Monitoring interactions in collaborative learning environments (CSCL): a tool kit for Synergeia

Antonio Calvani*, Antonio Fini*, Giovanni Bonaiuti* and Elvis Mazzoni**

*Università di Firenze; **Università di Bologna

antonio@calvani.it; anto@fininformatica.it

Abstract

Starting from the early 90ies computer supported collaborative learning has gained new interest thanks to the diffusion of networks and social constructivism. In order to carry it into effect, it is possible to make use of generic tools (web forum, space for the uploading of materials, etc.) or of specific environments produced within the researches on CSCL (Computer Supported Collaborative Learning).

While e-learning platforms pay much attention to tracing and readability of data and to interoperability standards, in the more specific field of CSCL critical elements remain unsolved in this regard: even if data are traced, they cannot be always easily read and consequently processed or exported for analysis purposes with other software (statistics, etc.).

Synergeia, a well-known environment for net collaborative learning born of a European project, though allowing the tracing of users' interactions, does not allow the extraction of this information for quantitative analyses or for the exportation outside the platform. This contribution, moving from experiences realized with this software in a university context, shows a specific tool (SIm – Synegeia Log Miner) which has been realized to extract important data for the monitoring and consequent evaluation of net collaborative interactions. The fitting of the specific tool is also the occasion to reflect on the indicators to be selected and on the interoperability perspectives linked to the format of this kind of data, through the identification of a generic data structure suitable for representing interactions in this sort of systems.

Edizioni Erickson – Trento
www.erickson.it

Je-LKS Journal of e-Learning and Knowledge Society - *Vol. 1, n. 1, aprile 2005 (pp. 63-73)*

1. Theoretical context and application

Collaborative learning has been gaining new interest since the early 90ies thanks to the diffusion of networks and social constructivism. For this kind of activity, besides generic tools (web forum, space for the uploading of materials, etc.) specific environments have been produced. They arose within researches on CSCL, *Computer Supported Collaborative Learning* (Koschmann, 1994; 1996; 2002) and nowadays they are enriched by interesting contributions of the artificial intelligence (Jerman et al., 2001).

Among the most known environments there are: Knowledge Forum, which is the evolution of CSILE (Scardamalia, Bereiter, 1989; 1994), one of the first and most known projects, SNS (Jonassen Redimez, 2002), Fle 3 (Lakkala, Rahikainen, Hakkarainen, 2001), Synergheia (Stahl, 2002).

The evaluation of collaborative activity can be dealt both with qualitative and quantitative methods. The first, that is qualitative methods, are generally based on the analysis of interactions (Gunawardena, Lowe and Anderson, 1997; Bocconi, Midoro and Sarti, 1999; Gnisci and Bakeman, 2000), on chat analysis, focusing on the talk turns (Galimberti, 1995; Bonaiuto, 2002), and on intervews with the participants as well (Light, Colbourn and Light, 1997; Ligth et al., 1998; Cacciamani, 2004). These data survey methods, like those based on «mixed» approaches, usually entail considerable burdens of manual labour on behalf of the evaluator, which could be partly reduced if the data were recovered automatically (Martinez et al., 2001; Sha and Van Aalst, 2003).

With regard to quantitative data, we remark that as a rule CSCL systems, as well as e-learning platforms, store a discrete quantity of information on events usually identified by the generic term of «tracing data». In some cases, these data are then made easily accessible; in others, they are held by the system for the only sake of information to users, with no statistical purpose. An example of easy access to tracing data is given by the Knowledge Forum platform (http://knowledgeforum.com). This platform has been endowed with a special tool (Analytic Toolkit, Burtis, 1998), which has been implemented to make readable, through a Web-based interface, a wide range of quantitative data, obtainable from the platform data base.

To support the activities of a post-lauream specialization course in the Laboratory of Education Technologies of the University of Florence, we adopted the collaborative environment Synergeia in the 2003-2004 academic year.

