Use of Categorization and Structuring of Messages in order to Organize the Discussion and Reduce Information Overload in Asynchronous Textual Communication Tools

Marco Aurélio Gerosa, Hugo Fuks and Carlos José Pereira de Lucena
Catholic University of Rio de Janeiro – PUC-Rio
Software Engineering Laboratory – Computer Science Department
R. M. S. Vicente 225 – Gávea – Rio de Janeiro – RJ – Brazil
{gerosa, hugo, lucena}@inf.puc-rio.br

Abstract

This paper shows how the use of categorization and structuring of messages in asynchronous textual communication tools could be useful in course deliver via the Internet in order to facilitate the argumentation and to guide the participants to reflect about their messages. Although the use of categorization caused an increase in the total number of messages, there was a reduction of the information overload and an increase in the quality of the discussion. We also present how the message categorization was used in a distance course delivered through the AulaNet environment, showing how we defined and improved the set of categories, and how it helped accompanying the learners.

1. Introduction

The new telecommunications technologies are changing the way humanity lives, learns and works [1]. With the Internet making it possible for people scattered around the world to communicate and share information, a number of communities with similar interests have emerged, and meet virtually to discuss a great variety of subjects at any time and physical distance.

These communities tend to generate a large volume of information and the communication tools must be adapted in a way, which facilitates and organizes the discussion, and supplies instruments to reduce the information overload on people [2]. Structuring the discussion and supplying simple and representative pieces of information about the messages' contents, in a way that helps the identification of their relevance and context, could achieve this reduction. Some of this information could be extracted automatically, such as the date it was sent and the sender, and other information, such as the title of the message and its priority, needs to be supplied by the author.

In order to aid the pre-identification of message content and its automatic classification, message categorization could be employed within the communication tools. The author should select from a pre-defined set, the category that is most appropriate for the message he is sending. If properly employed, besides supplying additional information about the content of a message, thus facilitating contextualization for readers, categorization also can help in structuring, organizing and focusing the discussion, making it more objective and explicit [3]. Upon supplying information about content such as the category and the title, the author is obliged to reflect on what he is writing, increasing the quality of the discussion and leading to greater learning [4]. Despite its advantages, the choosing of a category requires an additional effort in the preparation of a message and the author's concepts, points of view and ideas must be well thought out in order to express himself in a way that separates his discourse into different messages, each one with its own category and relationships [5].

In this paper we will analyze how message categorization and structuring have energized discussions and how the set of categories was defined and refined. For this study, we have used a course via Internet within the AulaNet, an environment for the creation, distribution and administration of Web-based courses, for which we will propose some improvements based upon the difficulties we have observed.

2. Message categorization

The first step of the message categorization implementation is to define the set of categories that will be used. This should be minimal, sufficient and unambiguous in a way that does not restrict people discourse and lets them categorize their messages without much of an effort [6]. In order to define the initial set of categories, an analysis should probe the participants' behavior, and the objectives and the form of the

0-7695-1351-4/01 $10.00 © 2001 IEEE

discussion [7,8]. Being careful to define clearly the categories can minimize doubts and confusions. Ambiguous categories lead to the spreading around of messages that should be in the same category, leading to erroneous automatic messages grouping.

3. The AulaNet learning environment

The AulaNet is a groupware for the creation, distribution and administration of Web-based courses. Its services are organized based upon the principle that, in order to learn in a group, an individual must share his ideas (to communicate), be in tune with the members of the group (to coordinate) and carry out the tasks in a satisfactory manner (to cooperate) [9].

The communication mechanisms of AulaNet are: a tool for electronic mail messages with the instructor (Contact with the Teachers); electronic mail with the group (Discussion Group); an asynchronous threaded text discussion (Interest Group); a synchronous text chat conference (Debate); and a tool for instantaneous exchange of messages with on-line participants (Contact with Participants). These mechanisms are put at the disposal of the teacher at the time of the creation of the course and while it is being delivered, making it possible for the teacher to select and re-configure which services he wants to make available for the learners.

The message categorization is implemented in the Discussion Group and Interest Group. The teacher defines which categories will be made available for the participants' choice at moment they post their messages. It is possible to create, rename, deactivate and activate categories at any time of the course.

