Group learning through Internet

Liane Tarouco University Federal of Rio Grande do Sul Graduate Program Informatics on Education Porto Alegre – RS - BRAZIL Liane@penta.ufrgs.br	Antonio Rodrigo de Vit University Federal of Rio Grande do Sul Graduate Program on Computer Science Porto Alegre – RS - BRAZIL arodrigo@inf.ufrgs.br

Abstract: Internet is being intensively used for distance education. Normal services like email, chat and videoconference are not enough to fully support for properly handle interactions without extensive non-automated work. To support group work related to learning activities it is necessary develop applications to handle participants contributions and consolidate contributions providing summaries of the discussion. At the same time it must be integrated with group decision support tools to help students to build results from what was discussed. This paper presents a leaning environment developed to support through Internet for distance education.

1. Introduction

Computer-mediated distance learning (CMDL) is an uncommonly bright star on the horizon of higher education innovation. According to a recent study by CCA and Associates, 30 percent of colleges and universities have at least one distance learning program in place and an additional 28 percent are planning such programs. (Tucker 1998)

As Internet use in distance education grows exponential nowadays, services used at the beginning was only the ones having strong relation with ancient forms of distance education. Distance education started using post services and even those days, many courses use in wide scale Post Office distribution of printed material. Other mass communication media, as television and radio or recorded tapes with video or audio where also very usual. Some Internet services started to be used in replacement or complementing this kind of services technologies used are shown in table 1.

Ancient forms of distance education	**Internet based environment**
Printed text books	HTML pages
Exercises	Forms and CGI
TV broadcast	MBONE
VHS tapes	Real Video, MPEG
Audio tapes	Real audio
Assesment	Forms and CGI

Table 1: Distance education services

Internet utilization adds more possibilities, for news forms of interaction like is shown in figure 1.

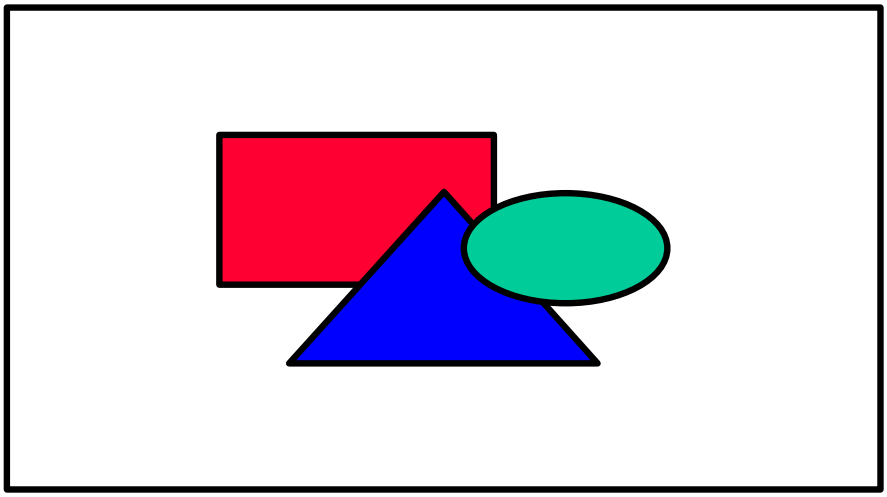

Figure 1: Interacting through Internet

The use of the Internet for distance education derived some advantages but one of the main is the support for human-to-human interaction that is as essential activity in learning. But long experience on using showed a need of a better level of automation on handling of group contribution. Huge number of contributions are some times just relegated to a historical register kept in log files. So the group activity may give opportunity for student to present contributions that some times are relegated to forgetfulness.

To support more good use of contributions derived from group work, related to learning activities, it is necessary the development of applications to handle participants contributions in such a way to provide support for group decision and utilization of contributions that came up during group activity. Collaboration strategies are requested to improve the information quality provided to support decisions. In distance education it shows up as an educational innovation.

