HAL
open science

A closer look at tracking human and computer interactions in web-based communications

Madeth May, Sébastien George, Patrick Prévôt

To cite this version:

Madeth May, Sébastien George, Patrick Prévôt. A closer look at tracking human and computer interactions in web-based communications. International Journal of Interactive Technology and Smart Education, 2008, Vol. 5 (Issue 3), p. 170-188. hal-00345591

HAL Id: hal-00345591

https://hal.science/hal-00345591v1

Submitted on 9 Dec 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Draft : May M., George S., Prévôt P., A closer look at tracking human and computer interactions in web-based communications, *International Journal of Interactive Technology and Smart Education*, 2008

A Closer Look at Tracking Human & Computer Interactions in Web-based Communications

Madeth May, Sébastien George, Patrick Prévôt

Université de Lyon, INSA-Lyon, LIESP Laboratory,
La Doua, 21 Avenue Jean Capelle – Villeurbanne, F-69621, France
Email : {Madeth.May, Sebastien.George, Patrick.Prevot}@insa-lyon.fr

Keeping track of user's communication activities in Web-based environments has always been considered a complex task. It requires tracking systems that are capable of efficiently tracking users' activities and producing tracking data that can be useful to various users. The objectives of this paper are twofold: (i) to present an approach for better observing the different levels of Human & Computer Interactions during a Computer Mediated Communication activity, and (ii) to present the technical aspects of a Web-based tracking system for communication tools such as discussion forums. Our research applications are applied to educational settings. The proposed approach focuses particularly on what, where, when and how to thoroughly collect the tracking data of user communication activity in distance learning situations. We also present the three cases of experiments that we have conducted, along with the results analysis, and some examples of how we can exploit the tracking data to support the participants in the learning process.

Keywords: Tracking system, Computer Mediated Communications, Human-Computer Interaction, E-Learning, Tracking Data Analysis and Visualization, Data indicators

1. INTRODUCTION

According to Berge and Collins (1995), Computer-Mediated Communication (CMC) promotes a type of interaction that is often lacking in the traditional teacher-based classroom. It allows learners the freedom to explore alternative pathways to find and to develop their own style of learning. CMC tools are also a part of distance learning environments. They are widely employed in distance learning platforms (Guzdial & Turns, 2000), to support communication activities between both learners and teachers.

CMC tools such as Chats, discussion forums, whiteboards, and collaborative learning environments, play a very important role in knowledge sharing between the participants in the learning process. They provide an important learning opportunity for students and an increase in the ease of exchanging ideas for subject that are more discursive (Guzdial & Turns, 2000). A lot of discussion topics related to learning contents, research projects, and so on, can be brought up and exchanged between the students and teachers with the assistance of CMC tools. Furthermore, using asynchronous communication tools, like

discussion forum has the potential to improve the teaching and learning experiences in traditional classroom formats (Groeling, 1999), as well as in distance learning (Corich et al., 2004). Thus, communications have become a big part of the students' activities in the distance learning process.

As a means of providing more convenient support to CMC-tool users, particularly to both teachers and students, we have conducted a research project, which is involved in tracking and exploiting the recorded traces of student's communication activities. Tracking data is the core item for this research work. One of our main objectives is to closely observe the students' communications when they are using CMC tools in their learning activities. We are interested in a study of how to efficiently track the different levels of Human and Computer Interactions (HCI) during CMC activities. We also aim at assisting both teachers and students, during and after their communications and providing them more useful information on their activities. As an example, by exploiting the recorded tracking data, we would like to keep the teachers informed of students' activities, and to give feedbacks to the students on what they have done during a learning session. The properties of the tracking data inside Web-based learning contexts have been particularly studied in this research work. However, due to their riches and complexities, some aspects like privacy issues regarding the users and the use of tracking data in academic institutions will not be discussed in this paper. It is very important to mention here that the participants we mention in this paper are clearly informed about the tracking processes on their communication activities and they are totally aware of the limitation of the use of any recorded tracking data.

The rest of this paper is structured as follows. In the second section, we discuss about the important research issues in collecting and exploiting learner's tracking data via a number of interesting related works. The third section is dedicated to a presentation of our approach, which mainly focuses on the different levels of Human and Computer Interactions to be observed during the tracking process. An example will be given to show how a user activity in discussion forums is being observed and how the tracking data at each level of interaction are being used for. We discuss in the same section the technical aspects of our Web-based tracking system. In section four, we present the three cases of our experiments, followed by the results analysis, and the use of tracking data to assist the participants in the learning process. An assessment on our approach is made in the same section as well.

2. TRACKING DATA IN WEB-BASED EDUCATIONS

Why Tracking Data?

We present in this section the use of student's tracking data in different learning contexts via a number of related works that are close to ours.

As mention Choquet & Iksal (2007), the tracking data of a user's activity can be referred to "traces of use", generally generated by tracking system in accordance with its defined trace format or model. In Web-based learning environments, the trace of learners' activities is a significant source of information that reveals not only the activities themselves, but also their outputs (i.e. the results of the activities that the learners carried through the learning process).

The traces of learners are known as important elements, keeping the teachers informed of different learners' activities and other aspects. Courtin & Talbot (2006) focus on the observation process in distance learning environments. They introduced a model for trace analysis that allows teachers to understand different sequences of learners' activities and the communications exchanged between learners. Després (2003) presented a synchronous "Student Monitoring Tool" that observes the participation rate and the progress of each student during a learning session. In the research work of Komis et al. (2002), the tracking process has been done through collaborative learning environments. This allows the instructors to supervise each individual as well as groups of learners while being in remote situations. Furthermore, by analyzing the traces in collaborative learning environments, the instructors could evaluate social and cognitive aspects of learners. Evidence of a similar approach can also be found in the research works of Riccardo & Dimitrova (2003), where teachers could view three main aspects of learners which include social, behavioural and cognitive aspects.

Donath et al. (1999) conducted one of the most interesting research works regarding the

learners' tracking data visualization. The synthetic information derived from the trace analysis and visualized in graphical representations could help learners review their own behavioral aspects and that of others. Regarding the tracking data visualization, Hardy et al. (2004) provided a set of "Learner Behavior Discovering" tools that allow teachers to review learners' behavioural aspects while accessing to the e-Learning materials and interacting with the associated delivery environment.

