Gerosa, M.A., Pimentel, M.G., Fuks, H. & Lucena, C.J.P. (2004), "Analyzing Discourse Structure to Coordinate Educational Forums", The 7th International Conference on Intelligent Tutoring Systems - ITS-2004, Maceió-AL, August 30th to September, 3rd 2004, Lecture Notes on Computer Science LNCS 3220, Lester, J.C., Vicari, R.M. & Paraguaçu, F. (eds), ISBN 3540-229485,ISSN 0302-9743, pp. 262-272.
Available at http://www.les.inf.puc-rio.br/groupware

Analyzing Discourse Structure to Coordinate Educational Forums

Marco Aurélio Gerosa, Mariano Gomes Pimentel, Hugo Fuks & Carlos 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 this paper, aspects related to discourse structure like message chaining, message date and categorization are used by teachers to coordinate educational forums in the AulaNet environment. These aspects can be computationally analyzed without having to inspect the content of each message. This analysis could be applied to forums in other educational environments.

1 Introduction

As an asynchronous communication tool, a forum makes it possible for learners to participate at their own pace while allowing them more time to think. However, educational environments still do not offer computational aids that are appropriate for coordinating forums. The majority of the environments present a typical implementation that does not take into account educational aspects and it remains up to the teacher (without specific computational support) to collect and analyze the information that is necessary to coordinate group discussion.

Coordination is the effort needed to organize a group to enable it to work as a team in a manner that channels communication and cooperation towards the group's objective [8]. When coordinating a group discussion in a forum, among other factors the teacher must be prepared to ensure that all of the learners are participating, that the contributions add value to the discussion, that the conversation does not go off on non-productive tangents and that good contributions are encouraged.

This article focuses on message chaining, categorization and timestamp. These message attributes help in the coordination of educational forums without the teacher inspecting the content of individual messages and in a manner that allows computational support for this

In a forum, where messages are structured hierarchically (tree), it is possible to obtain indications about the depth of the discussion and the level of interaction by observing the form of this tree. Measurements such as the average depth level and percentage of leaves provide indications about how a discussion is going. Message categorization can also help to identify the types of messages, making a separate analysis of each message type possible. By analyzing the date that messages were sent, among other factors it is possible to identify the amount of time between the sending of messages, the day of the week and the hour expected for messages to be sent. Comparing

this data also makes it possible to obtain 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. Based upon these aspects, the course coordinator can evaluate how a discussion is evolving, giving him enough time to redirect the discussion and, for example, to check up on the effects of his interventions.

The AulaNet environment supports the creation of educational forums, as presented in Section 2. The Information Technology Applied to Education (ITAE) course, which provided the data for the analyses presented in this article, also is discussed in this section. Section 3 shows the analyses about discourse structure. Section 4 concludes the article.

2 The Conferences Service in the AulaNet Environment

The AulaNet is an environment based on a groupware approach for teaching-learning via the Web that has been under development since June 1997 by the Software Engineering Laboratory of the Catholic University of Rio de Janeiro (PUC-Rio). The AulaNet is freeware and is available in Portuguese, English and Spanish versions at groupware.les.inf.puc-rio.br and www.eduweb.com.br.

The Information Technology Applied to Education (ITAE) course has been taught since 1998 on the Computer Science Department of PUC-Rio. This course is being taught entirely at a distance through the AulaNet environment. Its objective is to get learners to collaborate using information technology, becoming Web-based educators [2]. The course seeks to build a learning network [5] where the group learns, mainly, through the interaction of its participants in collaborative learning activities.

The ITAE is organized by subject, with one topic discussed per week. Learners read selected content relating to the topic, conduct research to expand their knowledge and participate in a discussion about specific questions of the subject being studied. The discussion is carried out over three consecutive days using the AulaNet's Conferences service.

In the ITAE, 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 of the role of the *seminar leader*, being responsible for preparing a seminar message followed by three questions, used by group members to develop their argumentation. During this phase, the seminar leader is responsible for keeping the discussion going and maintaining the conference's dynamics.