Synergeia is a client/server software developed starting from a groupware instrument, the BSCW system, of which it is an extension. It was born from the project ITCOLE (Innovative Technologies for Collaborative Learning and Knowledge Building) under the financing of the European Commission for IST (IST-00-III.2 'School of Tomorrow'). The main characteristic of Synergeia is its capacity of

supplying a group of people engaged in cooperative activities with tools to build and share informative resources. The adaptation to training needs of a tool for the development of productivity in the working field is realized in particular through the enriching of the discussion tools (knowledge building areas), with the addition of elements for the typifying (scaffold, thinking type) of messages. In order for the community of users to exchange useful information for the management of the resources produced, any other object of Synergeia (like folders, documents and links to external resources) can be commented and evaluated both by the author and by readers. Every action performed is traced. Unlike what happens in e-learning platforms, in this case the purpose is not to provide the teacher with the information for the evaluation but to give to each player the visibility on the development of actions and on the participation of all in the learning process.

Through proper overviews, users can approach the chronicle of each object and verify the author's name, the date, the time and the name of who has read or used it, etc. It is an useful function to strengthen the idea of an active and busy community.

This information, though regularly recorded, cannot be analysed in a synthetic way because there are not tools able to give quantitative or descriptive data processing. This lack can be found in many similar products, except perhaps Knowledge Forum which has an integrated analytic tool kit. That is why we decided to tackle the problem by developing a system allowing the extraction and standardized analysis of the interactions within collaborative platforms.

2. Data acquisition in Web-based collaborative environments

For the purposes of our research it was important to address two different issues:

1. the detail level of the traced data (what actions are actually traced?);
2. the easy access to tracing data (are data stored in the data base?).

As regards the first point, different detail levels in the tracing can or cannot allow to get minute information on the operations performed. As for the access to data, there are two possibilities: platforms store the tracings by means of relational data bases and/or by means of log files. Log files are simple text files on the web server, made up of a number of «lines» (generally one line for each action carried out).[1]

The use of logs to analyse interactions presents many difficulties, mainly linked to the interpretation of data (Mazzoni, 2003; 2004). The management of data

[1] They are «applicative» log files, different from «generic» log files, managed by any web server which simply stores the users' access to the different site pages. In this case the application log file reports detailed information which are typical of the specific application.

by means of a data base is certainly a better solution, on condition that access is allowed and the structure of data opportunely documented.

Regarding Synergeia, as already stated, the data base including tracing data does not allow the access with external tools and the technical documentation is not available.

3. What are the essential data for the analysis? Technological and methodological considerations

3.1 Technological considerations

From the technical point of view, the first objective for everybody wanting to develop a tool of analysis of wide range of use is the definition of a high standardized data structure containing all the information necessary for the study of interactions, included treatments and representations of interactions of a more complex level, such as those of SNA (Social Network Analysis, Freeman, 1986). An excellent proposal has been advanced by a group of researchers of the University of Valladolid (Spain). It leans to extreme generalization and to the standardization of the concept of «interaction» through the XML representation of the original data belonging to different collaborative situations (Martinez et. al., 2003). One of the possible developments of our work could be the definition of a subscheme to be placed within the structure proposed by Martinez and colleagues.

3.2 Methodological considerations

Since we had to choose in our analysis quantitative indicators, we built our model including five typologies of essential data in order to monitor interactions with the possible implications related to collaborative learning dynamics.

We answered the question «what is useful to survey» by proposing the following categories:

1. *Participation*: number of messages in a given lapse of time (average of the group, standard deviation, individual value, difference between «chatterboxes» and «absentees», etc.);
2. *Production:* number of documents or other products enclosed to messages (absolute numbers, individual differences, etc.);
3. *Reactivity:* times of latency between a message and the other (average times, maximum times, standard deviation, etc.);
4. *Reading* of messages and documents from the group users (per cent values, individual index of read messages/documents);
5. *Structure* (horizontal/vertical) of the web forum: the web forum can extend in depth (underlying degrees of messages with consecutive concatenate answers) or

in width (many parallel answers at the same level). This aspect was discussed in particular by David Wiley (2002) who proposed a synthetic numerical indicator called *Mean Replay Depth*, opportunely adjusted to take account of new threads without reply and defined «d». This value, according to Wiley, could let achieve a rapid indicator of the discussion level in a web forum.