Tracking data plays another role in helping both researchers and developers improve the learning platforms and develop adaptable educational tools that better match the needs of individual user (Avouris et al., 2003). Last but not least, the concept of trace was also the subject of various researches outside the educational settings such Human Computer Interaction application design and evaluation (Hilbert & Redmiles 2001). For an example, in order to evaluate the different aspects of an interactive application, system designer studies the ways that "traces of use" should be produced and exploited. Evidence can be found in (Tarby et al., 2007), the discussion of two different approaches on exploiting the recorded traces from the use of interactive applications to evaluate their usability and utility.

During the study of these related works, an important observation has been made. The lack of functionalities of existing CMC tracking systems and the ways tracking data have been manipulated will be illustrated in the next section. We will point out the important research issues related to (i) the observation of learners' activities in Web-based learning environments, (ii) the use of existing log files, and (iii) the data analysis and visualization.

Tracking Data: the Key Issues

A number of important research challenges in keeping track of learner's communication activities on CMC tools have been discussed in our previous work, presented in (May et al., 2006). In this section, we would like to address three main key issues related to use of CMC tracking data, from the production to the visualization of the data.

2.2.1 Observing learner's communication activities

In order to efficiently track learners' communication activities on CMC tools, the tracking system must closely follow where the activities will take place. However, in the existing approaches, most systems were designed to observe the user's activity on only the server side (e.g. where the communication platform is hosted), the user's interaction on the client side (e.g. user Web browser) is completely ignored. In this case, the granularity of traces should be rather large and the information returned from the trace analysis might not be accurate enough to reflect the complete activities of users during their communications.

An attempt is made to investigate "how to make tracking systems capable of producing traces that are not just a simple history of users' activities, but containing substantive information that can be useful to various users?". The approach described further on in this paper focuses on the observation of users' activities on both client and server sides, thus keeping track of (i) Human-Computer Interactions, (ii) Human-Human Interactions Mediated by Computer, (iii) Computer-Computer Interactions, (iv) Non computer mediated user action, and (v) Computer action without user action. This allows us to have various compositions of traces with finer granularity in which user interactions and semantic aspects of the communications can both be found.

Another observation we made about the existing tracking systems is that the activities of "lurkers" on CMC tools have never been tracked down. Let us consider an example of a discussion forum. A lurker is a type of user who does not participate in the communications with other users and who is not "visible" to other users when online. Lurkers might log on to the forum to view other users' discussions, but have no intention of exchanging any messages with the discussion groups. However lurkers are recognized as an important part of Internet community, as mentioned in (Smith, 1999). Hence, tracking lurkers' activities and analyzing their traces permit us to better understand lurker behavior and their influences in distance learning environments (Takahashi et al., 2003).

Besides the idea of observing communication activity on both client and server sides, tracking all other users and studying their traces is one of the particularities of our research. In doing so, we hope to provide more convenient support to

CMC-tool users, especially to increase teachers awareness of other users' activities, including lurkers'.

2.2.2 Structuring tracking data

Since each choice of modeling and structuring traces was made to match each individual need, traces of users' activities stored on existing CMC tools are often carried out in an ad-hoc manner, which either confines the reusability of data in different purposes or makes data exploitation difficult (i.e. traces can be hard exploited independently by different exploitation tools).

We noticed that in related studies, most tracking systems still used text log files to keep track of users' communications on CMC tools. Consequently, the traces stored in log files have rarely been exploited by the users (instructors and learners) either because of the ignorance of their existence, or because the traces do not match the demands of the users. Furthermore, the structure of traces in a log file varies from one CMC tool to another, due to the fact that each log file depends on how it was generated. Yet, there is a lack of semantic aspects for traces stored in log files (e.g. pure text log file).

To avoid this kind of situation, traces should be represented in a generic format from which standard or specific formats can be created and used for various communication tools. The main advantage of formalizing a generic format is to allow users to represent the identical traces in different formats and to manipulate the traces in a similar way (e.g. when using the same operations for different data processing in the phase of traces exploitation). In addition, we need to consider the possibilities of enriching the recorded traces; the fact that traces are being modified by adding more descriptive data (supplementary information) to their original representation, allowing traces to be restructured, transformed into another format, and reused for other types of CMC traces exploitation tool. Moreover, the repository type that stores the traces should be independent of the platforms that give access to it. This will ensure that the data are always accessible and the compatibility of the data (i.e. the reusability of the data in other platforms) is always present.

2.2.3 Analyzing and visualizing tracking data

The learner's tracking data are somehow made up of specific information which is not directly interpretable by the CMC-tool users without the assistance of the specific tools. The traces can be analyzed quantitatively and/or qualitatively with both interactions analysis methods (Pozzi et al., 2007) and content analysis methods (Wever et al., 2006). The major problems that we usually encounter are due to the effectiveness of the method used and the quality of the results returned from the traces analysis. In fact, the analysis can not be done efficiently if the recorded traces are not descriptive enough and when there is lack of traces indicators. Thus, the root of the problem seems to have a higher degree of correlation to the collection of the tracking data. Once again, the approach described in the following section shows how to efficiently track learners' communication activities.

With regards to the traces visualization, various visual forms of traces are to be considered: textual, histogram, or graph with multi dimension. In order to facilitate users in visualizing traces, the visualization tools must be equipped with friendly Graphic User Interface (GUI) components, by which users could easily interrogate the trace repository by simple formal query and transform traces into graphical representations. Yet some visualization tools provide only the overview of users' activities and often in a unique form. To improve upon the deficiency of such tools, the visualization tools should enable users to visualize, as rigorously as possible, the traces of users' activities and the same traces in different visual representations as well as in different scales.

3. TRACKING HUMAN-COMPUTER INTERACTIONS: A CLOSER LOOK

Proposed approach

We studied both synchronous and asynchronous Computer Mediated Communication tools. With the participation of researchers from different disciplines, including HCI and E-Learning specialists, we built our approach with a participative method. We adopted the 5W1H method (When, Where, Who, What, Why, and How) to build our approach. By answering the *who* and *why* questions, we were able to identify the real needs of different participants in the learning process and to contextualize the CMC activities to

be tracked. It is important to mention here that every answer to the questions was useful for us to define the objectives of the research as well as the tasks to be accomplished by the approach. Our goal is to make our approach able to be applied to various educational settings. Nevertheless, we came across a lot of challenges like to propose an explicit approach for different CMC tools or to make the tracking system flexible to the target learning environments. That is why we had to clearly define the context of users' communication activity and the nature of the tracking data to be collected.