Each Conference message is evaluated and commented upon individually by the course's mediators in order to provide guidance to learners about how to build knowledge and prepare their texts; the idea is to avoid the sending in of contributions that do not add value to the group. The problems that are encountered in the 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.

3. Coordination of Educational Forums

Analyses about message chaining, categorization and timestamp are presented in this section, showing how these factors can help in the coordination of educational forums. The data and examples were collected from five editions of the ITAE course.

3.1. Message Chaining

Communication tools have different ways of structuring messages: linear (list), hierarchical (tree) or network (graph), as can be seen in Figure 1. 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 always better than another. Linear structuring is appropriate for communication in which chronological order is more important than any eventual relationships between the messages, such as the sending of notices, reports and news. Hierarchical structuring is appropriate for viewing the width and the depth of the discussion, making it possible to structure messages sharing the same subject on the same branch. However, since there is no way to link a message between one branch and another, the tree can only grow and, thus, the discussion takes place in diverging lines [9]. Network structuring can be used to seek convergence of the discussion.

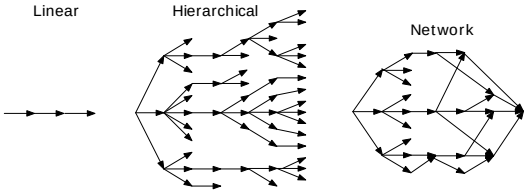


Fig. 1. Examples of discussion structure

The forum has a hierarchical structure. In the ITAE, the forum, based on the Conferences service, is used for the in-depth discussion of the course's subject matter. The AulaNet makes it possible for the author of the message, at the moment he or she is preparing it, to select a category from a set that have been previously defined by the course coordinator [3]. The available categories available in the ITAE course, used to identify the message type, are *Seminar*, *Question*, *Argumentation*, *Counter-Argumentation* and *Clarification*, originally based on the IBIS' node types [1]. According to the dynamics of the course, at the beginning of each week a previously selected learner 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 answer to and discuss these questions.

The format of the resulting tree indicates the depth of the discussion, thus, the level of interaction [7]. For example, a tree that has only three levels indicates that there was almost no interaction, given that level zero is the seminar, level one comprises the questions and level two comprises the answers to the questions. That means the learners only answered the questions without discussing the ideas with each other. The trees extracted from the conferences of the five editions of the ITAE course are shown in Figure 2.