We could add to these typologies of indicators, looking important for any platform, another one concerning a functionality which can be often found in these environments: the use of labels to mark the message, with the aim of developing metaknowledge activity *(Thinking type).*

The utility of information of this kind for an immediate evaluation of what is going on is evident. For instance, it is important to know how much messages and documents are read: it is not necessary that in a good group «everyone writes»; it is natural that new members, especially in the first phases, do not write but limit themselves to read others' texts. These indications can be easily transformed in indexes which, in some way, give the scope of the mutual attention in a group. In collaborative groups there are often phenomena of communicative centralization, which can be immediately signalled by an automatic monitoring. Even a web forum much extended in depth could reveal anomalous dynamics (for example endless «duets» because of formality or acrimony…).

4. The model of data

The proposed model of data is based on four main entities: USERS, that is people registered in the system (teachers, students, tutors); GROUPS, that is unions of participants, typical of collaboration phases; MESSAGES inserted in web forums and DOCUMENTS inserted in shared folders.

MESSAGES and DOCUMENTS are equipped with accessories related to the tracing of reading.

An in-depth study of data is not the object of this paper, yet it is necessary to specify some elements. Firstly, a CSCL environment can comprise many kinds of interactions (chat, whiteboard, surveys, etc.). Our project has focused on two basic aspects of CSCL activities: *production of documents* and *web forums*. This fact, already explained in the methodological considerations, is further highlighted by the data structure proposed. Secondly, we assume that users, participating in CSCL activities, are organized in «working groups». The group is the basic unit for the analysis of interactions and inside it users can play different roles. Thirdly, we thought it right to identify a distinction between mandatory information (M-*mandatory*) and optional information (O-*optional*). Mandatory information are necessary for processing and for the main analyses. For example, the ID of the message being answered (UPPER_MESSAGE_ID) is considered mandatory, while other data are not. It stands to reason that, when the time comes to choose

the collaborative platform, if we are going to use the kit, we will need to verify in advance that it manages at least the mandatory data. Finally, we included in the model some «derived» data, that is obtainable from the processing of available data, even though they were not present in the native format in the data base and/or in the log files. For example, information on threads are deduced by means of algorithms based on the upper message ID.