We give below the explanation of the approach in a general context. We focus particularly on where, what, when and how to track learner's communication activities on CMC tools. We started by distinguishing different levels of Human and Computer Interactions in a Computer Mediated Communication activity as shown in figure 1.

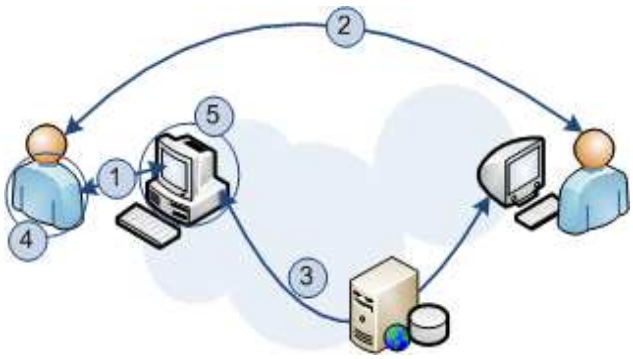

Figure1 Different levels of Human and Computer Interactions to be observed. (1) The interaction between user and the computer interface. (2) The interaction between users mediated by computers. The user's computer can be connected either with or without server. (3) The interactions between computers via physical P2P or Server-based network. (4) The user behavior while using computer to communicate with other users. It is a non-mediated interaction. (5) The computer action without user action

What and Where to Track?

During a Computer Mediated Communication activity, five levels of interactions can be observed:

(1) The Human-Computer Interactions refer to the user's actions while using the Graphic User Interface of CMC tools to communicate with other users. If we look at an example of an activity "Writing a new message", the interactions between a user and a CMC-tool interface can be "edit" message title or message content, "move" vertical scrollbars upward or downward, "drag & drop" smilies into the message, etc. All of these actions occurred only on the user interface. In a Web-based application, these actions occur on the user Web browser without sending any request query to the server or to any other user's machine. In our study, we define the traces of user's activities by a composition of two parts of the traces: a part that represents activities on the server side, which is collected at the moment of exchanging user queries between client browser and server, and another part that represents the Human-Computer Interactions on the client side (i.e. user Web browser).

Tracking Human-Computer Interaction on the client side allows us to follow the successive user's actions and events that occurred during a CMC activity. The main reason of having an observation system on the client side is that most of the time, users interact with the CMC tool interface by using the "cache" memory on the client side without having to exchange any request query with the server. Examples can be found when a user clicks on "backward" or "forward" button on the CMC tool interface to edit the message or other previous inputs. Such interactions represent a big part of the whole activity (i.e. Writing a new message) and the tracking data of these interactions are useful for rebuilding the successive processes of the past activity (e.g. what did a user do to write a new message).

We usually use the tracking data at this level to identify the user's behavior while using a CMC-tool for the communications.

(2) The Human-Human Interactions Mediated by Computer refer to the content of the interaction exchanged between users. With the same example of "Writing a new message"; all the written text on the user interface will be submitted to the server or directly to other user's machine so that the message can be read by other users. To do so, user has to click on "Send" or "Submit" button on the user interface. The message is being sent via a request query to the server or to the machine where the message must be stored or displayed. Tracking the content of the communication makes the tracking data more descriptive, thus enabling us to know how a user

writes a new message and what the message is about.

The collected tracking data at this level will be exploited along with the tracking data at the first level (i.e. Human-Computer Interaction). This usually leads us to easily reproduce not only the general context of the communication activity that describes the successive sequence of user's interactions, but also the content of the communication (i.e. the semantic aspect of the activity).

(3) The Computer-Computer Interactions: keeping track of meaningful events means to track also the computer input and output processes while a communication happens. The tracking data of Computer-Computer Interactions serve two main purposes: (i) evaluation of the quality of the computer processes in exchanging the communication data and (ii) monitoring the CMC-tool performance. The results are most of the time very useful for the developers who seek to improve the CMC tools, and the researchers who are involved in development experiences. As an example, we commonly use the tracking data at this level to debug our system and to strengthen the security of the communication.

(4) The non-computer mediated user action: in other words, it is all user's other actions outside the computer environment (e.g. a user makes a phone call during the learning session). In some circumstances, particularly in distance learning situations, it is not sufficient to track only the computer mediated activities of the learners. Video and audio recorders are more practical in observing the learner behavior. The audio-visual data can be then used for multipurpose, among which the analysis of user's behavior while working individually or collaboratively. It is important to mention here that we do not take into account the audio-visual data at the current stage of our research work. However, observing users' behaviors with the audio-visual tracking system has been considered for the future work.

(5) The computer action without user action: there are plenty of computer actions that occur automatically without the action of the user. Examples include a pop-up message indicating to the user that his/her session in the chat room will be expired in 5 minutes, or a jingle to alert that a new member has logged in to the forum. Tracking such computer actions can be done on both the client and server sides. On the client side, we can capture most of events that occurred and showed up on the user interface, as on the server side, the events will be captured once the request query has been launched and executed. The tracking data at this level will be used as supplementary information to complete the tracking data from the other previous levels. They usually reflect what else happens besides the Human-Computer Interactions.

When and How to Track?

Since there are a great variety of CMC tools in Web-based learning environments, the wisest solution is not to build a tracking system for each single tool. The most appropriate solution is to study the common points and the particularities of each tool and to propose tracking system architecture, which is applicable to a variety of CMC tools. For example, it is undeniable that every CMC tool provides a functional tool for "writing a message"; that is the common point. The dissimilarity is the possible ways a user can employ it to write a message. The particularities of CMC tools are mainly about the User Interfaces and the types of Human and Computer Interactions available in each tool – when a user writes a message in forum 1, placing the written message in to a thread category which is feasible through a multi-selected drop list. The user would do that otherwise in forum 2, because instead of multi-selected drop list, forum 2 proposes a set of checkboxes for the thread categories. The final results of that activity are the same; however, the way the user interacts with the two forums is different. Therefore, we started to formalize the use models to describe the way users employ each functional tool to perform their communication activities.

