible to all participants, so that the learners better understand where they can improve and what they have gotten right.

ANALYSIS OF MESSAGE CHARACTERISTICS

Analyses about message chaining, categorization, size and date are presented in this section, illustrating how these characteristics can help in the coordination of educational forums. The data and examples were collected from seven editions of the ITAE course (from the second semester of 2001 to the second semester of 2004).

Message Chaining

Communication tools have different ways of structuring messages: linear (list), hierarchical (tree) or network (graph), as can be seen in Figure 2. Despite the fact that a list is a specific case of a tree, and this is a particular type of graph, no one structure is better than another. Linear structuring is appropriate for communication in which the chronological order of the messages, such as the sending of notices, reports and news, is important. Hierarchical structuring on a forum, on the other hand, is appropriate when the relationships between messages, such as questions and answers, need to be quickly identified. However, it is relevant to point out that, since there is no way to link messages from two different branches, the tree can only grow wide and, thus, the discussion takes place in diverging lines [Stahl, 2001]. Network structuring can be used to seek convergence of the discussion.

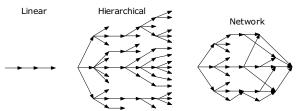


Figure 2. Examples of discussion structure

In the ITAE course, the forum, based on the AulaNet Conferences service, is used for the in-depth discussion of the course's subject matter. The format of the resulting tree indicates the depth of the discussion and the level of interaction [Pimentel, Fuks & Lucena, 2003]. For example, a tree like the one exemplified in Figure 3 has only three levels, which indicates that there was almost no interaction, given that level zero is the seminar message, level one comprises the three questions and level two comprises the answers to the questions. This means that learners only answered the initial questions.



Figure 3. Example of a corresponding tree

The trees extracted from the conferences of the five editions of the ITAE course are shown in Figure 4. It can be also noted that in some editions the depth of the tree become shallower over the semester while in others it becomes deeper. It can be seen that in 2003.1 and 2003.2 the trees gets deeper, indicating that learners improved their argumentation skills during those semesters. On the other hand, in 2002.1 the trees got shallower over the semester, most probably due to lack of coordination. It is also possible to observe that normally the first conference corresponding tree is the shallowest one. It is possible for example to visually compare the depth of the conferences of a given edition with those of other editions. However, in order to conduct a more precise analysis, it is also necessary to have statistical information about these trees, as presented in Figure 5.

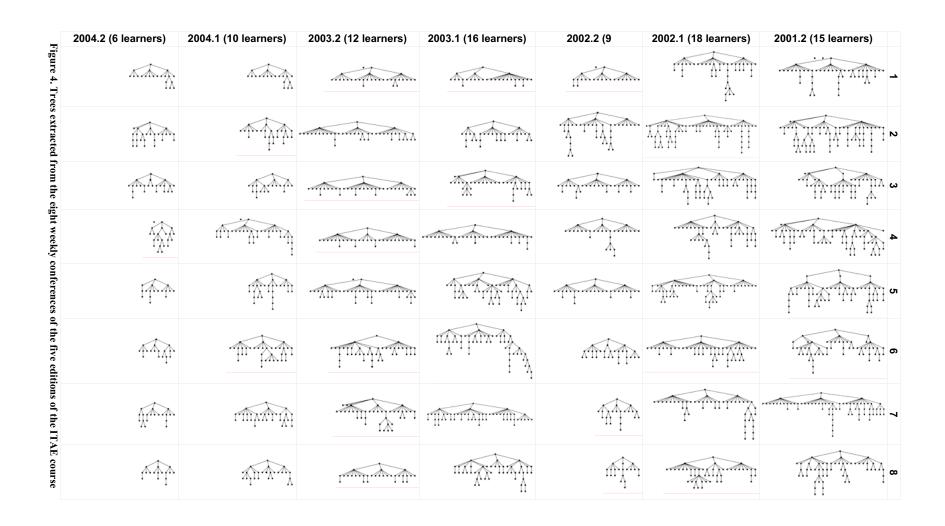
The trees shown in Figure 4 and the charts in Figure 5 indicate that the interaction on the 2002.1 edition declined over the course of the conferences, while the interaction on ITAE 2003.1 edition increased. It can be seen in Figure 5 that the average depth of the trees in the 2002.1 edition declined while the percentage of messages without answers (leafs) increased, which indicates that learners were interacting less as the course progressed. In the first four conferences of this edition, the average level of the tree was 3.0 and the percentage of messages without answers was 51%; in the last four conferences, the average tree level was 2.8 and the leafs were 61%. On the other hand, in the 2003.1 edition, learners' interaction increased over the course: the trees were getting deeper while the leafs were decreasing. The average depth level was 2.2 in the first four conferences, increasing to 3.0 in the last four, while the percentage of leafs went from 69% in the first four conferences to 53% in the last four. While the 2002.1 edition learners slowly decreased their interaction level, the 2003.1 edition learners quickly increased their interaction level. Another indication of the declining interaction level of the former is the continuous decline in the number of messages.