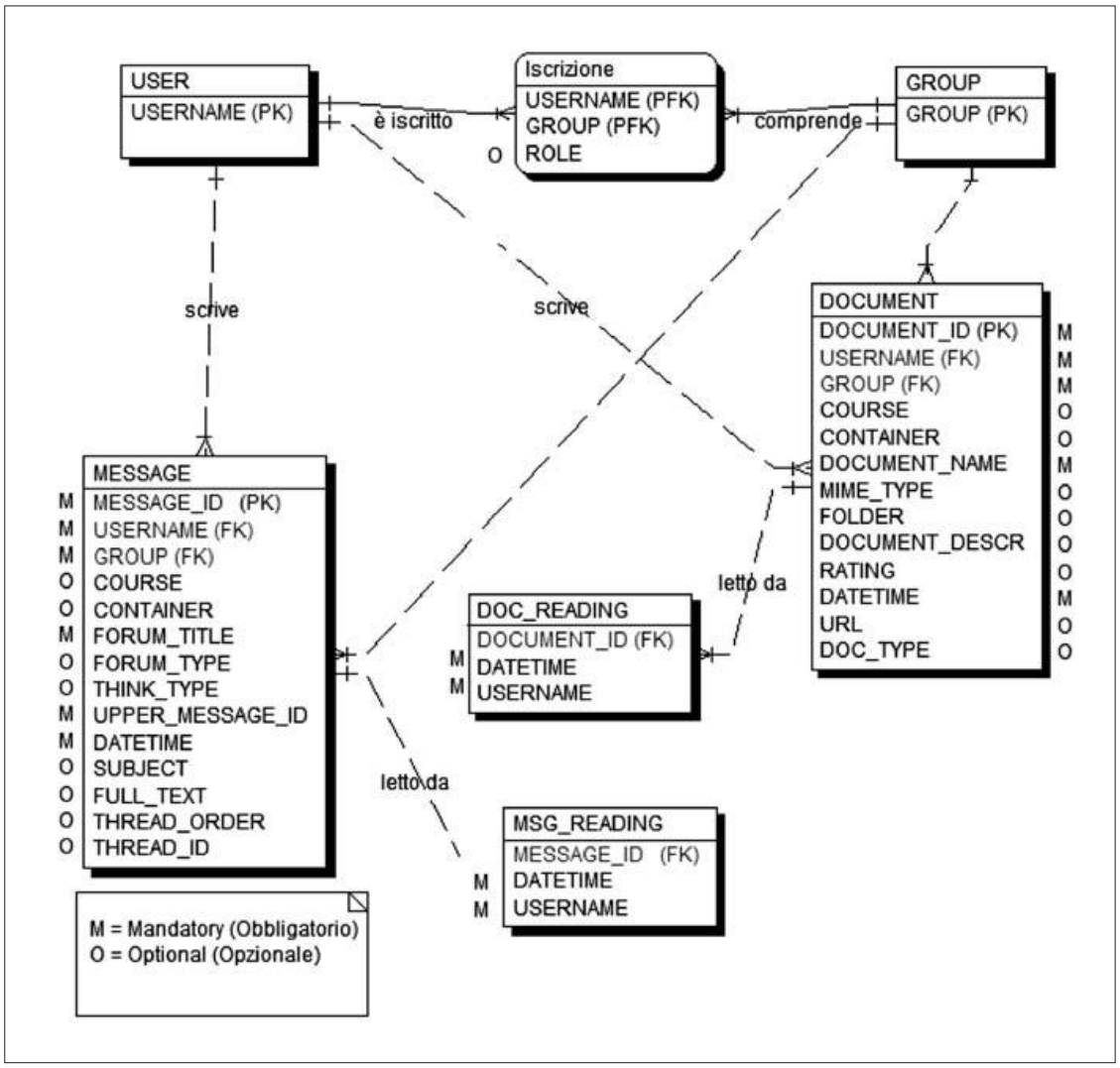

Figure 1 The model of data (ER Entity-Relationship diagram).

5. SIm (Synergeia Log Miner) kit

Synergeia makes available a daily report on the actions carried out. It can be automatically sent to users via e-mail, in addition to a general log file held on the server.

Yet, neither Synergeia nor the system on which it is based (BSCW) offer tools that can be compared to the Analytic Toolkit above mentioned for the analysis of the collected data. Even the direct access to the data base is not allowed, as it is organized in an owner format whose characteristics are not public.[2]

After having analysed the different options, we decided to use the log file to get the necessary information. This file contains specific information on every event (or action) accomplished by the system users. Once we had «decoded» the log file,[3] we managed to write a software module for the acquisition/filtering of data, from the log file to the format set by the model of data described in the previous paragraph. We so realized the pure kit, the module supplying statistics on quantitative data, allowing the easy access to recovered data and giving the possibility of executing filters, sortings and exportation of the selected data onto the electronic paper, so as to permit its use in a wide range of specific software instruments for statistical analysis.

The front-end gives a series of general descriptive statistics, at level of group, forum and user.

The summaries presented include:

- At Group level:
  - number of messages;
  - percentage of read messages;
  - written messages per user (MIN, MAX, AVERAGE, ST.DEV.);
  - use of thinking types;
  - number of written documents;
  - percentage of read documents;
  - written documents per user (MIN, MAX, AVERAGE, ST.DEV.).
- At Forum level (besides the indicators provided at group level):
  - date of the first message;
  - date of the last message;
  - days of activity;
  - answer time of messages (MIN, MAX, AVERAGE, ST.DEV.);
  - number of written messages;
  - maximum depth of threads;
  - value of Mean Reply Depth «d» (Wiley, 2002).
- At User level:
  - number of written and read messages;
  - number of written and read documents;
  - number of groups in which the user has acted.

[2] Synergeia is not an Open Source software but an extension of a commercial software (BSCW) given in free use license to education organizations. The technical specifications, however, are not available for public spreading.

[3] Unfortunately the documentation related to the structure of the log file is not publicly available.

The following figures show some screens taken from the front-end:

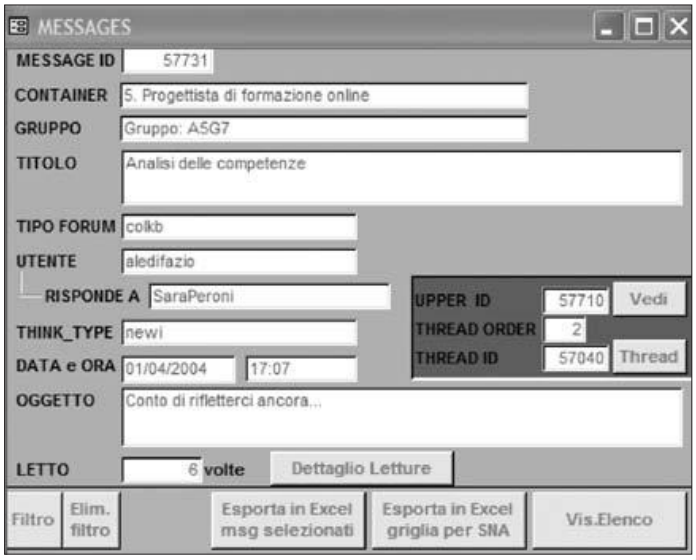

Figure 2 format of MESSAGE table.