A use model enables us to (i) define the context of a user's activities and (ii) identify every user action on the interaction objects and its associated events. As shown in figure 2, the identification of different user actions, interaction objects and the associated events of an activity «Write a message». We separated the Human-Computer Interaction from the whole activity to get a clearer view of what action a user can perform on a CMC-tool interface, what kind of interaction it is, and what happens when there is an interaction. The biggest advantage of doing so is to make the tracking system able to observe every Human-Computer Interaction (e.g. a user clicks on a button) and the associated event (e.g.

what happens when a user clicks on a button) – and that is what makes a tracking system efficient.

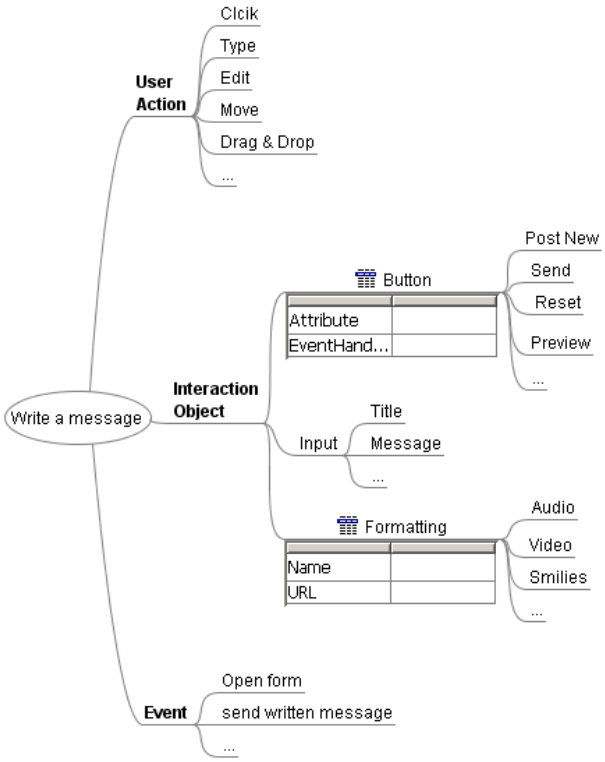

Figure 2 An example of the identification of User action, Interaction objects and the associated Events of an activity "Write a message"

To better understand when and how to track a CMC activity, an example of a use model for the activity «Post a new message» in a discussion forum (c.f. figure 3) is given below.

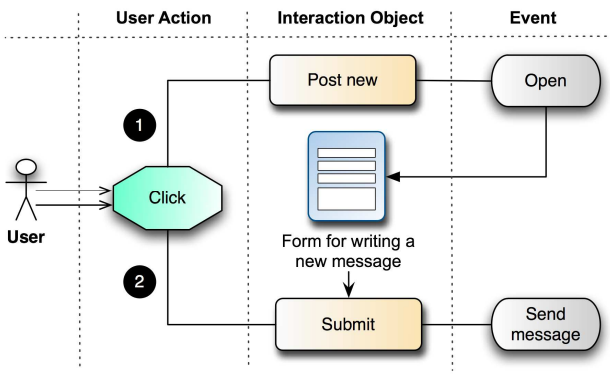

Figure 3 An example of a use model for an activity «Post a new message» in a discussion forum

The interaction objects in the context of this activity could be a «Post new» button, a «Form for a new message», and a «Submit» button, by which users employ to post a new message. The arrow (1) represents a sequence of events that happened when the user clicks on the «Post new» button to open the «Form for a new message» in order to write a new message. This form includes several other interaction objects, in this example, a «Submit» button. When the user clicks on the «Submit» button (arrow 2), there is another event called «Send message», representing the action that the user's message is being submitted to the server.

The identification of the interaction objects and the successive events to be observed let the tracking system take into account every user's action, and to simultaneously produce the tracking data of user's activities in accordance with its defined use model. In this way, each use model indicates how to observe, when to capture the user's actions and/or User-Machine Interactions, and what to generate as tracking data.

A Web-based Tracking System

In (May et al., 2007), we presented in detail the architecture of our tracking system for CMC tool and how each system component was built. The following is a brief description of the two principal components of the system, the observation component and the trace repository.

Observation components

The observation components were specifically designed with a number of "traces collectors", which ensure the observation of the user's interactions on the client side (Human-Computer Interaction) and the user's communications on server side (Human-Human Interactions Mediated by Computer). The observation component is attached with a number of use models which describe how each communication activity on the CMC tool can be performed by a user and how the trace collector generates instantaneously the tracking data representing the interaction of users and the associated communication content.

Via figure 4, we will look at an example of a tracking process, showing how an activity «Post a new message in a discussion forum» is being tracked and how the tracking data are being generated and stored.

The user's interactions on the forum interface, such as typing a message, drag & drop smilies into the message, moving the scrollbar up or down will be captured by traces collectors on client side. The tracking data will be generated and temporarily stored on user's workstation. When the user clicks on the «Submit» button, there is a HTTP server request query to submit the written message to the server. The trace collector on the server side captures that request query and generates simultaneously the tracking data to represent the communication activity (i.e. post a new message) as well as the content of the communication (i.e. written message).