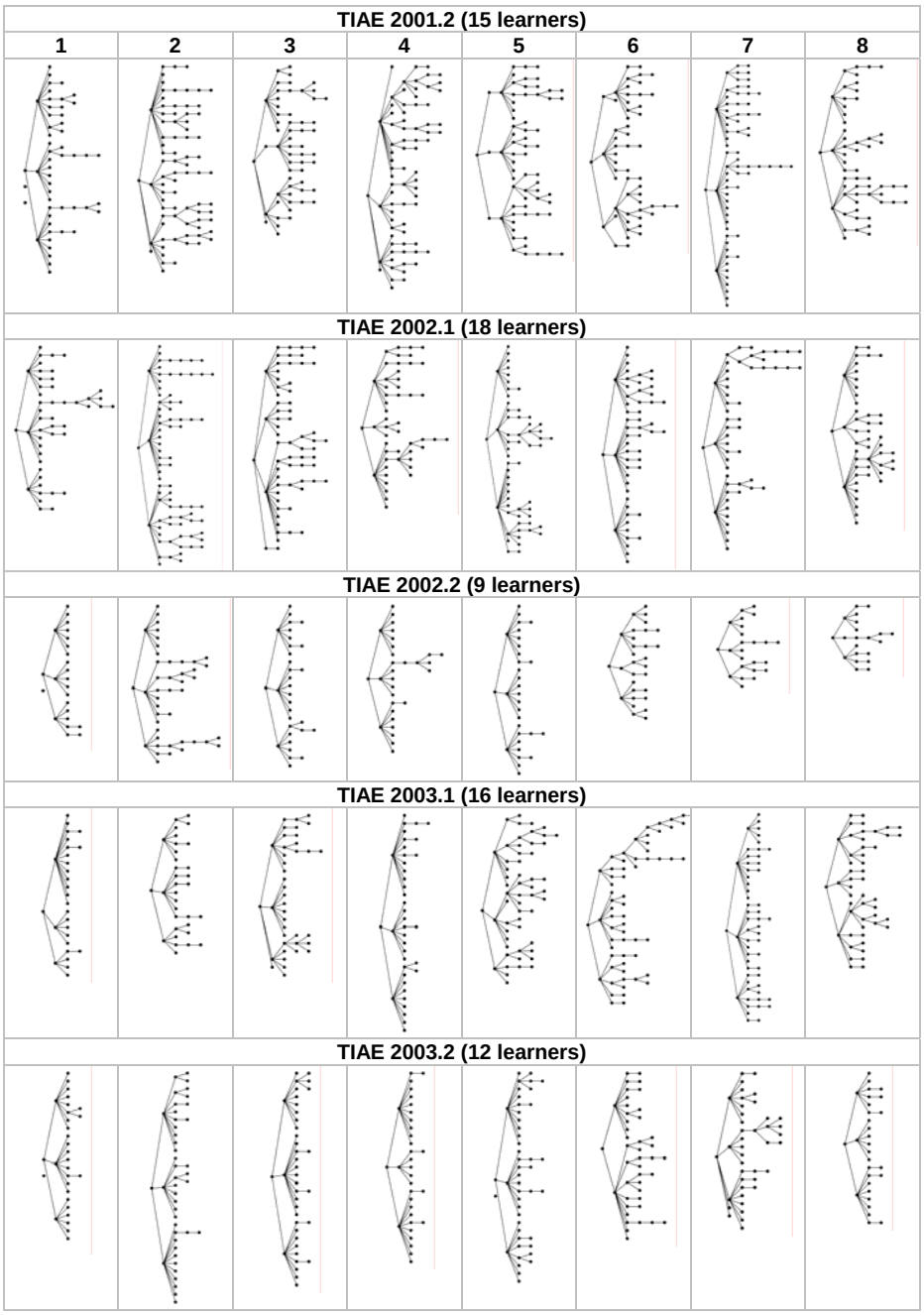


Fig. 2. Trees extracted from the Conferences of the five editions of the ITAE course

Visually, upon analyzing the trees in Figure 2, it can be seen that in ITAE 2001.2 and ITAE 2002.1 the tree became shallower over the period of time the course was being taught. In the ITAE 2002.2, the tree depth changed from one conference to

another. In the ITAE 2003.1 and ITAE 2003.2, the tree depth increased during the course, despite the fact that there were a number of shallow trees. It is also possible to observe in this figure that, in all editions, conference one corresponding tree is the shallowest. Although the depth of a tree does not in and of itself ensures that in-depth discussion took place, it is a good indication. The teacher, then, can initiate a more detailed investigation about the discussion depth. Based on the visualization of the trees, it is possible 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.