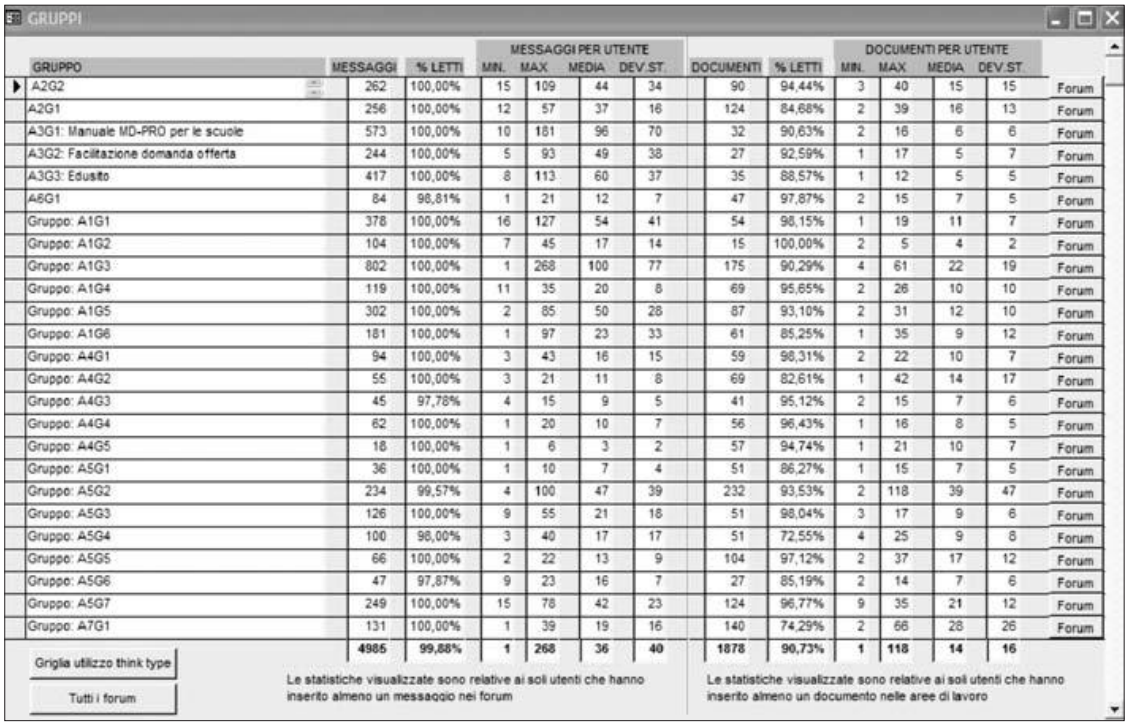

Figure 3 One of the synthetic displays per group.

The first version of the tools above described was realized in a unique software module comprising the data recovery from Synergeia and the front-end named

«SLM – Synergeia Log Miner».[4] The two modules are anyway independent: the section for the recovery of data can be repeated and realized as specific interface for other CSCL platforms, while the section for the front-end will remain exactly alike since it is based on the stated standardized data model.