All this information was obtained without having to read messages. Comparing the discussion trees during the progress of the course should be enough to let mediators intervene in order to guarantee a maximum depth level, a minimum number of leafs and a desirable amount of messages. Mediators may also intervene when a given conference is not following an expected quantity of messages per level. This pattern is shown in Figure 6.

In level 0, where just a seminar message is expected, there is an average of one message in each tree of the course editions. In level 1, there is an average of 3 messages, which are the three questions proposed by the seminar leader. Level 2 contains the arguments responding direct to a question, forming a peak in the quantity of messages. In level 3 and thereafter the number of message decreases. If the quantity of messages in a given level of a tree in a given conference departs significantly from this pattern, mediators should intervene.

Message Categorization

Upon preparing a message, the author chooses the category that is most appropriate to the content being developed, providing a semantic aspect to the relationship between messages. Looking at the categories, learners and mediators estimate how the discussion is progressing and the probable content of the messages. The AulaNet does not force the adoption of a fixed set of categories. The teacher who plans the course can change the category set to the objectives and characteristics of the group and their tasks.



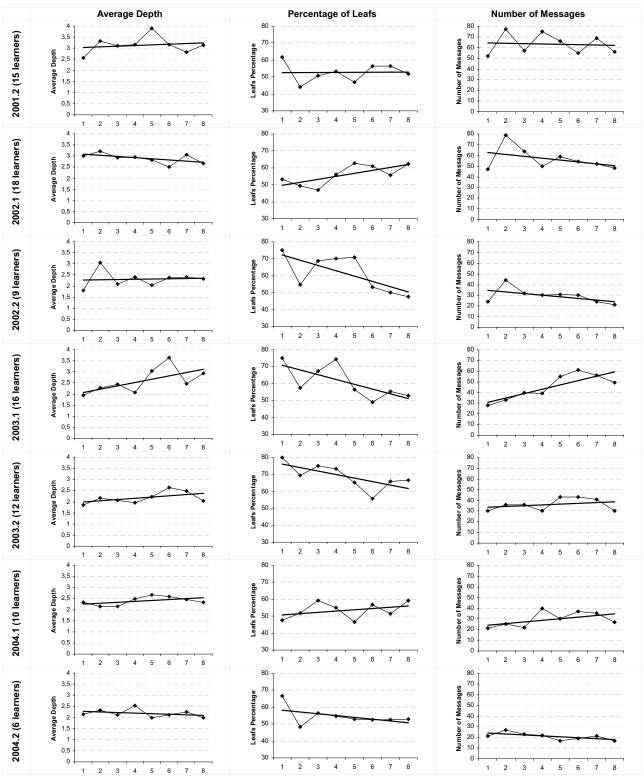


Figure 5. Average depth, leafs percentage and number of messages of each conference of the course editions

The categories adopted in the ITAE conferences were originally based on the IBIS' node types [Conklin, 1988]. The categories adaptation was made based on the information provided by follow-up reports furnished by AulaNet, like category usage per participant [Fuks, Gerosa & Lucena, 2002]. Currently, the categories defined in the course are: *Seminar*, for the root message of the discussion, posted by the seminar leader at the beginning of the week; *Question*, to propose discussion topics, also posted by the seminar leader; *Argumentation*, to answer the questions, posing the author's point of view in the message subject line and the arguments for it in the body of the message; *Counter-Argumentation*, to be used when the author states a position that is contrary to an argument; and finally, *Clarification*, to request or clarify doubts about a specific message.