If the learner's choice of category has not been appropriate, the instructor may change it even after it has been sent. When this happens, the author of the message receives an electronic mail notification.

Upon viewing the list of messages of a service, the participants check which category the message belongs through its name in brackets, as well as other information such as the title, the author and the date, as shown in Figure 1 and Figure 2. Thus it is possible to quickly identify how the discussion is going on and what is the probable content of each message. The AulaNet also supplies reports that group messages by category, helping the instructors and the students to follow the quantity of messages in each category.

4. The course Information Technology Applied to Education

The ITAE course has been offered since 1998 by the Catholic University of Rio de Janeiro and currently is taught entirely via the Internet within the AulaNet environment. The objective of the course is to capacitate educators to work in groups, using the new information technologies in teaching/learning. Since the focus of this paper is on communication tools using message categorization, only the use of Discussion Group and Interest Group services will be analyzed.

The Discussion Group is used for communication with the entire class. In this service, when a message is posted, besides being stored in the environment it is also sent to the electronic mailbox of all members of the group. On the environment, the messages are shown in the form of a list sorted chronologically, as shown in Figure 1. This service is used in the course mainly for coordination messages from the instructors.

The Interest Group operates like a conferencing system where it is possible to answer a message and the answers are nested below it forming a threaded discussion. This structure permits the organization of a discussion by topics, with related messages remaining below the original topic message, as shown in Figure 2. The Interest Group is used in the ITAE to develop the course's themes, as well as those selected by the participants.

According to the course methodology, for each course topic a learner is designated as seminar leader. He is responsible for carrying out research on the subject and preparing the Seminar, a text that reports about what already exists and what is being researched, showing his or her point of view. The other learners are responsible for sending contributions about the seminar topic, going into greater depth on the subject matter.

5. Categorization of messages in ITAE

The experience with message categorization on the ITAE began during the first semester of 2000. In this section, it is reported how was the implementation of the message categorization, how the categories set was defined and refined, and how it was associated with the structuring of message.

5.1. First semester of 2000 (2000.1)

In the first semester of 2000, the messages of the seminar leader and the contributions were posted on the Discussion Group and the Interest Group was used for the in-depth discussion of other topics that emerged during the course. The categories that were originally defined for the Discussion Group were: Presentation, for the presentation of the participant at the start of the course; Seminar and Seminar Contribution, for the messages from the seminar leader and the contributions; Operational Problems, to report problems; and Generic, for messages that do not fit into any of the previous ones.