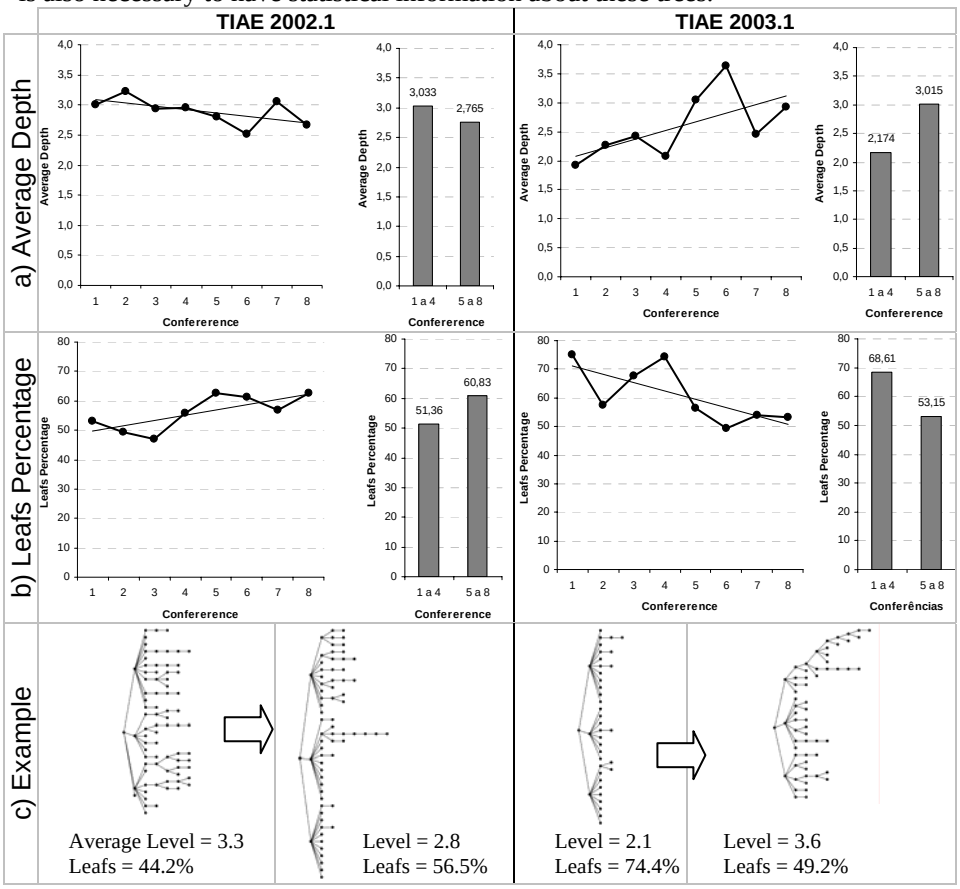


Fig. 3. Comparison of the Conferences of the ITAE 2002.1 and 2003.1 editions

It can be seen in Figure 3 that the average depth of the tree in the ITAE 2002.1 edition declined while the percentage of messages without answers (leaves) increased, which indicates that learners were having diminishing interaction as the course advanced. In this edition, in the first four Conferences 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 leaves were 61%. For its part, in the ITAE 2003.1, learners interacted more over the course of the conferences: the tree corresponding to the discussion was getting deeper while the percentage of messages

without answers was decreasing. The average level was 2.2 in the first four Conferences, increasing to 3.0 in the last four Conferences, while the percentage of messages without answers went from 69% in the first four Conferences to 53% in the last four. Figure 3 also presents a comparison between a conference at the beginning and another at the end of each one of these editions, emphasizing their difference. The trees shown in Figure 2 and the charts in Figure 3 indicate that the interaction on ITAE 2002.1 edition declined over the course of the conferences, while the interaction on ITAE 2003.1 edition increased.

All of this data was obtained without having to inspect the content of the messages. Comparing the evolution of the form and of the information about the trees in the course allows teachers to intervene when they perceive that the level of interaction has fallen or when the Conference is not reaching the desired depth level. Next in Figure 4 is shown the expected quantity of messages per level.

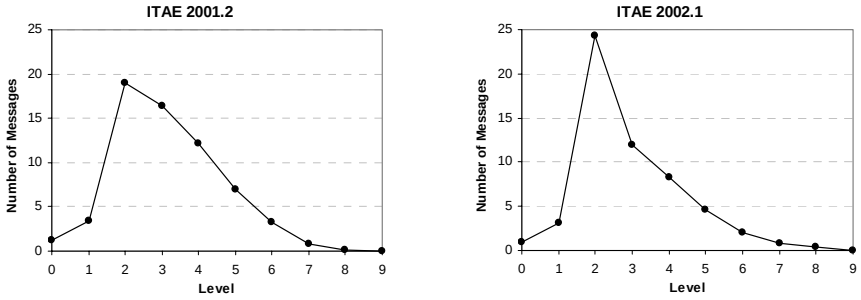


Fig. 4. Average quantity of messages per tree level corresponding to the conferences