For all the bells and whistles of groupware (Bannon 1992) , e-mail continues to be the most widely used groupware component So one can count on all the automated work provided by email or news service servers to receive, store or/and properly forward the messages. But no matter if only **email** or **news** service is used for group communication, as well as other forms of interaction, like chat combined with videoconference or with other multimedia environment (as The Palace or Virtual Reality based environments), the results from group activity are huge amounts of text derived of participants contribution. When it comes to the consolidation phase of the work, usually it is needed a lot of non-automated task to extract main ideas from the many ones presented almost randomly during the discussion. It is also needed a voting system to support group decision about the consolidated results from discussion (ordering, reordering, selecting, discarding etc).

This work describes strategies and tools designed to accomplish automatic ideas grouping (with focus on common themes) for organization of contributions resulted from steps in a group decision. A set of tools was designed and developed for to support distance education cooperative work.

2. Group tool

The work start on the supposition that group activity results on a log file containing interchanged texts (synchronous or asynchronously) during a group debate. These inputs are submitted set of routines that group ideas automatically. The system uses a WWW based interface and log files derived from the group interchange of ideas. The actual results where built from existing previous works at UFRGS.

First considered work to give a start to present project was a tool denominated "Issue Analyzer" that provides natural language recognition. It also contains an algorithm for calculation of similarity between contributions of participants from a group discussion. This prototype was implemented on Sun stations and the software used RPC (Remote Procedure Protocol) for intercommunication between a server and client applications. It provided the grouping of similar ideas generated in the phase of brainstorming and organized them according with definitions of the group.

Another previously developed work, named Eurekha, showed the importance of information retrieval techniques, to support textual information organization. It pointed the usefulness of this kind of support in an

environment with information overload like Internet. This solution uses methods for grouping textual objects. Objects are organized automatically in similar objects groups, facilitating their location, manipulation and analysis.

Both systems are directed to natural language (in written form) processing. In natural language processing, one of the more important aspects is the formal language representation. In this work it was used Discourse Representation Theory. This kind of system faces big challenges: linguistic aspects of natural language representation and handling of problems like ambiguity, references to names, pronouns, etc. So another aspect present in this solution was the grammatical classification of sentences.

The present work aimed a result in terms of group decision support system tool that is able to handle natural language input.

2.1. Proposed Model

The designed system used existing results derived from previous works, referred above. It starts with am Module I that handle log files (derived from chat sessions using software like CuSeeMe, The Palace and similar ones). Files are split up in smaller units (files with only one sentence or contribution). Those smaller units are submitted to a routine that performs similarity calculation. This module, process each sentence resulting in properly labeled version according with a Syntactic Derivation Tree (ADS). Stop words are removed from sentences because they do not add significance. It is possible to customize the system adding or removing stop words to the dictionaries. Figure 2 shows types of stop words (adverbs, pronouns, cardinals, prepositions etc)