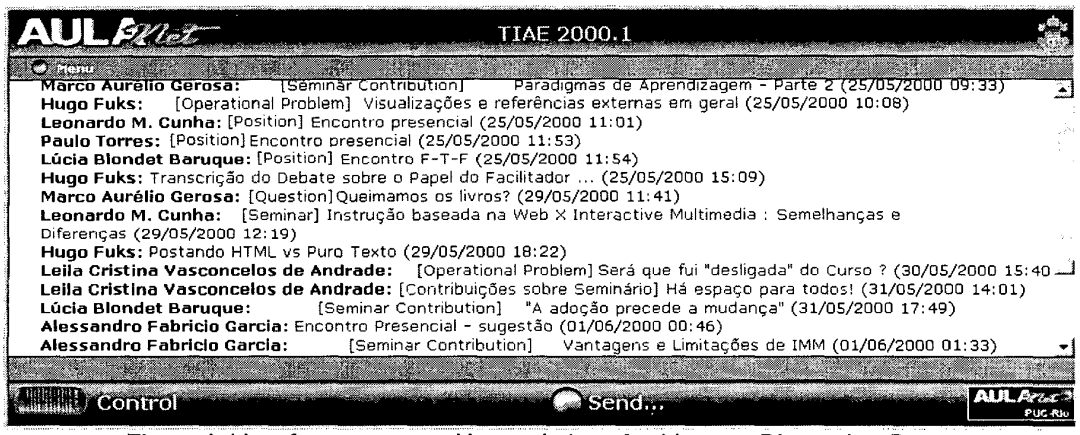

Figure 1. List of messages and intercalation of subjects at Discussion Group

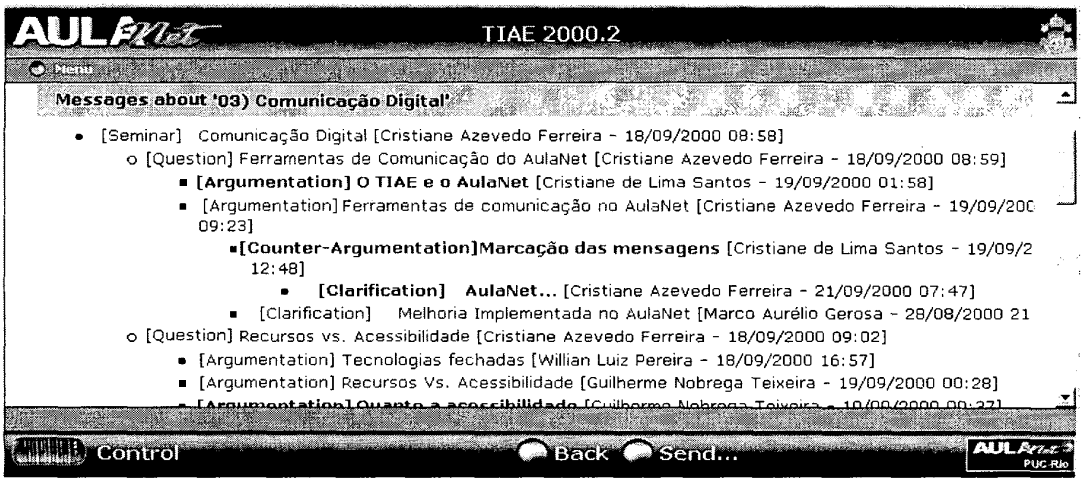

Figure 2. A dialogue within an Interest Group showing the structuring of the messages

During the semester, it was noted that this set of categories were used basically for exhibiting ideas and posting notices, not offering mechanisms for debating the topics in depth. To solve this difficulty and to stimulate interaction in the Discussion Group, three more categories were offered. Based upon the IBIS proposal [10, 11], the categories offered were Question, to propose questions and discussion topics; Position, to express a point of view answering a question; and Argumentation, to supply the reasons for the position.

Since these new categories presumed short messages, delimited by the scope of the category and were strongly inter-related (a Position is always about a Question and an Argumentation is about a Position), the linearity of the Discussion Group showed itself to be an obstacle (Figure 1). It was not possible to know to which Question a Position referred and about which Position an Argumentation was referring, without reading the title or the body of the message. In addition, the learners had difficulties in using them, mixing more than one concept in a message [10].

Looking for data that could base a reformulation on the use of the categories for the subsequent semester, the purpose of each one of the 205 Discussion Group messages was analyzed and classified (see Table 1).

It can be noted that there is two groups of messages in Table 1: 107 messages regarding group coordination (Items 1 to 6) and 98 messages regarding content of the course (Item 7). Once again, these two types of messages were sent to the same list and they were intercalated with each other, making the list disorganized. The table also shows that the majority (94%) of the messages send by the instructor was about group coordination. Finally, 35% of all messages pertained to the Generic category, which indicates that the set of categories adopted was not adequate.

Table 1. Classification of the Discussion Group messages of 2000.1

Purpose	Total	Role		Category Used							
		Instructor	Learners	Presentation	Seminar	Contribution	Op. Problem	Question	Position	Argumentation	Generic
1. Presentation of the participants	**10**	1	9	9	-	-	-	-	-	-	1
2. Motivation and activity follow-up	**9**	9	-	-	-	-	-	-	-	-	9
3. Information about methodology	**16**	16	-	-	1	-	2	-	-	-	13
4. Notices about the course	**31**	31	-	-	-	-	2	-	-	-	29
5. Operational problems	**8**	2	6	-	-	-	8	-	-	-	-
6. Discussion about the course running	**33**	5	28	-	-	-	-	9	5	2	17
7. Discussion of course content	**98**	4	94	-	17	75	-	2	1	-	3
TOTAL	**205**	**68**	**137**	**9**	**18**	**75**	**12**	**11**	**6**	**2**	**72**

5.2. Second semester of 2000 (2000.2)

In order to solve the problem of lack of structure and organization of the Discussion Group messages, which were not divided into topics and which were presented in chronological order, the discussion about the content of the course was transferred to the Interest Group. A new forum was created for each class so that the messages were organized and compartmentalized within it.