A peak in the average quantity of messages at level 2 can be seen in Figure 4. In level 0, where just a seminar message is expected (sent by a learner at the beginning of the week), there is an average of one message in each tree of the course editions analyzed. In level 1, there is an average of 3 messages, which are the three questions proposed by the seminar leader. In level 2, where the arguments are sent in response to the questions, there is a peak in the quantity of messages. In level 3 and thereafter if the quantity of messages of the tree in any given Conference departs significantly from this standard, the teacher should investigate to check what is happening.

3.2. 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 the messages. The AulaNet does not force the adoption of fixed sets of categories. The coordinating teacher—the one who plans the course—can adjust the category set to the objectives and characteristics of the group and the tasks.

Upon viewing the messages of a Conference, participants immediately realize the category to which the message belongs (between brackets) together with its title, author and date. Thus, it is possible to estimate how the discussion is progressing and what is the probable content of the messages. The AulaNet also implements reports about the utilization of the categories per participant, in order to facilitate the future refining of the category set and to obtain indications about the characteristics of the participants and their compliance with tasks. Categorization also helps organize the

discussion in a manner that favors decision making and maintenance of communication memory [2].

The categories adopted in the ITAE Conferences reflect the course dynamics. They 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, offering 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.

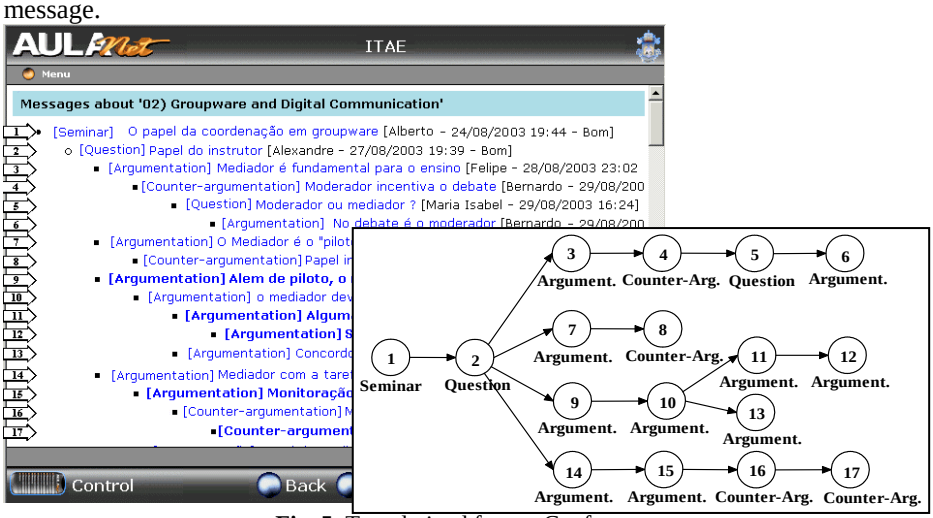


Fig. 5. Tree derived from a Conference

Figure 5 presents a portion of dialogue from a Conference with numbered messages and a tree equivalent of this portion. Looking at the categories, it is possible to perceive the semantics of the relationships between these messages. For example, message 4 is a counter-argument to message 3; 5 questions 4; 6 answers to the question posted by 5 through an argument; and so forth.

It is also possible to identify the differences between messages of different categories. For example, in this article, level three categories were analyzed taking into account their grades and quantity of characters.

As previously explained, the *Seminar* category is used in the first message of the Conference (level 0); next, three messages from the *Question* category are associated with level 1; and the answers to the *Argumentation* category appear on level 2. As of level 3, messages from all of the categories can appear, with the exception of the *Seminar* category. Figure 6 presents the percentage of messages of each category on the different tree levels of the ITAE course edition. 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; 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 the wrong category.