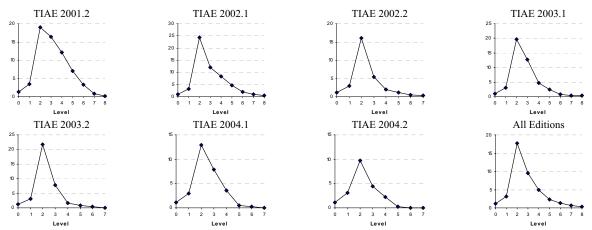


Figure 6. Average quantity of messages per tree level corresponding to the conferences

Figure 7 presents the percentage of messages of each category on the different tree levels of the course. As expected, one can observe that on level 0 (the tree root), the predominant category is *Seminar*, on level 1 it is *Question*, and on level 2 it is *Argumentation*. The *Counter-Argumentation* category begins to appear on level 3 and its relative usage increases; the use of the *Clarification* category begins to appear as of level 1 (it is possible to clarify a seminar or a question). Those messages whose relationship between the category and the level differ from what has been described, normally, derive from choosing a wrong category.

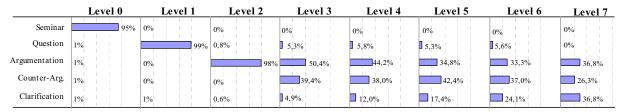


Figure 7. Percentage of utilization of the categories per tree level

The category also helps to identify the direction that the discussion is taking. For example, in a tree or a branch that only contains argumentation messages, there is no idea confrontation taking place, which is bad for the argumentation practice. Similarly, excessive counter-argumentation should attract mediators' attention. The group might have got into a deadlock or, even worst, there may be interpersonal conflicts taking place.

Message Size

As each category has a different semantics and influences the way that messages are composed, message size analysis was made for each category. Figure 8 presents the average values of characters for each category and average deviations. In this figure, one can see that the *Seminar* category is the one having the largest messages, followed by *Argumentation* and *Counter-Argumentation*. The shortest messages are those belonging to the *Question* and *Clarification* categories.

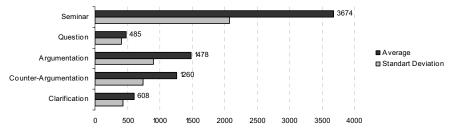
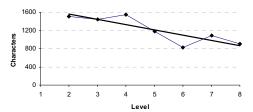


Figure 8. Quantity of characters per category

During an edition of the course, one of the learners said: "When we counter-argue we can be more succinct, since the subject matter is already known to all." If the subject is known to all (because it was presented during the previous messages) the author can go directly to the point that interests him or her. This can be noted in the chart in Figure 9, which shows a decline in the average quantity of characters per level in the *Argumentation* (correlation coefficient = -92%) categories. The analysis was restricted to these two categories because *Seminar* and *Question* do not show in higher levels and *Clarification* seems to be level independent regarding size (correlation coefficient = 48%).



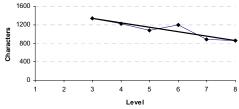
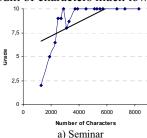
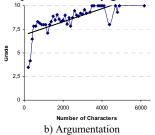


Figure 9. Quantity of characters in the messages per level of the argumentation (on the left) and counter-argumentation (on the right)

The expected amount of characters for a given message helps mediators identify problematic situation. Figure 10 presents a chart showing the amount of characters versus the average grade of the messages in the *Seminar*, *Argumentation* and *Counter-Argumentation* categories. It can be the seen that messages having an amount of characters much lower than the expected normally get a lower than average grade.





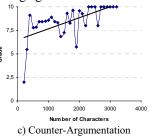
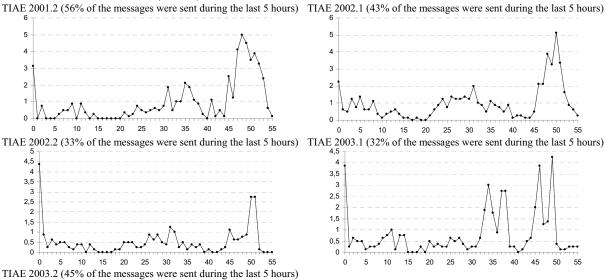


Figure 10. Amount of characters versus grade per category

Message Date

On the ITAE course, the conference goes for 50 hours: from 12 noon Monday to 2 pm Wednesday. Figure 11 presents the hourly rate of messages sent during conferences. Until the 2003.2 edition, it can be seen that there is a burst during the last five hours of the conference. In some cases, more than 50% of the messages are sent during this period of time. This phenomenon of students waiting until the last possible moment to carry out their tasks is well known and has been dubbed "Student Syndrome" [Goldratt, 1997]. The act of sending contributions near to the deadline disturbs an in-depth discussion, given that last-minute messages will neither be graded nor be answered during the discussion. This might be the reason for an excessive amount of leafs on the trees in some conferences, hence, less interaction.