In the Discussion Group, which was used for the group coordination, the categories Presentation, for the participants present themselves; Operational Problems, to report about problems; and Generic, to be used when the message did not fit other categories were maintained. Since the discussion about the content of the course was moved to the Interest Group, the categories Seminar, Seminar Contribution, Question, Position and Argumentation were taken out of the Discussion Group.

In order to reduce the number of messages in the Generic category, new categories were created to isolate the messages about the group coordination, which represented 96% of the generic messages on the previous semester, as observed in Table 1. The categories created were Notices, for notifications; Monograph, for messages related to the final work of the course; and Evaluation, for learners to evaluate the course and the AulaNet environment.

The categorization of the messages also was adopted for the Interest Group, using some of the Discussion Group categories: Seminar and Seminar Contribution for the seminar and contributions of the learners, and Question for discussion of topics. Due to the difficulty of using the IBIS methodology noted in the previous semester, it was decided to eliminate the Position category and change the concept of the Argumentation category. This category would respond directly to a Question, providing the opinion of the author on the title of the message, and the explanation and arguments on its body, joining the old concepts of the Position and Argumentation categories.

For greater clarity about the objective of the Argumentation, a category called Counter-Argumentation was created, with the same structure of Argumentation, but used to disagree with messages of that category. Furthermore, the categories of Case, to relate experiences, and Generic, for messages that did not fit any other category, were created.

During the semester, it was noted that the Question category was used for two different purposes. Besides being used to propose topics for discussion, as it was initially planned, it also was used to post doubts about the course or about other messages. This second form of use generally generated just one answer that did not necessarily had a sense of argumentation or point of view. To make the discussion more clear, it was decided, with the help of the learners, to offer two more categories: Doubt, for simple questions that did not generate debate; and Clarification, to solve doubts and misunderstandings.

This semester, the ITAE had a group of seven learners and three instructors, while in the last semester were seven learners and one instructor. The summary of the number of messages in each category and the comparison with the previous semester is shown in Table 2, where DG means a Discussion Group category, IG is an Interest Group category and the number in parentheses is the quantity of messages for the category.

Table 2. Comparative table about the use of categories in the two semesters analyzed

2000.1	2000.2
DG – Seminar (18)	IG – Seminar (13)
DG – Seminar Contribution (75)	IG – Seminar Contribution (33)
DG – Presentation (9)	**DG – Presentation (12)**
DG – Operational Problems (12)	**DG – Operational Problems (14)**
DG – Question (11)	IG – Question (65)
DG – Position (6)	IG – Argumentation (129)
DG – Argumentation (2)	IG – Counter- Argumentation (26)
DG – Generic (72)	**DG – Generic (20)**
IG – Generic (90)	IG – Generic (12)
	IG – Doubt (7)
	IG – Clarification (25)
	IG – Case (2)
	DG – Evaluation (18)
	DG – Notice (50)
	DG – Monograph (20)
Total: 288	**Total: 446**

It can be seen that because of the structuring of the discussion through the nesting of the messages and the refining of the sets of categories adopted, learner participation increased in the discussion about the course

topics, increasing the average number of messages per class from 7 to 24. According to the instructors, there was also a gain in the quality of the work caused by an improvement in the learning process in relation to the previous semesters, mostly because of self-discipline imposed by the use of the categories and structuring.

The learners chose to use the Question, Argumentation and Counter-Argumentation categories instead of using the Seminar Contribution category, whose number of messages fell from 75 to 33. The quantity of messages in those categories in the second semester (220 messages) is 11 times greater than the quantity of messages of the Question, Position and Argumentation categories in the first semester (19 messages). The quantity of messages in the Seminar, Presentation and Operational Problems categories did not change from one semester to another, because there were no changes in methodology that would benefit or restrict the use of these categories. There was a significant reduction in the quantity of messages in the Generic category of the Discussion Group (72 to 20) through the adoption of the Notice category. Finally, the Case category practically was not used and the Generic category of the Internet Group had a low level of utilization (4% of the messages).

