Creating discussion threads graphs with Anagora

Emmanuel Giguet, Nadine Lucas GREYC Cnrs, Université de Caen Basse-Normandie France, ENSICAEN
Emmanuel.Giguet @info.unicaen.fr, Nadine.Lucas@info.unicaen.fr

Abstract: We present Anagora, a graphic tool tracing discussion threads along a time axis. Anagora displays overlapping discussion threads over time on a single screen. Its special feature is to calculate the best resolution for a forum to fit on a screen by choosing the most appropriate time scale. Anagora is used to generate views of fora or forum thumbnails. Several discussion fora coming from e-learning platforms illustrate how Anagora is used by tutors and moderators to monitor students' collaborative work.

Introduction

In the framework of the Calico[1] project, a French collaborative research project joining researchers, teachers and teacher trainers, a number of tools were designed to study and monitor computer mediated communication. Bulletin boards transcripts, also called fora or computer conferences, can be loaded on the Calico platform and processed by quantitative and qualitative tools. Among them is Anagora, tracing discussion threads along a time axis. A number of solutions are proposed in literature to visualize discussion threads, social interactions and sometimes provide link to content (see Dimitracopoulou *et al.*, 2005 for a review, and Bratitsis & Dimitracopoulou, 2008 for methodological issues).

Discussion threads are convenient to assess activity on an educational forum and are used to monitor them (e. g. Gerosa, 2005; Chen & Vassileva, 2006). Beyond the educational settings, researchers tried to provide a view of discussion threads with scalable tools, to cope with very large Usenet fora (Turner *et al.*, 2005), or to display discussion images on small screens (Engdahl *et al.*, 2005). Some researchers devised ways to cope with hierarchical, multilevel contexts in education (Enriquez, 2007). Taking the time axis in consideration to show the overall activity of a community by showing concomitant active threads was also tempted (e.g. Huynh Kim Bang & Bruillard, 2005).

Anagora provides bar graphs to visualize overlapping discussion threads over time on a single screen. Its special feature is to calculate the best resolution for a forum to fit on a screen by choosing the most appropriate time scale (corresponding to days, decades, months or more) according to data. It thus provides a kind of thumbnail image, which is used to represent a forum on the Calico platform in reduced format. The full-size image is also used to monitor discussions in computer-mediated communication. Anagora is also fit to deal with multilingual data through Unicode. It effectively handles French, Greek, Vietnamese and English language.

In this paper we present this visualisation tool and show some examples of Anagora use to explain how it helps tutors in interpreting small and large educational discussion fora, in distance education.