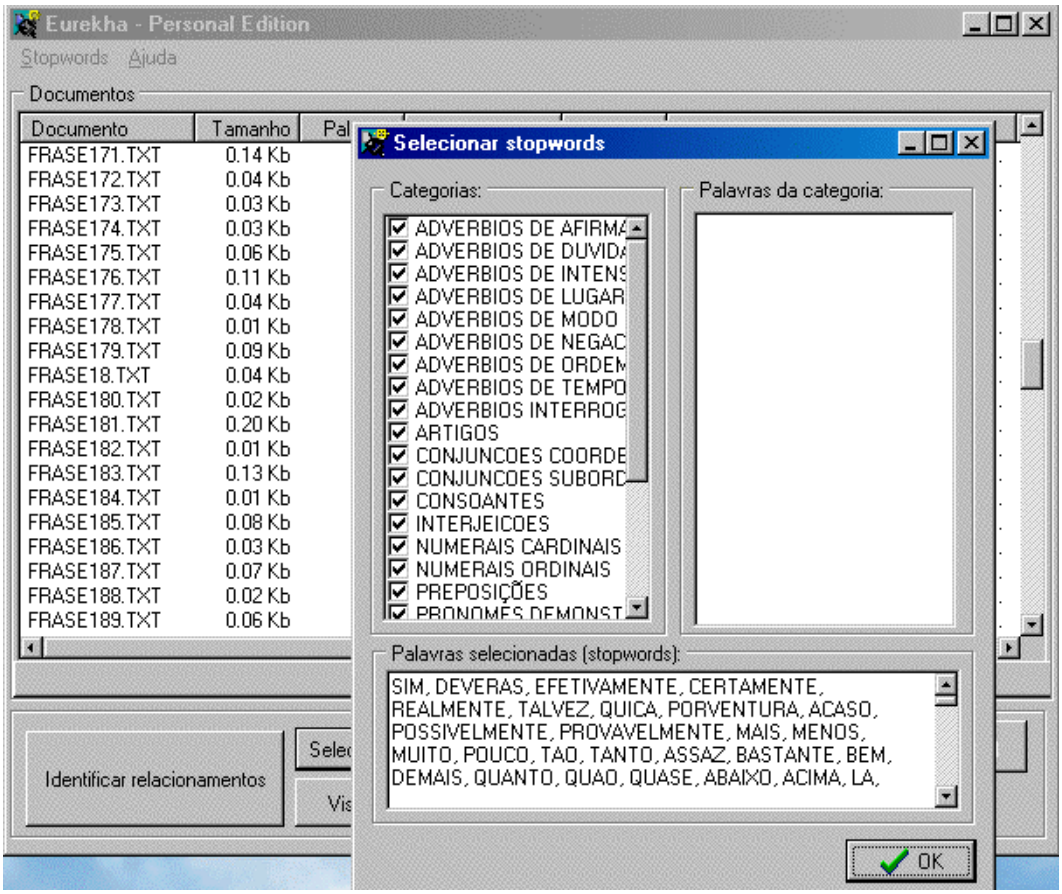

Figure 2: Editing stop words

Module II (previous work named Eurekha) handle sentences grouping them in cluster according similarity. In this step pre-processing of natural language is done. It is considered that the documents have already been corrected with the use of an orthographic checker (task also performed by the module I). Besides, is interesting that the texts have suffered some normalization of terms. However, these procedures are not conditionings Figure 3 show na example of results using Eurekha.