Table 3. Classification of the Discussion Group Messages 2000.2

Purpose	Total	Role		Category Used					
		Instructor	Learners	Presentation	Notice	Op. Problem	Evaluation	Monograph	Generic
1. Presentation	**12**	3	9	12	-	-	-	-	-
2. Motivation	**19**	16	3	-	14	-	3	2	-
3. Informs and explanations	**5**	5	-	-	5	-	-	-	-
4. Notices about the course	**21**	20	1	-	14	-	3	2	2
5. Operational problems	**19**	9	10	-	3	12	-	4	-
6. Running of the course	**56**	8	48	-	12	2	12	12	18
7. Course content	**2**	-	2	-	2	-	-	-	-
TOTAL	**134**	**61**	**73**	**12**	**50**	**14**	**18**	**20**	**20**

For a more detailed analysis of the Discussion Group, we classified the messages according to their purpose, showing the role of the sender and the categories used, as it can be seen in Table 3. We can note that the Notice category was used for practically all purposes. This indicates that it is too generic and that is necessary to separate it into categories that are more specific. Its concept was also overlapped with other categories, like the Monograph. For example, when the instructor published the grades for the group's monographs, one time he used the category Monograph and other he used the Notice. Both uses may be considered correct, since the publishing of grades is a notice but also is related to the course monographs. The fact of having two correct categories for the same message confuses the author at the moment of choosing a category and the participants when looking for a message.

6. Improvements on the environment

Based upon the difficulties that were observed and reported by the learners, we will now describe some proposals for improving the AulaNet environment to minimize the problems encountered.

A difficulty encountered by the learners was to know the purpose and the form of use of the categories. In order to minimize this difficulty, the environment now shows, with the name of the category, a brief description previously supplied by the teacher explaining its purpose.

In order to guide learners who did not properly choose or use the categorization, now the teacher is able to comment on the message, giving instructions and explanations to both the author and the group.

It was noted that some categories could be used only in given phases of the course, as it is the case of the Presentation category during the first weeks. Despite this fact, they were available the entire time allowing a wrong choice. In order to reduce the number of categories available to participants when they have to choose one, the teacher can deactivate and reactivate the categories at any time. Upon deactivating a category, it continues to appear in the reports but it becomes no longer available for new messages.

Another way of limiting the number of categories available to the participants at the moment of choice is the teacher indicating what are the valid sequences [12]. For example, the Argumentation category is used only in response to a Question and, consequently, it only needs to be available to respond to a message in this category.

Other functionalities suggested are the possibility of changing the position of the Interest Group messages, the use of icons to identify the category and the possibility of the author to correct the categories of his messages.

7. Conclusions

The success of the application of message categorization depends in part on the set of categories adopted. This set must be defined by estimating the objectives of the discussion and of the messages, as well as the behavior of the participants. Once the initial set has been defined, it should be refined, observing each category's use and the purposes of the messages in the generic category. Even if a satisfactory set is obtained for one group, there is no guarantee that it will work for other groups, so the teacher should pay attention to adequate it. Upon the definition of a set of categories, the teacher should avoid those that are ambiguous or with similar meanings. This kind of category could confuse the

participants when categorizing their messages and skew the results of the report tool, since the messages that should be grouped together will, in fact, be dispersed. The set of categories could be based in argumentation and communication models, interaction patterns and structured ethnographic analysis. The set adopted in the ITAE course was based in the communication, coordination and cooperation model [13], and in the IBIS model [10].

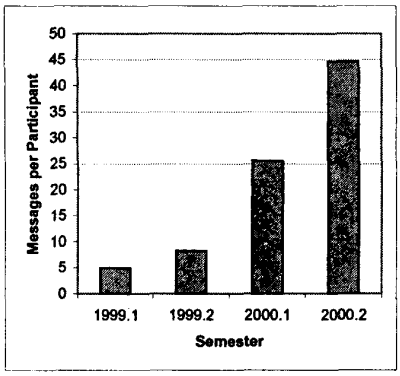

Figure 3. Chart showing the increase of the number of messages per participant

The introduction of categorization and the structuring of messages in an Internet course facilitated and organized the discussion, which is a fundamental factor in active and numerous communities. The discussion flowed better, with greater learner participation. As observed in Figure 3, there was an increase in the number of messages in 2000.1, when it was adopted the message categorization, and mainly in 2000.2, when it was associated with the message structuring. Moreover, according to the instructors, the quality of the contributions increased and the discussion's focus was maintained due to the self-discipline imposed by the use of categories. This indicates that the message categorization and structuring caters for discussion of the topics in death, leading in a gain in the learner's work and in the learning process.