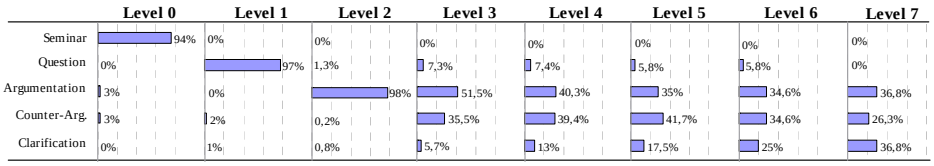


Fig. 6. Percentage of utilization of the categories per tree level

Message size also has a different expected value for each one of the categories, given that each category has its own objectives and semantic. Figure 7 presents the average values of characters for its 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 in the *Question* and *Clarification* categories.

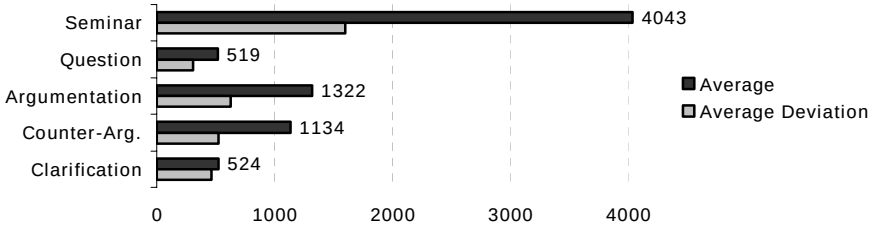
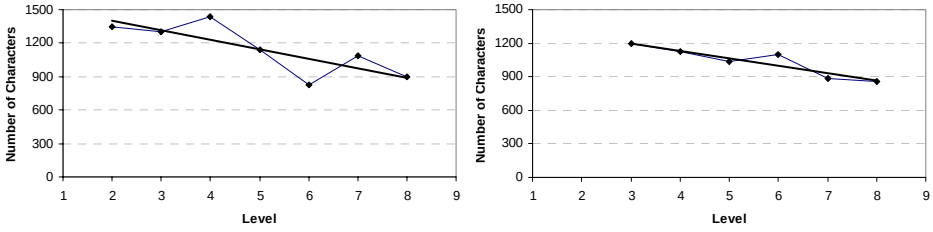


Fig. 7. Quantity of characters per category

At some point, during the course, one of the ITAE learners said: "When we counter-argue we can be more succinct, since the subject matter already is known to all." This statement is in keeping with the chart in Figure 7. If the subject is known to all (it was presented during the previous messages) the author can go directly to the point that interests him or her. Somehow, this also can be noted in the chart in Figure 8, which presents a decline in the average quantity of the characters per level in the *Argumentation* (correlation = -80%) and *Counter-Argumentation* (correlation = -93%) categories.

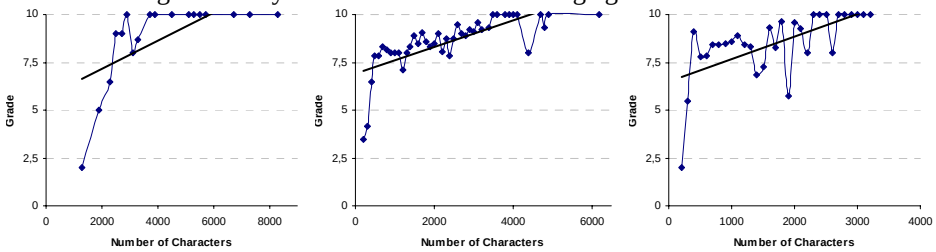


a) Argumentation Category b) Conter-Argumentation Category

Fig. 8. Quantity of characters in the messages per level

Knowing in advanced the quantity of characters expected for a given message (based on its category and the level) helps the teacher evaluate the message and orient the learners, giving them an idea about how much they should write in their messages. Figure 9 shows a chart about the quantity of characters versus the average grade of the messages in the *Seminar*, *Argumentation* and *Counter-Argumentation*

categories. It can be the seen that messages with a quantity of characters much lower than the average normally receive a lower than average grade.