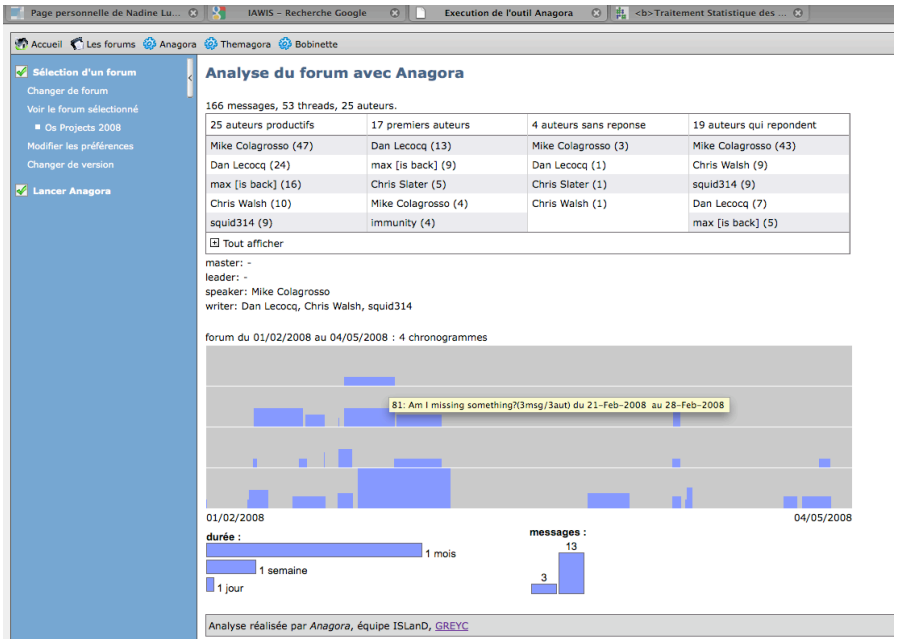

Figure 1. The Calico platform interface with name of thread activated

Discussion threads and chronograms

Anagora highlights high activity in a forum, through discussion overlap. A discussion thread (on the same topic) is shown as a red block, horizontally spreading according to its duration, and vertically spreading according to its number of messages. When dragging the mouse on a block, the title of the thread appears along with dates as shown in figure 1 showing a general view of the French interface. Discussion threads are displayed on rows, while they are sequential. The first row is placed at the bottom of the screen, simultaneous discussion threads are placed above. One row depicting temporally distinct threads is called a chronogram. There are as many chronograms as overlapping discussions within a given time frame. The vertical axis is then called the chronogram axis. In figure 2 there are at most 5 ongoing discussion threads at the same time, during the first month and then the third one. The time scale indicates that the minimal unit in this case is the day.

Graph scalability

The need to address scalability came with textual analysis of fora, which can vary to large extents in duration, number of threads, number of participants and number of posts (Lucas & Giguet, 2008). The same concern prevailed when designing Anagora.

To draw Figure 2, discussion threads including at least 3 messages are represented, otherwise, they are omitted for the sake of clarity. This threshold was chosen because the canonical exchange pattern is made of three posts: a question, an answer and an acknowledgment (thanks or OK message). The actual number of threads is indicated above the graphs with some other quantitative metadata. Depending on the original configuration of the forum, discussion thread metadata are used when available, else, in case of flat list fora, replies with the same message title are considered as forming a thread.

The scale is calculated to fit in the screen, so that, as will be seen in examples, a short forum over a month or so will be drawn in the same screen window, thus it will look expanded as compared to Figure 1, while a forum spreading on years will seem shrunk.

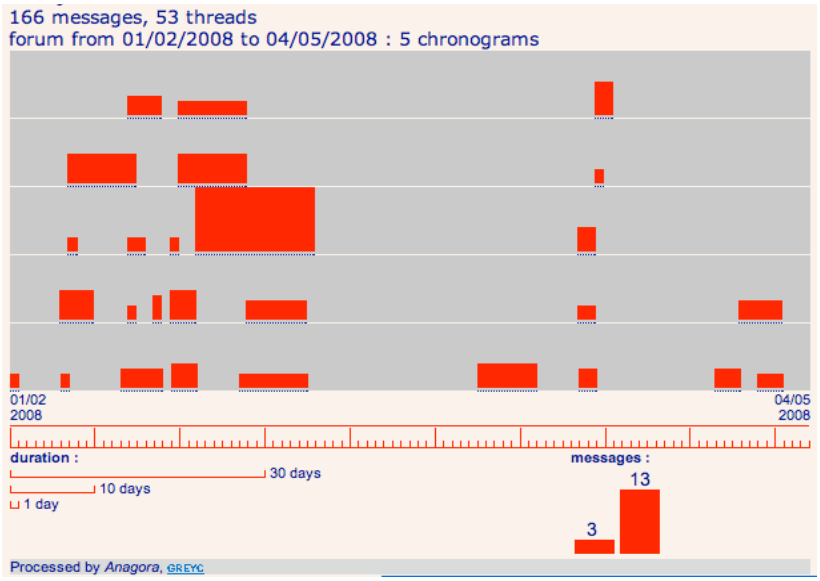

Figure 2. Chronograms for a discussion forum over 3 months (OS projects 08)

Interpretation of users' behaviour

Interpretation varies with the nature of tasks, number of participants, expected behaviour, time allocated for each task etc. We report on two experiments in entirely distant education, using Anagora along with other tools.

Small group collaboration

Anagora was used to monitor collaborative activity in small educational fora, with three students cooperating for an assigned task for about one month. Two groups are compared for two tasks each.

The first group, called DUTBM, exhibits a fairly typical behaviour in collaboration (Fig. 3 and 4). For the first task, discussion threads tend to overlap at the start of the forum, with many topics being discussed at the same time, thus creating 4 chronograms. This pattern is common when students share ideas on work packages. Later, students start working on their allotted part and activity decreases in the forum. Three overlapping discussions are seen in the final period before the assignment is sent to the tutor.

For the second task, this group encounters more difficulties at the beginning and discussions pile up for nearly a month, before consensus is reached (Fig. 4). Interactions are short thereafter and only brief threads (not drawn) are needed before the assignment is sent to the tutor.