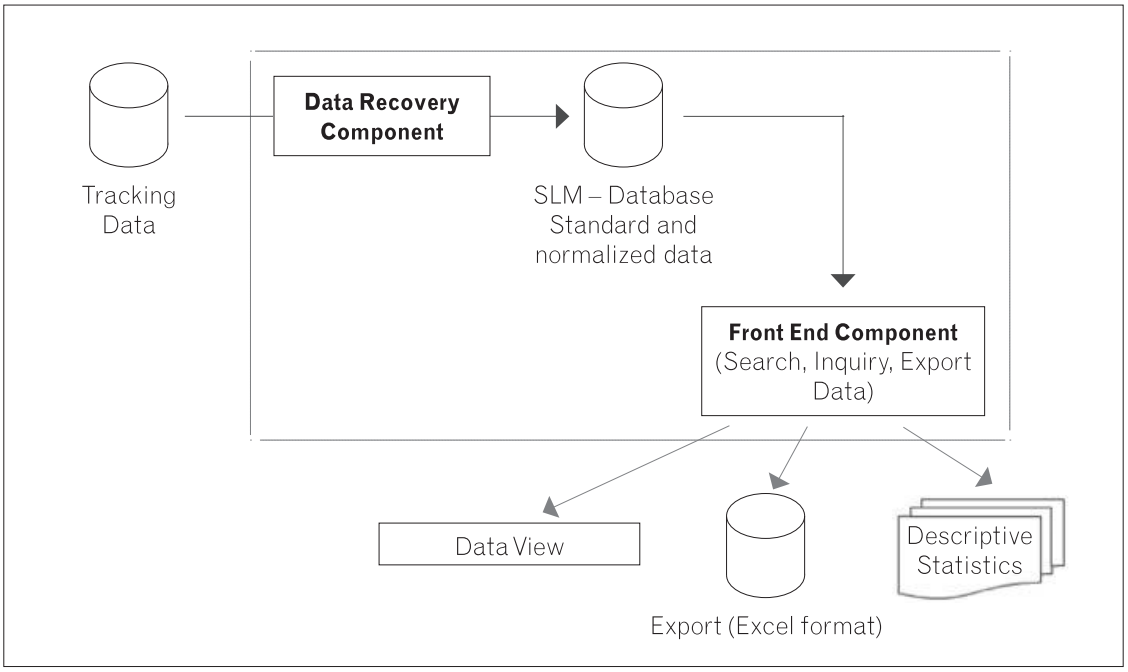

Figure 4 The logic structure of the kit (data mining and front-end).

6. Conclusions

Coordinators, facilitators and net tutors have to meet nowadays news roles required by the new conditions of net interactions. They often have to manage, coordinate and moderate collaborative dynamics and enable groups of people, with different degrees of previous familiarity, to work together, share experiences and integrate their own attitudes in common products.

The activity of university and post-university courses frequently requires to manage simultaneously a variety of collaborative groups. From this point of view the readability of tracing data has been neglected in CSCL environments, though it has been well supervised in e-learning platforms which are more directed toward managing great numbers of users.

We have proposed a model for the reading of significant data, obtainable from tracing data not necessarily available in an «explicit» way, and we have implemented a version of it on Synergia.

[4] The SLM software has been developed, using Microsoft Access™, by Antonio Fini.

We deem that a platform like Synergia deserves to be used more, even in the university context, and to be integrated with instruments enabling tutors or facilitators to have an immediate picture of the interactions.

Future developments should concern different aspects, among which the possibility to realize modules for the recovery of data from other collaborative platforms, the implementation of further statistical indicators, the realization of a totally web based version of the kit. Finally, independently of the possible future implementations of Synergeia, we think that the interoperability development of the obtainable data should be the core of a further study following this first step of working-out.

BIBLIOGRAPHY

Bocconi S., Midoro V. e Sarti L. (1999), *Valutazione della qualità nella formazione in rete,* «TD (Tecnologie Didattiche)», vol. 16, n. 1, pp. 24-40.

Bonaiuto M. (2002), *Conversazioni virtuali.* Milano, Guerini & Associati.

Burtis J. (1998), *The analytic Toolkit,* The Ontario Institute for Studies in Education, The University of Toronto: knowledge Building Research Team

Cacciamani S. (2004), *Elementi per la costruzione di un protocollo di analisi dell'attività di un webforum,* «Form@re», n. 27. Recovered on 15th November 2004 from: http://formare.erickson.it/archivio/maggio_04/1cacciamani.html

Freeman L. C. (1986), *The impact of computer based communication on the social structure of an emerging scientific speciality*, «Social Networks», n. 6, pp. 201-221.

Galimberti C. (1995), *L'analisi delle conversazioni: una pratica di ricerca psicosociale*. In G. Gulotta (a cura di), *La scienza della vita quotidiana,* Milano, Giuffré.