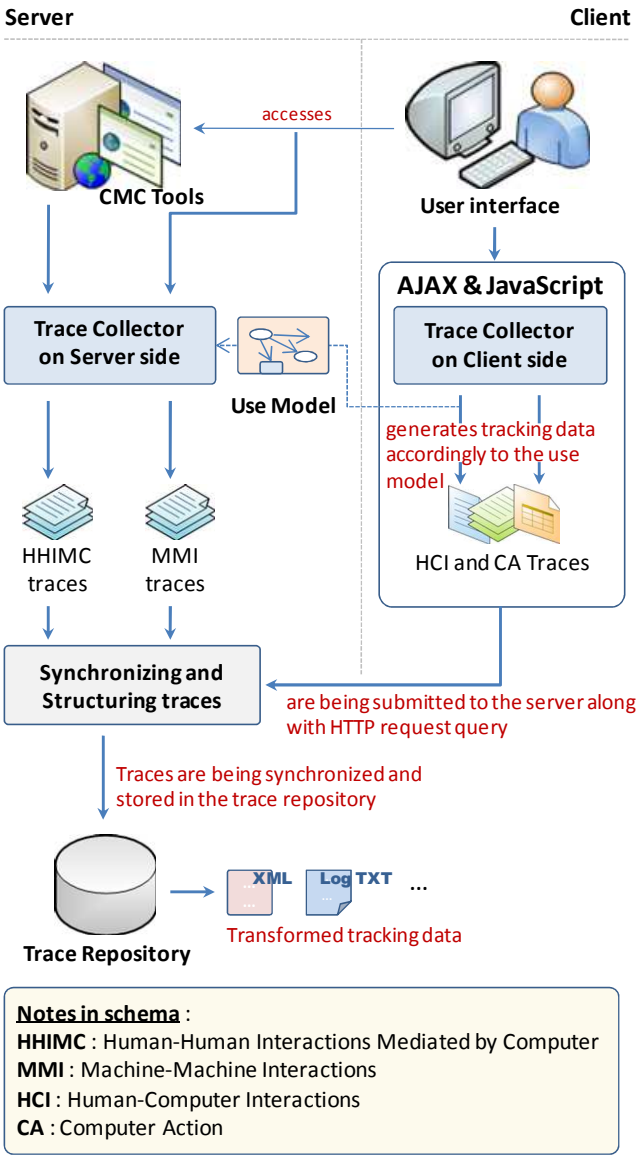

Figure 4 General architecture of Web-based tracking system on Client and Server sides

At each HTTP request, the temporary tracking data, previously stored on client workstations, will be submitted to the server. These data will be next synchronized with those on the server, structured and stored in the trace repository.

The given example showed only how tracking the Human and Computer Interactions at the first and the second levels works, the tracking process at other levels are based on the same concept.

We had developed the trace collector on the client side by using JavaScript and AJAX technologies. JavaScript is a lightweight scripting language which is executed on user's Web browser (client side) and supported by any kind of Web browser. AJAX (Asynchronous JavaScript And XML) is a cross-platform technique usable on many different operating systems and Web navigator as it is based on open standards such as JavaScript and XML. It should be noted that we are not developing any spyware-type application and we do not need to install any tracking application on the user machine. By using AJAX technologies, we are able to make our Web-based tracking system more flexible in term of manipulating the tracking data directly at the client side, the generation and sending of the tracking data to the server in the background are done without interrupting the user's navigation. The predefined use model of each communication activity allows the traces collectors on both client and server sides to exchange the information and to make the information coherent, e.g., the server is capable of synchronizing the tracking data that are being submitted from clients with those on the server.

Trace repository

The trace repository was implemented as a centralized database server with MySQL. It contains the meta data that are used to structure the tracking data from the traces collectors of both client and server sides. The choice of using a relational database like MySQL for storing CMC tracking data has several advantages such as, (i) traces are structured in a rich format, (ii) traces can be easily restructured and transformed into another format (e.g. XML/RDF or TXT), and (iii) the operations for traces manipulation such as insertion, modification, etc., can be easily performed with simple SQL queries.

4. EXPERIMENTS AND RESULTS

We have conducted three experiments with the participation of students, teachers, and researchers from different disciplines. The main objective of our experiments is to evaluate the efficiency of our Web-based tracking system and at the same time the quality of the tracking data. We also would like to study the real needs of the participants from these experiments, so that we can make use of the collected data to support the CMC-tool users, particularly teachers and students in their communication activities.

Experiment Setup

We employed a free and Open Source Course Management System, Moodle (http://moodle.org/) for our experiments. Since Moodle possesses a log database to keep track of user's navigation, we were able to use it as reference trace repository and we could look at some important aspects such as the quantity, quality, and data indicators of the recorded traces in both Moodle log database and our trace repository.

A random course was picked up by our research team and added to the platform Moodle so that the participants could carry out the discussions by concentrating on its contents. In order to make the discussions more relevant to the learning activities (e.g. the educational scenario that is implemented in the chosen course), a contextual forum, CONFOR (George & Labas, 2007) was integrated in to the Moodle platform. Thus, the communications between the participants were done on CONFOR and not on the Moodle built-in discussion forum. Thanks to the tracking system architecture that is based on "Observation components" (c.f. section 3.4) and "Use model" (c.f. section 3.3), new "Traces Collectors" to CONFOR and Moodle were easily added.

It is important to mention here that CONFOR is a particular discussion forum that was designed and developed to connect the learning activities to the discussion activities. CONFOR has been employed in distance learning platforms to incite the learners to use forum as a tool for discussion and for sharing knowledge between them. This means that the students' discussions are strongly based on their apprenticeship (e.g. questions and answers on their lessons and assignments, etc.). The reason of using CONFOR is that we are strongly interested in the communication activities that are only related to the course contents and the given learning scenarios.

In each experiment, two different use scenarios have been prepared to guide the participants in their communication activities. The two use scenarios consist of both learning and discussion activities on different topics and different orders. They focus mainly on the following activities: (i) Browsing the forum structure, (ii) Viewing lessons, (iii) Posting new messages in the forum, (iv) Replying to messages in the forum, and (v) Reading message in the forum. Each use scenario was formulated to incite the participants not only to go through the important steps as described in the learning scenario, but also to share their knowledge by posting messages on the discussion forum.

The first use scenario was distributed to the half of the participants, and the second one to the other half. The main idea of having two use scenarios is to make sure that a big number of participants will not do the same actions, on the same things, or at the same time. By doing so, we came up at the end of each experiment with a large number of records which somehow describe the various activities of the participants, and from which we could study different learning situations of the participants while they are using CMC tools, like discussion forum.

Results Analysis

We present below some quantitative analysis of the experiments and two examples of the recorded tracking data. Table 1 gives us a summary on the experiment conditions and the number of records we have stored in the trace repository after each experiment.

Table 1 Information on the experiment set up and the records from the experiments

No	Number of Participants	Number of Messages	Duration (minutes)	Number of Actions Recorded
1	8	62	30	1037
2	26	83	45	1399
3	26	147	45	2421

The numbers of records from each experiment represent both Human and Machine Interactions during the communication activities.

In the first experiment, with only 8 participants and within 30 minutes, we arrived up with 1037 records. This number is almost two-third higher than the records in Moodle logs database (in Moodle 1.5, table mdl_log is for storing user logs). The explanation is quite simple. Our Web-based tracking system is capable of observing the Human-Computer Interactions which have been carried out on the client side (i.e. user Web browser). Every other user action that has not been gone through Moodle server has been tracked by the "Trace Collector" on the client side and the tracking data have been synchronized with the server side data, structured, and stored in the trace repository, as we have described earlier in section 3.4 of this paper. Thus, each record has finer granularity that contains different actions of user and machine during the user's navigation on both Moodle platform and CONFOR discussion forum.