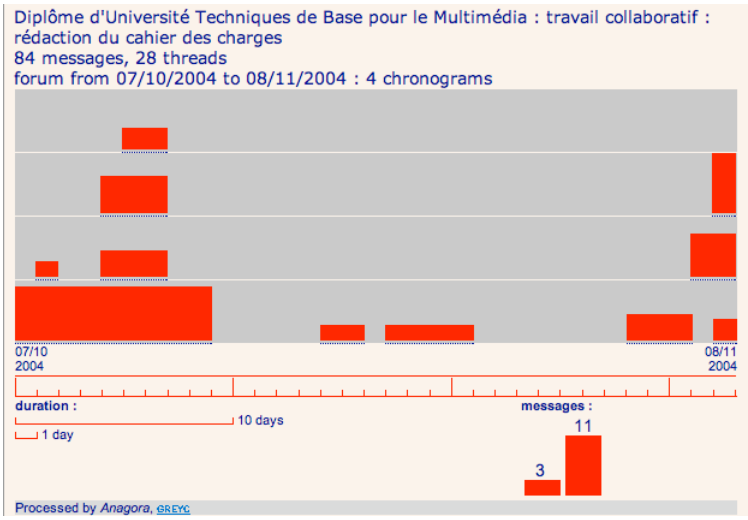

Figure 3. Chronograms for a small group (DUTBM task 1) with typical peaks of simultaneous discussions at start and before the end of the forum

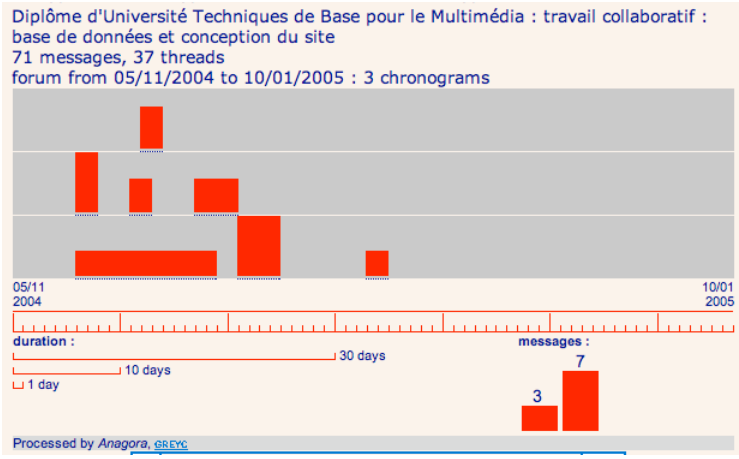

Figure 4. Chronograms for a small group (DUTBM task 2) with simultaneous discussions at the beginning

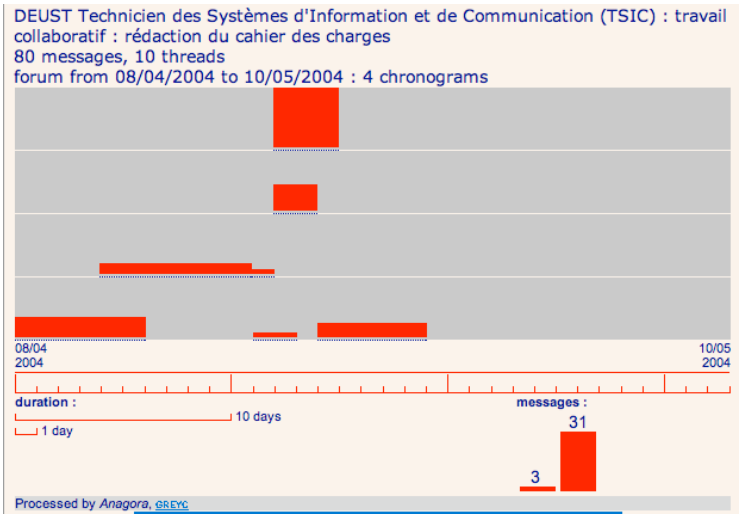

Figure 5. Chronograms for a small group (DEUST task 1) with simultaneous discussions in the middle

A different behaviour can be seen from the simultaneous discussions going on in another group called DEUST (Fig. 5 and 6). For task 1, the number of chronograms is four, same as for the first group. But these different discussion threads start in the middle of the task, hinting that tuning between participants has not been successful. For task 2 this pattern is reinforced with as many as 8 chronograms shown in the middle of the task. Dissension occurs in this case. Interpretation is backed by access to the discussion thread.

Moderators appreciate the juxtaposition of figures 3 and 5 and 4 and 6 (for the same task) to judge the distribution of chronograms for different groups.