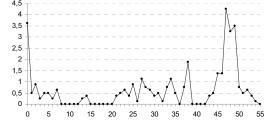


Figure 11. Average hourly rate of messages of the 8 conferences from 2001.2 until 2003.2 edition

In order to avoid this unwelcome behavior mediators have to encourage the earlier sending in of contributions. Unfortunately, our experience with this course has shown that this encouragement does not work. In the 2004.1 edition, the following experiment was conducted. The last 4 conferences had a different assessment rule than the first 4 conferences: if until the 25th hour the learner had not sent half of the expected amount of messages, the grade of all the messages sent during the following 25 hours would be divided by 2. The new rule seemed to work.

First 4 conferences (33% messages sent during the last 5 hours)



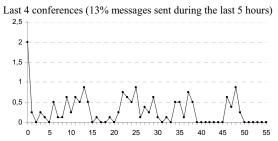


Figure 12. Average hourly rate of messages of the 4 first and 4 last conferences of the 2004.1 edition

In Figure 12, the second chart does not show message burst indicating that the rule has worked. The percentage of messages sent during the last 5 hours of conference felt from 33% in the first half of the course to 13% in the second half. Nevertheless, there are lower 25th and 50th peaks. However, now mediators and learners have room to access and answer the first batch of messages. The same thing can be seen in Figure 13, where all 8 conferences of the 2004.2 edition were assessed based on the aforementioned rule. In this conference, an average of 18% of messages was sent during the last 5 hours.

TIAE 2004.2 (18% of the messages were sent during the last 5 hours)

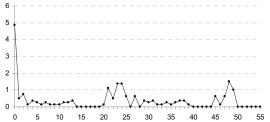


Figure 13. Average hourly rate of messages of the 8 conferences of the 2004.2 edition

For the moment, the no-need-to-read-messages-right-now measurements stop there. From this point on, mediators have a clear picture of what is going on in the conferences, but there is no way of avoiding the reading of messages. Anyhow, the conferences seem to be more balanced as an effect of the steering done by the mediators based on the use of statistical and visual information.

CONCLUSION

Communication among learners takes a fundamental role in the learning process through the exchange of information and points of view and interconnecting the group. Groupware technology supports collaborative learning activities, providing an environment where group interaction takes place. A forum provides a valuable tool that can be used to foster reflection in a paced learning. However, educational environments still do not offer computational aids that are appropriate for coordinating forums.

Message chaining, categorization, size and message timestamp are characteristics that help in the coordination of educational forums within ITAE. Based upon the form established by message chaining, it is possible to infer the level of interaction among course participants. Message categorization provides semantics to the way messages are interconnected, helping to identify the accomplishment of tasks, incorrect message nesting and the direction that the discussion is taking. The analysis of message timestamp makes it possible to identify the Student Syndrome phenomenon, which gets in the way of the development of an in-depth discussion.

By analyzing the characteristics of messages, mediators are able to better coordinate learners, having indication of when to intervene in order to keep the discussion from moving in an unwanted direction. Furthermore, these analyses could be used to develop filter for intelligent coordination and mechanisms for error reduction. It should be emphasized that these quantitative analyses provide indications and alerts about problematic situations, but also show whether the discussion is going well.

A well coordinated forum does not necessarily imply in learning taking place, it is still up to the teacher to insert forum-based relevant activities in the course dynamics and to motivate learners to ensure learning outcomes. The statistics and analysis presented in this paper help to better mediate the discussion process and to

identify uncommon situations, which does not necessarily means problems; it requires the teacher to check them out, inspecting the content of the discussion. In addition, although this paper is focused on the teacher mediation, the information presented may also be used by learners to better coordination themselves.