In the second and third experiments, there is a big difference for the number of records even the two experiments were conducted with the same number of participants and the same amount of time. Here is the explanation. In the third case, the participants have exchanged a lot of messages. This usually happens when there are interesting posts on a particular topic. By looking at the recorded traces, we found out that, the participants have done some other activities, which are not described in the given use scenario. Since every user action has been tracked, it is usual to find out at the end of the third experiment, the number of records that high. There were almost 73% of extra recorded data comparing to the second case. Interesting information (e.g. data indicator) can be extracted from these records. First of all, two different communication behaviors corresponding to the two given scenarios have been identified. However, from the trace analysis, we also found out that more than 80% of the participants from the third case did not really follow what exactly described in the use scenario.

Besides the quantitative analysis, we have particularly looked for other significant data indicators regarding the semantic aspects of the users' activities. Such indicators should be able to tell us whether or not the participants really followed what they were asked to do, or if a particular participant started a discussion topic after viewing a lesson, and what the discussion was all about, etc. Before having a look at a few examples of how we can exploit those tracking data to support the participants in the learning process, we give below two examples of the recorded tracking data, respectively shown in figures 5 and 6. We have mentioned in section 3.4 that the tracking data are originally in relational format and stored in MySQL database server (i.e. trace repository). The given examples show the recorded tracking data, transformed in to an XML format.

The tracking data of each communication activity is described by a set called "Activity". The "Attribute" and "Delay" properties are the two major compositions of tracking data. The "Attribute" property describes the data attributes associated to each individual communication activity. It contains the information that link to other information which is not recorded in the trace repository. The "Delay" property represents the duration of an activity. If we look attentively at the recorded tracking data, we can find some significant information that reflects the whole activity of a user within a communication activity.

```
- <HCI id="8099">
    <User>Lucas</User>
  - <Activity>
      <Title>Reply to a message in forum</Title>
      <Attribute>IDForum=4586,IDMsg=263,IDMsgParent=68</Attribute>
      <Date>19/06/2007</Date>
      <Time>02:09:35</Time>
      <Delay>00:04:39</Delay>
      <RefAct />
    </Activity>
  </HCI>
```

Figure 5 An example of tracking data for an activity "Reply a message in the forum". Inside the "Attribute" tag, we find all the necessary information regarding which message Lucas wrote (i.e. IDMsg=263), replying to which message (i.e. IDMsgParent=68) and in which forum (i.e. IDForum=4586). In the "Delay" tag, we can see that Lucas spent 4 minutes and 39 seconds to write the message

Figure 5 is an example of tracking data for an activity "Replying a message in the forum". The particular indicators we could find in this record are the complete information on the message itself and the total time that spent a user for writing the message.

In figure 6, the tracking data HCI=8096 explains that user (Lucas) displayed a message IDMsg=68, in the forum IDForum=4586. This user spent 3 minutes and 26 seconds to display (or probably read) this message before performing another activity. From the tracking data HCI=8097 and HCI=8098, we are able to know that the user moved the vertical scrollbar to reach the bottom of the message. Such data indicators

are very substantial in automatic traces analysis. As an example, the recorded traces can be computationally analyzed to extract the significant data indicators and the time-consuming tasks, such as manual or semi-automatic traces analysis can be reduced.


```
- <HCI id="8096">
    <User>Lucas</User>
  - <Activity>
      <Title>Display a message in forum</Title>
      <Attribute>IDForum=4586,IDMsg=68</Attribute>
      <Date>19/06/2007</Date>
      <Time>02:06:03</Time>
      <Delay>00:03:26</Delay>
      <RefAct>0</RefAct>
    </Activity>
  </HCI>
- <HCI id="8097">
    <User>Lucas</User>
  - <Activity>
      <Title>Move scrollbar downward</Title>
      <Attribute>IDForum=4586,IDMsg=68,Scrollbar=V</Attribute>
      <Date>19/06/2007</Date>
      <Time>02:07:18</Time>
      <Delay />
      <RefAct>8096</RefAct>
    </Activity>
  </HCI>
- <HCI id="8098">
    <User>Lucas</User>
  - <Activity>
      <Title>Move scrollbar downward (Reach the bottom)</Title>
      <Attribute>IDForum=4586,IDMsg=68,Scrollbar=V</Attribute>
      <Date>19/06/2007</Date>
      <Time>02:08:49</Time>
      <Delay />
      <RefAct>8096</RefAct>
    </Activity>
  </HCI>
```


Figure 6 Tracking data for an activity "Read a message in the forum". Two types of tracking data can be found here, the communication activity on the server side (i.e. HCI id=8096), and the user interaction on the client side (i.e. HCI id=8097 and HCI id=8098)

Exploiting Tracking Data

Apart of the traces exploitation is to extract from the traces the useful "data indicators", with both quantitative and qualitative information that allows the participants in the learning process to analyze and/or to evaluate different aspects of their activities. We chose to represent the data indicators for the communication activities in graphical representations as shown in figure 7, 8 and 9. Those data indicators are generated by our platform TrAVis (Tracking Data Analysis and Visualization).

TrAVis was designed and developed through our research to help both instructors and students to exploit the tracking data of synchronous and asynchronous communication activities. In order to provide ease of use and the flexibility of TrAVis platform, each functional tool was implemented with customizable and friendly Graphic User Interface, allowing the users to easily access to the traces, to analyze, and to visualize the traces in different visual forms as well as in different scales.

It is important to mention that the given examples of exploiting students' tracking data below are inspired by the real needs of teachers in distance learning situations. They also reflect how useful it is to have traces with finer granularity and how CMC tracking data can be analyzed in order to help the teachers evaluate student's communication activities.