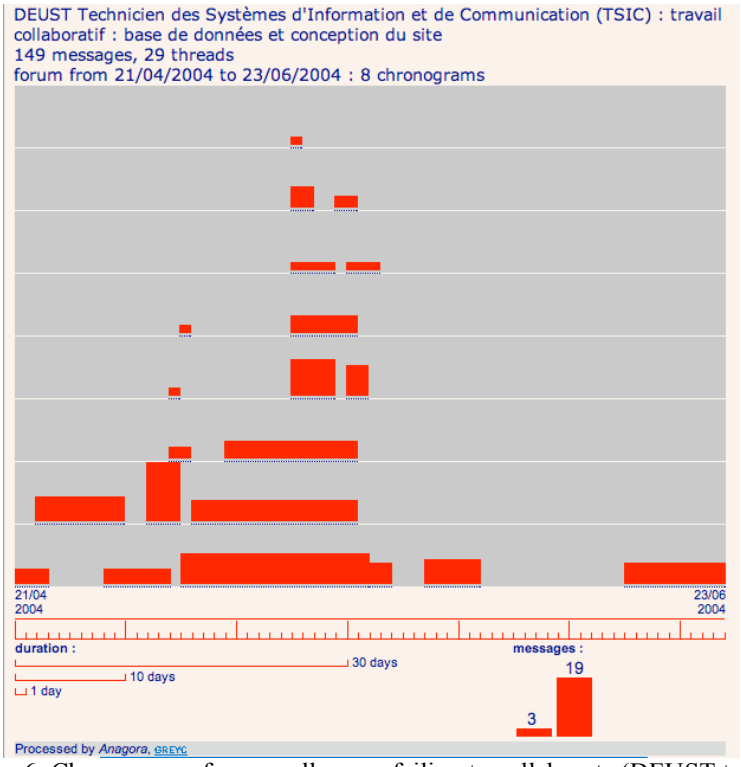

Figure 6. Chronograms for a small group failing to collaborate (DEUST task 2)

Long term collaboration

In a different experiment, a group of student in distance education was followed over three years, the average time for them to complete a standard two-year course. Participation in discussion forum for each curriculum was not compulsory, but was active. In figure 7, the scale is different from the short discussions just explained above. The minimal unit of time here is one month (30 days) instead of 1 day.

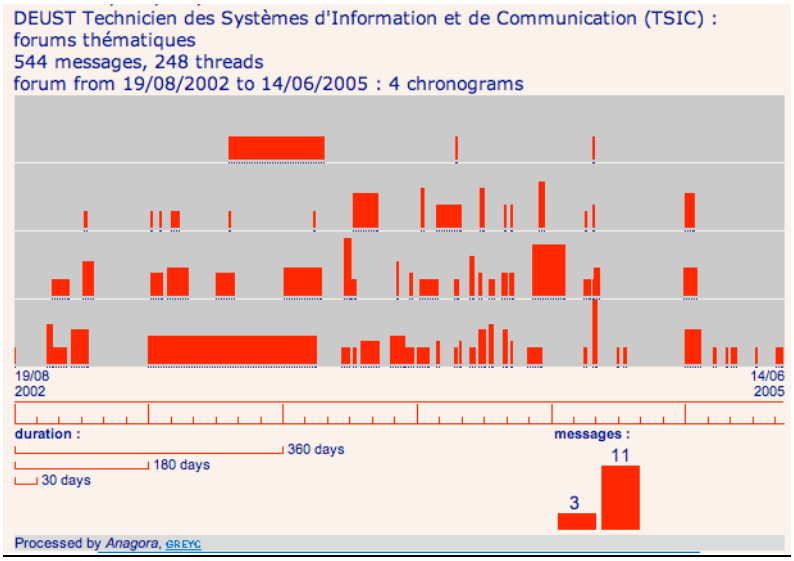

Figure 7. Chronograms for a 3 year period (DEUST)

In this case, the number of simultaneous threads reached four after eight months. The length of threads after roughly six months is fairly long, which is explained in some instances by recurrent complains about course organization. The length of threads tends to decrease in the second year, while the number of exchange per thread increases (one can see many tall and thin threads). In the last year, three simultaneous threads are seen, corresponding to partners joining into project taskforces.