There is no ideal visual and statistical outcome that educators should steer their course towards. The teacher should interpret the statistical and visual information taking into consideration the course's and participants' characteristics. Final decision and judgment are still up to mediators. Discourse structure and message categorization also help to organize the recording of the dialogue, facilitating its subsequent recovery. Based upon the tree form, with the help of the categories, it is possible to obtain visual information about the structure of the discussion [Kirschner et al, 2003]. Mediators using collaborative learning environments to carry out their activities should take these characteristics into account for the better coordination of educational forums.

ACKNOWLEDGEMENT

The AulaNet project is partially financed by Fundação Padre Leonel Franca and by the Ministry of Science and Technology through its Project ESSMA, grant nº 552068/2002-0. It is also financed by individual grants awarded by the National Research Council to: Carlos José Pereira de Lucena nº 300031/92-0, Hugo Fuks nº 303055/02-2 and Marco Aurélio Gerosa nº 140103/02-3. Mariano Pimentel received a grant from CAPES.

REFERENCES

- Benbunan-Fich, R. & Hiltz, S. R. (1999): Impacts of Asynchronous Learning Networks on Individual and Group Problem Solving: A Field Experiment, Group Decision and Negotiation, Vol.8, pp. 409-426.
- Bull, S., Greer, J., McCalla, G. & Kettel, L. (2001) "Help-Seeking in an Asynchronous Help Forum", Proceedings of International Conference on Artificial Intelligence in Education 2001.
- Conklin, J. (1988) "Hypertext: an introduction and Survey", Computer Supported Cooperative Work: A Book of Readings, pp. 423-476
- Ellis, C.A., Gibbs, S.J. & Rein, G.L. (1991): Groupware Some Issues and Experiences. Communications of the ACM, vol. 34, no. 1, pp. 38-58.
- Fuks, H., Gerosa, M.A. & Lucena, C.J.P. (2002), "The Development and Application of Distance Learning on the Internet", Open Learning Journal, V.17, N.1, pp. 23-38.
- Funaro, G.M. & Montell, F. (1999), "Pedagogical Roles and Implementation Guidelines for Online Communication Tools", ALN Magazine, Volume 3, Issue 2.
- Gerosa, M.A., Fuks, H. & Lucena, C.J.P. (2001), "Use of categorization and structuring of messages in order to organize the discussion and reduce information overload in asynchronous textual communication tools", CRIWG 2001, Germany, pp 136-141.
- Goldratt, E.M. (1997) "Critical Chain", The North River Press Publishing Corporation, Great Barrington.
- Graham, M., Scarborough, H. & Goodwin, C. (1999) "Implementing Computer Mediated Communication in an Undergraduate Course A Practical Experience," JALN, Vol.3 No.1.
- Harasim, L., Hiltz, S. R., Teles, L., & Turoff, M. (1997) "Learning networks: A field guide to teaching and online learning", 3rd ed., MIT Press, 1997.
- Hiltz, S.R. (1994) "The Virtual Classroom: Learning without limits via computer networks," Norwood, New Jersey: Ablex Publishing Corporation.
- Kirschner, P.A., Shum, S.J.B. & Carr, C.S. (eds), Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making, Springer, 2003.
- LaGrandeur, K. (1996) "Using Electronic Discussion to Teach Literary Analysis," Computers & Texts, 12.
- Laurillau, Y. & Nigay, L. (2002): Clover architecture for groupware, Proceedings of the ACM Conference on Computer Supported Cooperative Work, Louisiana, USA, pp. 236-245.
- Pimentel, M.G., Fuks, H. & Lucena, C.J.P. (2003) "Co-text Loss in Textual Chat Tools", CONTEXT 2003, LNAI 2680, Stanford, CA, USA, June, pp 483-490, 2003.
- Raposo, A.B. & Fuks, H. (2002) "Defining Task Interdependencies and Coordination Mechanisms for Collaborative Systems", Cooperative Systems Design, IOS Press, 88-103.
- Salmon, G. (2000), E-moderating: the key to teaching and learning online, London, Kogan Page
- Stahl, G. (2001) "WebGuide: Guiding collaborative learning on the Web with perspectives", Journal of Interactive Media in Education, 2001.
- Straus, S.G. (1996) "Getting a clue: the effects of communication media and information distribution on participation and performance in computer mediated and face-to-face groups," Small Group Research, 27 (1), 115-142.