"How do we know whether or not a displayed message in a forum is read?". This question has been frequently asked, particularly by the teachers who regularly use discussion forums in their teaching activities. We are not pretending that we can prove if a message was really read by the user who displayed it, but we can tell if a message has not been entirely read. For example, if a user has only rapidly displayed the message (e.g. less than 3 seconds) without touching or moving the vertical scrollbar downward the bottom of the message, but performing another activity instead (e.g. clicked on another message), the displayed message must not have been entirely read by the user. Back to the recorded tracking data shown in figure 6, a user (Lucas) might have read till the end of the message since he has not only displayed the message, but also moved twice the vertical scrollbar downward and to the bottom of the message, and besides, he has spent 3 minutes and 26 seconds on it (i.e. the windows that displayed the message remained active right after the message was displayed and the user did not perform another activity within 3 minutes and 26 seconds). Figure 7 represents these data indicators in graphical representations that allow teachers to easily visualize and interpret information on an activity "Reading a message in the forum".

Each sphere shown in figure 7 represents an activity of displaying a message and the diameter of the sphere is proportional to the time spent by each user reading the displayed message. The distance between two spheres represents the time gap between two different readings. A sphere can be one of the following four colors: orange, blue, green, or grey. The green sphere notifies us that the user read the message by having moved the vertical scrollbar downward to the end of the page

(reading till the end of the message). The orange sphere indicates the fact that the user has simply displayed the message contents without moving the scrollbar. The blue sphere signifies that the user has displayed the message contents and has moved the vertical scrollbar downward, but not to the bottom of the page (i.e. partial reading). The grey sphere indicates that the message was only displayed; the window that displays the message has immediately been left inactive.

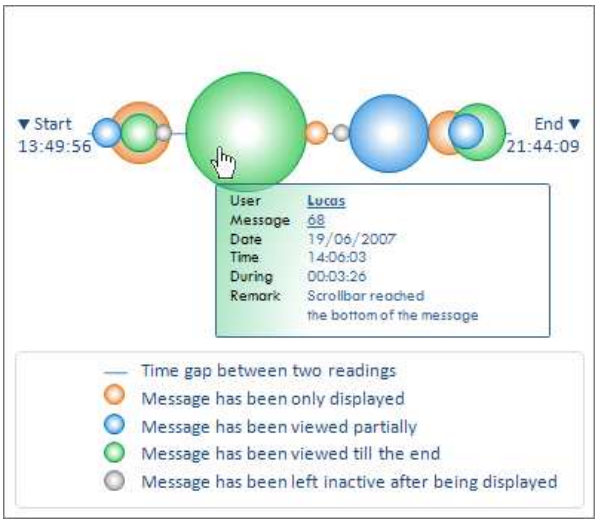

Figure 7 An example of data indicators for an activity « Reading messages in a discussion forum »

Providing such significant data indicators with every little detail of the activity is very useful for the teachers and a lot better than giving only the statistical data like the number of hits on the message, or which user clicks on which message, etc.

Here is another example about viewing lesson contents. The teachers would be very interested to know more than just how many times a student has visited their online lessons. The substantive data indicators to be provided to the teacher should be able to describe which student clicks on which lesson and when, whether or not the student has viewed the lesson, and for how long he/she spent to read it, etc. In each experiment, the recorded tracking data contain the information regarding the different levels of Human & Computer Interactions of a student activity, in which we can easily find all the necessary data indicators to give to the teachers. As shown in figure 8 the indictors for a student profile that give an overview of the four main activities of a student.

Figure 8 An overview of the learning and communication activities of a student named "Lucas"

We are trying to provide not only the quantitative information on a student activity (e.g. how many times a student has visited this lesson), but more importantly the semantic aspect of the activity (e.g. the discussion on a topic featured in the lesson). As shown in figure 9, the indicators for an activity "Writing messages" that allow the instructors to access to more information on a posted message in the discussion forum such as who wrote the message, how long it took, if the message was a reply to other messages versus a brand new message, in which forum does it belong, etc. Such detailed information can be very useful for the instructors, especially during the observation of student communication activities. The instructors could see how often a student participates in exchanging ideas on a certain topic and whether the posted messages are for seeking for help, debating or arguing on a learning topic, etc.

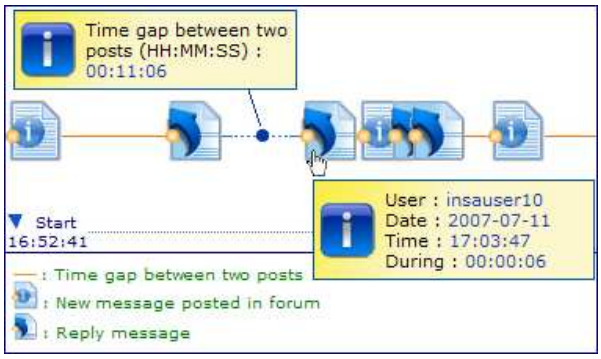

Figure 9 Data indicators for an activity « Posting and replying messages on a discussion forum »

Last but not least, another use of tracking data to assist a teacher who wishes to monitor the student activities within a learning session. Imagine that a teacher has distributed a learning scenario to his classroom and he expects that he will learn from the tracking data about what his

students have been doing and whether or not they have been following the given learning scenario. With the tracking data we have, the teacher could look at the successive activities that the students have carried through. The teacher could also evaluate different aspects of each individual student, such as the social aspect of a student. For example, the teacher could be aware of how a student communicates with the others in the classroom, the discussions that a student has made so far, and the discussion topic he/she usually brought up during the session, etc. In this case, tracking data will be the core element to be exploited by the teacher to keep himself informed about every single activity of the students.

Assessment

At the end of each experiment, the participants were asked to evaluate whether our approach met their expectations. They were also invited to answer an online questionnaire on the proposed data indicators and to precise their real needs in tracking and exploiting the traces of Computer Mediated Communication activities. The questionnaire can be accessible via this link: http://www.madethmay.com/travis/survey/

As an overall evaluation from the point of view of the participants, we could say that the approach was successfully in achieving its goals. In addition, the Web-based tracking system that we implemented for the three experiments has not only proved the feasibly of the approach in an efficient way, but also justified the reason why we need CMC-tool users' tracking data with finer granularity that contain substantial information on CMC activities and their outputs.

