



HAL
open science

Building Intelligent Web-based System for Efficiently Tracking Learner's Communication Activities on Discussion Forums

Madeth May, Sébastien George, Patrick Prévôt

► **To cite this version:**

Madeth May, Sébastien George, Patrick Prévôt. Building Intelligent Web-based System for Efficiently Tracking Learner's Communication Activities on Discussion Forums. 4th IEEE International Conference on Innovations in Information Technology (Ino 2007), Nov 2007, Dubai, United Arab Emirates, France. p. 148-152. hal-00265667

HAL Id: hal-00265667

<https://hal.science/hal-00265667v1>

Submitted on 19 Mar 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.

Building Intelligent Web-based System for Efficiently Tracking Learner's Communication Activities on Discussion Forums

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

Abstract

The tracking data of the learners' activities is always known as significant source of information, allowing the instructors to be aware of the activities that the learners carried through, along with their outputs. It is at the same time considered a privileged element, exploited by the researchers and the developers to improve the learning environments. Yet, collecting and exploiting tracking data in Web-based learning environment are still among the research challenges. In this paper, we present the latest results of our research works, regarding the approach adopted for efficiently tracking the learner's activities on discussion forums.

1. Introduction

Our research works are particularly involved in keeping track of the learner's activities through the distance learning platforms and exploiting the tracking data in different learning contexts. The main objective of the project is to use the tracking data, collected during the learner's communication activities on discussion forums, to assist both learners and teachers in their distance learning activities. Tracking data are too called "traces", generally generated by tracking system in accordance with its defined trace format or model [1]. The collected traces are used to depict the activities of the users and/or the events occurred during the usage of the systems in a specific context.

In Web-based learning situations, the trace of learners' activities is a significant source of information that reveals not only the activities themselves, but also their outputs (the results of the activities that the learners carried through the learning process). Therefore, traces are known as important

elements, keeping the instructors informed of different learning aspects of learners, like the progress of the learners' activities [2] and the exchanged communications between learners [3][4]. 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 [5]. The synthetic information derived from the analysis could help learners review their own behavioral aspects and that of others [4]. Last but not least, tracking data plays another role in helping both the researchers and the developers improve the learning platforms and develop adaptable educational tools that match better the necessity of each user [6].

This paper presents an approach for efficiently tracking the learner's communication activities on different types of discussion forum. We begin section 2 by uncovering the principal research issues in tracking user's activity and exploiting the tracking data in Web-based learning environments. In section 3, we present the different levels of Human and Machine interactions to be observed. We show how our Web-based tracking system was built in section 4, followed by an example of how a user's communication activities in discussion forums are being captured and how tracking data are being generated and stored in the trace repository.

2. Research issues in collecting and exploiting learner's tracking data

Based on the results of our first Web-based tracking system [7], we would like to pin-point the lack of functionalities of existing Computer Mediated Communication (CMC) tracking systems and the ways traces have been exploited.

2.1. Keeping track of learner's activities

In order to track efficiently the learners' communication activities on discussion forums, the tracking system must observe closely where the activities are going to be carried on. However, most systems were designed to observe user's activity only on the server side, the user's interaction on the client side 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. The approach we present in section 3, focuses on different levels of Human and Computer Interactions, which means that the observation of users' activities is done on both client and server side. This allows us to have different compositions of traces with finer granularity.

Another remark we made about the existing tracking systems is that the activities of lurkers on discussion forum have never been tracked down. Even lurkers do not participate in the communications between other users and who are not visible to other users when online, they are recognized as an important part of Internet community, as mentioned in [8]. Hence tracking the lurkers' activities and analyzing their traces permit us to understand more about the lurker behavior and their influence in distance learning environments [9]. This is one of the particularities of our research, tracking every other user and studying their traces by hoping to address more convenient supports to forum users, especially the teachers so that they can be more aware of other user's activities, including lurkers'.

2.2. Structuring tracking data

Since each choice of modeling and structuring traces was made just to match each individual need, traces 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). During the study of related works, we notice that most tracking systems still use text log files to keep track of users' activities on Computer Mediated Communication tools. Consequently, the traces stored in log files have been rarely exploited by the users (instructors and learners) either because of the ignorance of their existence, or because the traces do not match the exigency of the users. Moreover, the structure of log file varies from one forum 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 (i.e. pure text log file).

To avoid this kind of situation, traces should be represented in a generic format, from which another standard or specific format can be created to represent the identical trace. In addition, it is needful to consider the possibilities to enrich the recorded traces: the fact that traces are being modified by adding more descriptive data (supplementary information) to its original representation, allowing traces to be restructured, transformed into another format, and reusable for other types of traces exploitation tool.

2.3. Exploiting learner's tracking data

The learning's tracking data are somehow made up of specific information which is not directly interpretable by the forum users without the assistance of the specific tools. The traces can be analyzed quantitatively and/or qualitatively with both interactions analysis methods [10] and content analysis methods [11]. The major problems that we faced are due to the effectiveness of the used method and the quality of the results returned from the traces analysis. As matter of fact, the analysis can not be done efficiently when the recorded traces are not descriptive enough and when there is lack of traces indicators. As an example, the analysis that looks for a piece of information in the recorded traces to be extracted in order to answer explicitly *what exactly happens after this student logged on the forum, or if a message in the forum has not been read by users after its displays*. Thus, the root of the problem seems to be much more correlated to the collection of the tracking data. The approach discussed in section 3 aims at bringing the solutions to such problem.