Gnisci A. e Bakeman R. (2000), *L'osservazione e l'analisi sequenziale dell'interazione*, Milano, LED.

Gunawardena C. N., Lowe C. A. e Anderson T. (1997*), Analysis of a Global Online debate an the Development of an Interaction Analysis Model for Examining Social Construction of Knowledge in Computer Conferencing*, «Educational Computing Research», vol. 17, n. 4, pp. 397-431.

Jermann P., Soller A. e Muehlenbrock M. (2001), *From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning.* Proceedings of the First European Conference on Computer-Supported Collaborative Learning, Maastricht, The Netherlands, pp. 324-331.

Jonassen D. e Remidez H. (2002), *Mapping Alternative Discourse Structures onto Computer Conferences.* Paper presented at the CSCL 02, recovered on 15th November 2004 from: http://newmedia.colorado.edu/cscl/12.pdf

Koschmann T. D. (1994), *Toward a Theory of Computer Support for Collaborative Learning,* «The Journal of the Learning Sciences» (special issue), vol.3, n. 3, Lawrence Erlbaum Associates, New York.

Koschmann T. D. (1996), *CSCL: Theory and Practice of an emerging Paradigm,* New Jersey, Lawrence Erlbaum Associates.

Koschmann T. D., Hall R. e Miyake N. (2002), *CSCL 2: Carrying Forward the Conversation,* New Jersey, Lawrence Erlbaum Associates.

Lakkala M., Rahikainen M. e Hakkarainen K. (2001), *Perspective of CSCL in Europe: a review,* recovered on 15th November 2004 from: www.euro-cscl.org/site/itcole/D2_1_review_of_cscl.pdf

Light P., Colbourn C. e Light V. (1997), *Computer mediated tutorial support for conventional university courses*, «Journal of Computer Assisted Learning», n. 13, pp. 228-235.

Light P., Crook C., White S., Nesbitt E., Barrowcliff D. e Light V. (1998), *Learning sites: networked resources and the learning community,* «Journal of Computer Assisted Learning», n. 14, pp. 166-167.

Martinez A., De La Fuente P. e Dimitriadis Y. (2003), *Towards an XML-based model for the representation of collaborative action, in Proceedings of the Computer Support for Collaborative Learning,* CSCL, Bergen, 14-18 junio, June - 2003, recovered on 15th November 2004 from: http://ulises.tel.uva.es/uploaded_files/CSCL03_camerarready.pdf

Martinez A., Dimitriadis Y., Rubia B., Gómez E. e de la Fuente E. (2003), *Combining Qualitative Evaluation and Social Network Analysis for the Study of Classroom Social Interactions,* «Computers & Education», n. 41, pp. 353-368.

Mazzoni E. (2003*), Studenti nella rete: il Dire, il Fare e il Pensare dei comportamenti on-line,* PhD thesis (not published), University of Bologna.

Mazzoni E. (2004), *Strumenti per un approccio quantitativo allo studio delle interazioni. Il software Net Miner e i Log File,* «Form@re», n. 27, recovered on 15th November 2004 from: http://formare.erickson.it/archivio/maggio_04/5mazzoni.html

Scardamalia M. e Bereiter C. (1989), *Intentional Learning as a Goal of Instruction.* In L. B. Resnick (a cura di), *Knowing Learning and Instruction, Hillsdale, New Jersey, Lawrence Erlbaum Associated inc.*

Scardamalia M. e Bereiter C. (1994), *Computer Support for Knowledge-Buiding Communities,* «The Journal of the Learning Sciences», vol. 3, n. 3, New York, Lawrence Erlbaum Associates.

Sha L. e van Aalst J. (2003), *An Application of Social Network Analysis to Knowledge Building.* Paper presented at the structured poster symposium Probing individual, social, and cultural aspects of knowledge building at the annual meeting of the American Educational Research Association, Chicago, April 21-25, 2003.

Stahl G. (2002), *CSCL goes to school.* Paper presented at the CRIWG '02, La Serena, Chile, recovered on 15th November 2004 from: http://www.cis.drexel.edu/faculty/gerry/cscl/papers/ch11.pdf

Wiley D. (2002), *A proposed Measure of Discussion Activity in Threaded discussion Spaces.* Paper not published, Recovered on 15th November 2004 from: http://wiley.ed.usu.edu/docs/discussion09.pdf