Discussion and perspectives

Visualization tools for educational fora become more and more sophisticated (Mazza & Dimitrova, 2007). Related work in the domain of adaptive scalability for quantitative analysis of educational fora is recent. May *et al.* (2007) work on log traces of students' activity and address the same problem of meaningful units. Cress (2008) offers an elaborate mathematical approach to deal with levels and scale.

While Anagora is much less sophisticated, it is also fairly easy to use. Users in the Calico group generally liked the Anagora view of fora: this representation allows display of several groups at a time, allowing comparisons of group progress in computer based education. They also used these compact representations as thumbnails to represent the files on the Calico platform. This platform can be accessed at http://www.crashdump.net/calico/, and Anagora can be tested on external data as well.

However, some improvements are needed. The number of chronograms is decided with a fixed threshold of 3, but this value should also be calculated by the program, according to length and number of participants in the forum. The thumbnail effect should be applied both on the time axis and on the chronogram axis.

Anagora provides the best resolution for a forum to be seen on a single screen. Alternatively, visualisation for comparison of (images of) discussion threads in a fixed span of time could be provided. This would amount to give a constant ratio for geometric representations and duration, like in cartography for space. For instance, in the present state in figure 3, a thread spreading on ten days has the same dimension as a thread spreading on 28 days in figure 4 and this could be misleading. For tutors who manage fairly fixed time allotments in course management, it would be useful to compare the progress of different groups by keeping geometrical dimensions anchored on a constant representation of time. Last, interactivity should be provided through hyperlinks allowing thread content visualization and individual messages popping up inside the thread.

Endnotes

(1) Communautés d'apprentissage en ligne, instrumentation, collaboration, supported by the French Ministry of Education, Research and Technology for 2007-2009.

References

Bratitsis, T., & Dimitracopoulou, A. (2008). Interpretation issues in monitoring and analyzing group interactions in asynchronous discussions. International Journal of e-Collaboration, 4(1), 20-40.

Chen, H., & Vassileva, J. (2006). Design and evaluation of an adaptive incentive mechanism for sustained educational online communities. *User Modelling and User-Adapted Interaction*, 16 (2/3), 321-348.

Cress, U. (2008). The need for considering multilevel analysis in CSCL research: An appeal for the use of more advanced statistical methods. *International Journal of Computer-Supported Collaborative Learning*, 3(1), 69–84.

Dimitracopoulou, A., *& al.* (Eds.) (2005). *State of the art on interaction analysis for metacognitive support and diagnosis* European Community Report Kaleidoscope Network of Excellence (No. JEIRP D.31.1.1).

Engdahl, B., Köksal, M. & Marsden, G. (2005). Using treemaps to visualize threaded discussion forums on PDAs. CHI '05 extended abstracts on Human factors in computing systems, (pp. 1355-1358). New York, NY, USA : ACM.

Enriquez, J. G. (2008). Translating networked learning: un-tying relational ties *Journal of Computer Assisted Learning* 24 (2), 1365-2729.

Gerosa, M. A., Pimentel, G. P., Fuks, H. & Lucena, C. (2005). No need to read messages right now: Helping mediators to steer educational forums using statistical and visual information. In T. Koshman, D. Suthers & T-W. Chan (Eds) *Computer Supported Collaborative Learning 2005: The Next Ten Years!* Mahwah, NJ: Lawrence Erlbaum Associates.

Huynh Kim Bang, B., & Bruillard, E. (2005). Vers une nouvelle interface de lecture pour des forums de discussion dédiés à des élaborations collectives. In Saleh I. & Clement T. (Eds.), *Créer, jouer, échanger, Actes H2PTM'05*, (pp. 43-56). Paris: Lavoisier.

Lucas, N. & Giguet, E. (2008). Robust adaptive discourse parsing for e-learning fora. In P. Diaz, Kinshuk, I. Aedo & E. Mora (Eds) *8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008)*, (pp. 730-732). Los Alamitos, CA: IEEE Computer Society.

May, M., George, S. & Prévôt, P. (2007). Tracking, Analyzing, and Visualizing Learners' Activities on Discussion Forum. In V. Uskov (Ed) *Web-based Education (WBE),* (pp. 649-656), Arnheim CA: Acta Press.

Mazza, R. & Dimitrova, V. (2007). A graphical student monitoring tool for supporting instructors in web-based distance courses. *International Journal of Human-Computer Studies*, 65(2), 125-139.

Turner, T., Fisher, D., Smith, M., & Welser, T. (2005). Picturing Usenet: Mapping Computer-Mediated Collective Action. *Journal of Computer-Mediated Communication*, 10(4), article 7.