The results from the three experiments are very encouraging. We have received a lot of positive comments and feedbacks from the participants on our Web-based tracking system. Aspects the participants appreciated were the effectiveness of our tracking system and the significant data indictors that we could extract from the recorded traces. Besides, the teachers expressed their expectations on how the tracking data should be later on analyzed and visualized to match the needs of the intended participants in a particular learning situation. As for students, they were impressed that their actions on the client sides (i.e. on their Web browsers) could be tracked without interrupting their navigations. We also gathered a number of interesting remarks about what need to be improved in our approach from the researchers who have development experiences.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we present an approach for efficiently tracking user's communication activities on CMC tools, by looking very closely at the different levels of Human and Computer Interactions, where the tracking process must be carried out if we need to collect as thoroughly as possible the necessary information about the CMC activities, including the content of the exchanged communications. We also pointed out the key issues related to the problems in tracking user's activities on Computer Mediated Communication tools and in exploiting the traces that are collected within learning environments. The technical aspects of our Web-based tracking system were presented as well. With the experiences we had from the three experiments, we are currently working on making our approach explicit that can be applied to both synchronous and asynchronous CMC tools to support the tracking process in various learning platforms.

Furthermore, we are willing to conduct more experiments with other communication tools and in different learning situations in order to produce more significant data indicators for the participants in learning process. The result analysis from the questionnaire helps us to provide a support that better matches the real needs of both teachers and students, who often request an increase in the ease of exploiting the CMC tracking data. That is why we are improving TrAVis platform, making it accessible by both teachers and students, and allowing them to fulfil the important tasks in exploiting CMC tracking data, such as transformation, analysis, and visualization.

ACKNOWLEDGMENT

The present research work is a part of the research project "Personnalisation des EIAH" of "Cluster 2", with the financial support of the "Région Rhône Alpes". It has been conducted at the LIESP Laboratory of INSA-Lyon, France. The authors gratefully acknowledge the support of both institutions and researchers who have been actively participating in the cooperative research tasks.

REFERENCES

Avouris, N., Dimitracopoulou, A. Komis, V., Margaritis, M. (2003) Participatory analysis of synchronous collaborative problem solving using the OCAF methodology and tools. In Proc. CSCL , pp. 232-234.

Berge, Z., Collins, M. (1995) Computer-Mediated Communication and the Online Classroom in Distance Learning, Computer-Mediated Communication Magazine, Vol 2, No. 4.

Choquet, C., Iksal, S. (2007) Modeling tracks for the model driven reengineering of a TEL System, Journal of Interactive earning Research (JILR), Special Issue Usage Analysis in Learning Systems: Existing Approaches and Scientific Issues, Vol. 18, No. 2, pp. 161-184.

Courtin, C., Talbot, S. (2006) Trace analysis in instrumented collaborative learning environments, In Proc. IEEE Advanced Learning Technologies.

Corich, S., Kinshuk, Hunt, L. (2004) Assessing discussion forum participation: In search of quality, International Journal of Instructional Technology and Distance Learning, TEIR Center, Duquesne University, Pittsburgh

De Wever, B., Schellens, T., Valcke, M., Van Keer, H. (2006) Content analysis schemes to analyze transcripts of online asynchronous discussion groups: A Review, Computers & Education, Vol. 46(1), pp. 6-28.

Després, C. (2003) Synchronous tutoring in distance learning", In Proc. Artificial Intelligence in Education, AIED , pp. 271-278.

Donath, J. Karahalios, K., Viegas, F.B. (1999) Visualizing conversation, System Sciences, Vol.4, No. 4, 1999.

George, S., Labas, H. (2007) E-learning standards as a basis for contextual forums design, Computers in Human Behavior, http://www.sciencedirect.com

Groeling, T. (1999) Virtual discussion: Web-based discussion forums in political science. Paper presented at the American Political Science Association, Atlanta, Georgia.

Guzdial, M., Turns, J. (2000) Effetive discussion through a computer-mediated anchored forum, Journal of the Learning Sciences, Vol. 9(4), pp. 437-469.

Hardy, J., Antonioletti M., Bates S. (2004) E-Learner Tracking : Tools for discovering learner behavior, In Proc. IASTED Web-based Education, Actapress.

Hilbert, D.M., Redmiles, D.F. (2001), Extracting usability information from user Interface events. ACM Computing Surveys Vol. 32, pp. 384–421

Komis, V., Avouris, N., Fidas, C. (2002) Computer-supported collaborative concept mapping: Study of synchronous peer interaction, Education and Information Technologies Vol. 7, pp. 69-188.

May, M., George, S., Prévôt, P. (2006) A Web-based system for observing and analyzing computer mediated communications. In Proc. IEEE/WIC/ACM Web Intelligence, pp. 983-986.

May, M., George, S., Prévôt, P. (2007) Tracking, analyzing, and visualizing learners' activities on discussion forums. In Proc. IASTED Web-based Education, pp. 649-656.

Moodle, Free and Open Source CMS official web site [online] http://moodle.org/

Pozzi, F. Manca, S., Persico, D., Sarti, L. (2007) A general framework for tracking and analysing learning processes in computer-supported collaborative learning environments, Innovations in Education and Teaching International, Vol. 44(2), pp. 169-179.

Madeth MAY is currently doing his third year PhD research program at the National Institute of Applied Sciences of Lyon, in France (INSA-Lyon). He is working with the "Mediatised Collaborative Interaction " research team of the Laboratory of Computer Science for Industry and Production Systems (LIESP). His research work is particularly involved in Web-based educations, and more precisely in communication activities between learners and instructors. Tracking and exploiting traces of communication activity is one of the interested research subjects that Madeth May has been working on for two years now. Besides his two years of teaching experiences at the National Institute of Technologies of Cambodia, he has also co-supervised a number of trainees who had their internships in the LIESP laboratory.

Riccardo, M., Dimitrova, V. (2003) CourseVis: Externalising student information to facilitate instructors in distance learning, In Proc. Artificial Intelligence in Education, AIED, pp. 279-286.

Smith, A. (1999) Invisible crowds in cyberspace : Measuring and mapping the social structure of USENET, In Communities in Cyberspaces Perspectives on New Forms of Social Organization, Routledge Press.

Takahashi, M., Fujimoto, M., Yamasaki, N. (2003) The active lurker: influence of an in-house online community on its outside environment, In Proc. ACM SIGGROUP Supporting group work.

Tarby, J-C., Ezzedine, H., Rouillard, J., Tran, C.D., Laporte, P., Kolski, C. (2007) Traces using aspect oriented programming and interactive agent-based architecture for early usability evaluation: Basic principles and comparison, J. Jacko (Ed.), Human-Computer Interaction, Part I, HCII, Lecture Notes in Computer Science (LNCS), Springer-Verlag, pp. 632-641.