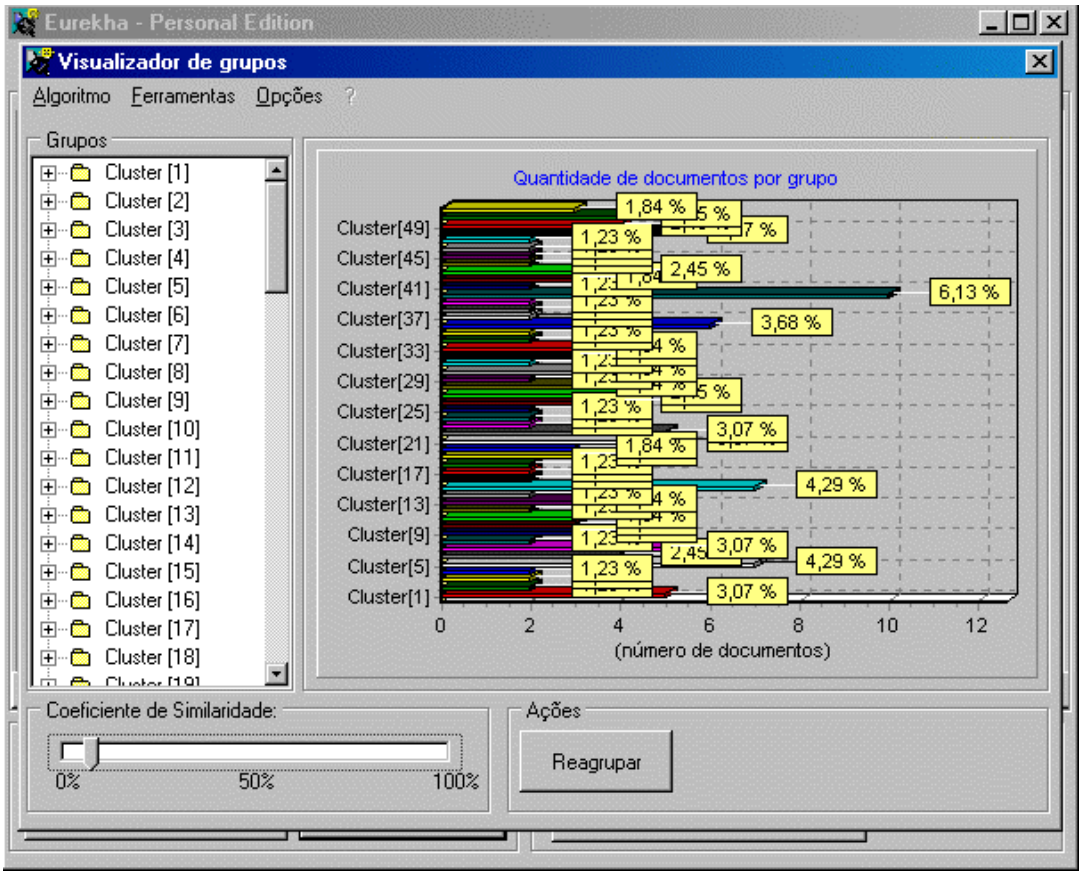

Figure 3: Output of Eurekha.

It is important to play with similarity coefficient to get better results. If one chooses a very high coefficient of similarity will result on high number of cluster with few sentences. If one choose a very low coefficient of similarity will result on few cluster but full of non related sentences. To handle files with student contributions we find out that 15% is a good number for the desired coefficient of similarity.

Other modules, built during present development of the project (Module III to V) handle the resulting cluster of sentences, identify the most frequent/relevant words that will be used to compose a set that represents each cluster. Module III performs this task.

The module IV index words and links them to original sentences for further information retrieval. So one can wants to know which sentences provided the most words resulting as representative for clusters (and of course who created that contributions.

Each group of words became an alternative in a voting process. The voting results in a hierarchy of representative ideas by the group. Module V implements this voting process. An HTML page with the results of this stage of the discussion is presented to the participants of the group to subsidize one more discussion round.

2.2 Results

A prototype version of the system is running. It provides all the funcions described above:

- clustering of sentences
- generation of consolidated idea for each cluster
- integration with a voting system to allow participant to weight and vote consolidated ideas

The results of this process can be used as starting of a new step of the group work where new discussions will use the results of previous consolidated work. This cycle can be repeated as many times as desired.

The tool will be used not only to support building conclusions from consolidated sentences but as it also

keeps the original sentences and their originators, it is possible to derive conclusion about productivity of each participant. Not only the amount of contributions can be counted but it is also possible to identify contributions that originated consolidated clusters selected by participants as more relevant and representative of the discussion conclusions. So it is possible to identify those students whose contributions are more significant, being more often included in the clusters more accepted by the group.

3. Conclusion

Tools described in this work where designed to support group activity in distance learning because it is believed that cooperation and collaboration are key components in distance education. Natural language processing was needed to handle contributions of group activity resulting in clustering of sentences and summarizing cluster to came up with representative ideas of each cluster that are submitted to a voting process and resulting ordered representative sentences are the essence of the group activity. These results may be partial being used as input for a new cycle of discussions.

4. References

Bannon, Liam. 1992 Schmidt, Kjeld. **CSCW: Four characters in serach of a context**. In: Readings in Groupware and Computer-Supported Cooperative Work: assisting human-human collaboration. San Francisco: Morgan Kauffmann Publishers, Inc.pp 50-56

Tucker. R. 1998 **Assessing the Virtual Classrooms: A Progress Report**. http://www.intered.com/edv5n2.htm