a) Seminar b) Argumentation c) Conter-Argumentation

Fig. 9. Quantity of characters versus grade per category

The category also helps to identify the direction that the discussion is taking. For example, in a tree or a branch only containing argumentation messages, there is probably no idea confrontation taking place. It is expected that the clashing of ideas helps to involve more participants into the discussion, thus, bringing up confronting points of view. Similarly, excessive counter-argumentation should attract mediator's attention. The group might be getting too involved into a controversy or, even worst, there may be interpersonal conflicts taking place.

3.3. Message Timestamp

On the ITAE course, the Conference takes place over 50 hours: from 12 noon Monday to 2 p.m. Wednesday. Over the course of these hours, learners post messages answering questions and arguments and counter-arguments to previous messages.

Figure 10 presents the frequency of the messages sent during the Conferences of the ITAE 2003.2 edition. On this edition, it can be seen that almost half of the message total was sent during the last five hours of the Conference. This phenomenon of students waiting until the last moment possible to carry out their tasks is well known and has been dubbed "Student Syndrome" [4].

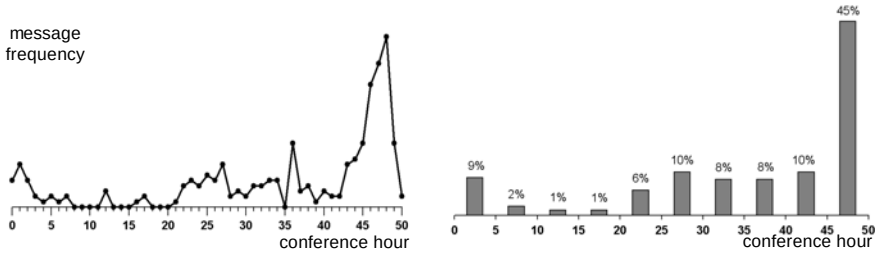


Fig. 10. Frequency of messages over the course of the conferences of the ITAE 2003.2 edition

The last-minute behavior observed in Figure 10 reminds the teacher to encourage earlier sending in of contributions. The act of sending contributions near to the deadline disturbs an in-depth discussion, given that last-minute messages will neither be graded during the discussion nor be answered. This might be the reason for an excessive amount of leaves on the trees in some conferences.

4. Conclusion

Message chaining, categorization and message timestamp are factors 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 connected, helping to identify the accomplishment of tasks, identification of incorrectly message nesting and the direction 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 and the orientation provided by an evaluation of the messages.

By analyzing the characteristics of the messages, teachers are able to better coordinate learners, knowing 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 to the teachers indications and alerts about situations where problems exist and where the discussion is going well. However, final decision and judgment are still up to the teacher.

Finally, 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 [6]. Teachers using collaborative learning environments to carry out their activities should take these factors into account for the better coordination of educational forums.

References

1. Conklin, J. (1988) “Hypertext: an introduction and Survey”, Computer Supported Cooperative Work: A Book of Readings, pp. 423-476
2. 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.
3. 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.
4. Goldratt, E.M. (1997) “Critical Chain”, The North River Press Publishing Corporation, Great Barrington.
5. 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.
6. Kirschner, P.A., Shum, S.J.B. & Carr, C.S. (eds), Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making, Springer, 2003.
7. Pimentel, M. G., Sampaio, F. F. (2002) “Comunicografia”, Revista Brasileira de Informática na Educação - SBC, v. 10, n. 1. Porto Alegre, Brasil.
8. Raposo, A.B. & Fuks, H. (2002) “Defining Task Interdependencies and Coordination Mechanisms for Collaborative Systems”, Cooperative Systems Design, IOS Press, 88-103.
9. Stahl, G. (2001) “WebGuide: Guiding collaborative learning on the Web with perspectives”, Journal of Interactive Media in Education, 2001.