The message categorization also supplied subsidies for the automatic classification and grouping of the messages through reports that the instructors used to follow-up learner participation and to understand how the discussion is progressing, identifying its central elements. Finally, the message categorization helped the reduction of the information overload for the participants, since it supplied complementary information that helped the identification of the content and structure of the discussion without read the messages, which is fundamental in numerous and active groups.

From this experience with categorization in a learning environment, we are trying to obtain a general guideline to create an initial set of categories, find a compromise between a fluid discussion and a structured one, and obtain findings that can be extrapolated to other asynchronous environments, synchronous tools, and environments designed to support collaborative work.

8. Acknowledgements

The AulaNet project is partially financed by the Fundação Padre Leonel Franca, by the Ministry of Science and Technology through its Program of Excellence Nuclei (PRONEX) grant nº 76.97.1029.00 (3366) and also through individual grants awarded by the National Research Council to: Carlos José Pereira de Lucena nº 300031/92-0 and Hugo Fuks nº 524557/96-9. Marco Aurélio Gerosa received a grant from Council for Improvements in the Teaching of Higher Education.

9. References

[1] H. Fuks, "Groupware Technologies for Education in AulaNet", *Computer Applications in Engineering Education*, Wiley-InterScience, NY, 2000, pp 170-177.

[2] S.R. Fussel, R.E. Kraut, F.J. Lerch, W.L. Scherlis, M.M. Mcnally & J.J. Cadizz, "Coordination, Overload and Team Performance: Effects of Team Communication Strategies", *Proceedings of CSCW*, 1998, pp 275-284.

[3] Selvin et al. "Compendium: Making Meeting into Knowledge Events", *Knowledge Technologies 2001*, Austin.

[4] D.A. Schön, *The reflective practitioner: How professionals think in action*, Basic Books, New York, EUA, 1983.

[5] S. Buckingham & N. Hammond, "Argumentation-Based Design Rationale: What Use at What Cost?", *International Journal of Human-Computer Studies*, 1994, pp. 603-652.

[6] S.J.B. Shum, A. MacLean, V.M.E. Bellotti & N.V. Hammond, "Graphical Argumentation and Design Cognition", *Human-Computer Interaction*, 1996.

[7] S.J.B. Shum, "Negotiating the Construction and Reconstruction of organizational Memories", *Journal of Universal Computer Science*, 2 (8), 1997, pp. 899-928.

[8] K. Delvin & D. Rosenberg, *Language at work: analyzing communication breakdown to inform system design (Csli lecture notes)*, CSLI Publications, 1996.

[9] H. Fuks, C. Laufer, R. Choren & M. Blois, "Communication, coordination and cooperation in distance education", *proceedings of the 1999 Americas Conference on Information Systems*, Milwaukee, USA, pp 130-132.

[10] J. Conklin & M. Begeman, "gIBIS: A hypertext tool for exploratory policy discussion", *Conference on Computer-Supported Cooperative Work*, Oregon, 1988, pp 140-152.

[11] K. Becker & A.P.T. Bacelo, "The Evaluation of GRADD: a GDSS supporting asynchronous and distributed meetings", *proceedings of sixth International Workshop on Groupware*, CRIWG 2000, Madeira, Portugal. pp. 19-26

[12] C. Laufer & H. Fuks, "ACCORD: Conversation Clichés for Cooperation", *proceedings of the international workshop on the design of cooperative systems*, pp 351-369.

[13] H. Fuks, M.A. Gerosa & C.J.P. Lucena, "The Development and Application of Distance Learning Courses on the Internet", *The Journal of Open and Distance Learning*, February 2002.