3. Approach for efficiently tracking communication activities

We used the 5W1H method (When, Where, Who, What, Why, and How) to build our approach. We started by answering the *who* and *why* questions in order to identify the real need of different participants in the learning process and to contextualize the communication 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.

Due to the scope of this paper, we present only the crucial parts of the approach and in brief, the explanation of *Where, What, When* and *How* to track learner's communication activities on discussion forums.

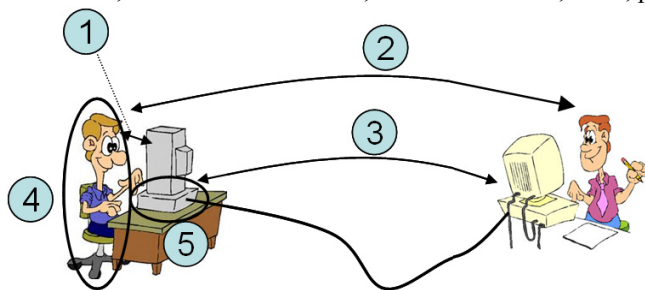


Figure 1. Different levels of Human and Computer Interactions to be observed

We started by distinguishing different levels of Human and Computer Interactions in a communication activity as shown in figure 1. We also took into account the dependency of the tracking data to be produced at each level so that they can be either analyzed independently or dependently.

3.1. What and where to track?

As shown in figure 1, during a 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. Let us give an example of an activity "Writing a new message". The interaction between user and a forum interface can be "edit" message title or message content, "move" vertical scrollbars upward or downward, "drag & drop" smiles into the message, etc. All of these actions occurred only on his/her user interface, more precisely on the user browser without sending any request query to the server or to any other user's machine.

(2) The Human-Human Interactions Mediated by Computer refer to the content of the interaction exchanged between the 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. 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, with what we are able to know how a user writes a new message and what he/she writes.

(3) The Computer-Computer Interactions: keeping track meaningful events means to track also the input and output process of the computer 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 communication tool performance. The results are most of the time very useful for the developers who seek to improve the communication 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 user behavior/attitude while using communication tool is a non-mediated interaction. It means every other actions of the user outside the computer environment (i.e. user drinks a cup of tea while writing a message, or 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 activity 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.

(5) The computer action without user action: there are plenty of computer actions that occur automatically without the action of the user. As an example, a pop-up message telling 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 been logged in to the forum etc. Tracking such computer action can be done on both client and server side. On the client side, we can capture most of events occurred and showed up on the user interface, as on the server side, the event will be captured once the request query has been lunched and executed.

3.2. When and how to track?

Since there is a variety of discussion forums used in Web-based learning environments, the wise solution is not to build for every single forum a tracking system. The key to the solution is to study the common points and the particularities of each forum and to propose tracking system architecture, which is applicable in different types of discussion forums. Here is an example; it is undeniable that every forum provides a functional tool for "writing a message" and that is the common point. The dissimilarity is the way a user can employ it to write a message. The particularities of discussion forums are mainly about the user interfaces and the types of Human and Computer interactions available in each forum. When a user writes a message in *forum 1*, placing his/her message in to a thread category 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 but the way the user interacts with the two forums is different. Thence, we started to formalize the **usage models** to describe the way users employ each functional tool to perform their communication activities. We give below the explanation of what a usage model is and what purpose it is used for.

4. A Web-based tracking system for discussion forums

4.1. Usage model

A usage model enables us to (i) define the context of user's activities and (ii) identify every **user action** on the **interaction object** and its associated **events**.

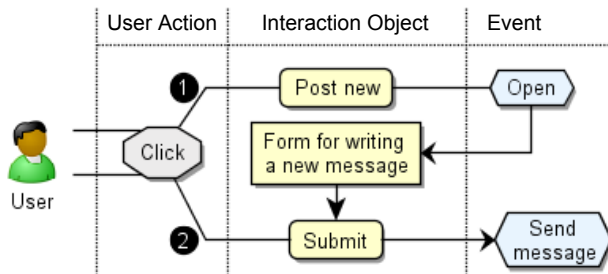


Figure 2. An example of a usage model for an activity «Post a new message» in a discussion forum

As shown in figure 2, a usage model for an activity called «Post a new message» on the 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 event that happens when the user clicks on the «Post new» button to open the «Form for a new message» in order to write a new message. 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 interactions with those objects, and to produce simultaneously the tracking data of user's activities in accordance with its defined usage model. In this way, each usage model indicates **how** to observe, **when** to capture the user's actions and/or User-Machine interactions, and **what** to generate as tracking data.

4.2. Observation component

We presented in detail in [12], the architecture of our tracking system for discussion forums and how each system component was built. **The observation components** were specifically designed with a number of **traces collectors**, which take care of observing user's interaction on client side (Human-Computer Interaction) and user's communications on server side (Human-Human Interactions Mediated by Computer). The observation component is attached with a usage model, which describes how a communication activity on the forum can be performed by a user and how traces collectors generates instantaneously the tracking data representing the user's activities and its associated communication content.

Via figure 3, let us give an example of a tracking process, showing how an activity «Post a new message» is being tracked and how the tracking data are being generated and recorded.

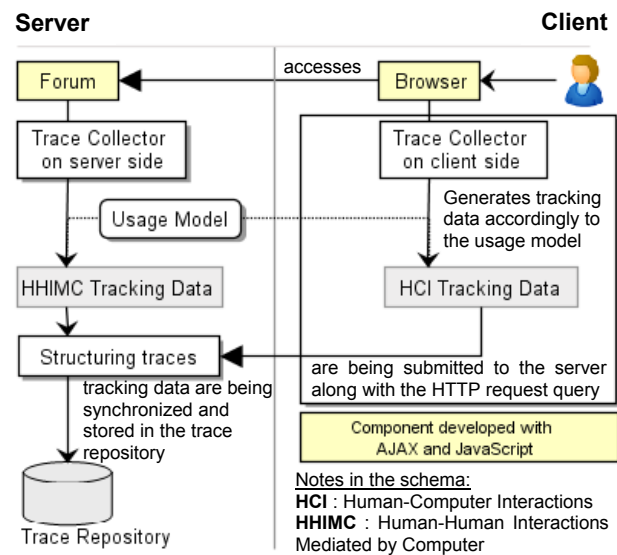


Figure 3. General architecture of Web-based tracking system on client and server side

The user's interactions on their browser interface, such as typing a message, drag & drop smiles into the message, moving scrollbar upward or downward, etc., 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 collectors on the server side capture that request query and generate simultaneously the tracking data to represent the communication activity (i.e. *post a new message*) as well as the content of the

Author version submitted to

The 4th IEEE International Conference on Innovations in Information Technology (Inno 2007),

Dubai, United Arab Emirates, November 18-20, 2007, pp. 148-152

communication (i.e. *written message*). At each HTTP request, the temporary tracking data, previously stored on client workstation, will be submitted to the server. These data will be next synchronized with those on the server, structured and stored in the trace repository.

We had developed the trace collector on the client side by using JavaScript language 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. With AJAX, we were able to make our Web-based tracking system more flexible in term of manipulating the tracking data directly at the client side, such as generating and sending the tracking data to the server in the background without interrupting the user's navigation. The predefined usage model of each communication activity allows the traces collectors on both client and server to exchange the information and to make the information coherent, e.g., the server is capable of synchronizing the tracking data submitted from clients with those on the server.

5. Conclusions and future work

In this paper, we pointed out the key issues related to the problems in tracking user's activities on CMC tools and in exploiting the traces collected within learning environments. Our principal contribution to this research challenge is an approach for efficiently tracking the user's communication activities on different types of discussion forums. The suggested approach is for building Web-based tracking system, which takes into account different levels of Human and Machine Interactions, where the observation must be carried out in order to collect as rigorously as possible the necessary information about the Computer Mediated Communication activities, including the content of the exchanged communications. The tracking data at each level contain different indicators that will be served by different users, especially the participants in the learning process as well as the researchers and the developers of Web-based communication tools.

Our future work is to make the approach become general so that we can provide an explicit framework to support the tracking system of different CMC tools. Besides, we are now working on a development of a platform for traces management, in order to analyze and visualize the traces of learner's communication activities in CMC tools.

6. References

- [1] Choquet C., Iksal S., "Modeling Tracks for the Model Driven Reengineering of a TEL System". *Journal of Interactive Learning Research*, Special Issue Usage Analysis in Learning Systems: Existing Approaches and Scientific Issues, Edited by Chesapeake, Vol. 18, n°2, pp. 161-184.
- [2] Després C., "Synchronous Tutoring in Distance Learning", *AIED Sydney (Australia)*, Edited by IOS Press, *Artificial Intelligence in Education 2003*, R. Mizoguchi (ed.), 20-24 July 2003, pp. 271-278.
- [3] Komis V., Avouris N., Fidas C., "Computer-supported Collaborative Concept Mapping: Study of Synchronous Peer Interaction", *Education and Information Technologies*, Vol.7, 2002, pp. 169-188.
- [4] Donath J., Karahalios K., Viegas F.B., "Visualizing Conversation", *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences*, Vol. 4(4), 1999.
- [5] Riccardo M., Dimitrova V., "CourseVis: Externalising Student Information to Facilitate Instructors in Distance Learning", *Proceedings of the International Conference in Artificial Intelligence in Education*, Sydney July 20-24, 2003, pp. 279-286.
- [6] Avouris N., Dimitracopoulou A., Komis V., Margaritis M., "Participatory Analysis of Synchronous Collaborative Problem Solving Using the OCAF Methodology and Tools", *CSSL Interactive Event Proposal*, 2003. pp. 232-234.
- [7] May M., George S., Prévôt P., "A Web-based System for Observing and Analyzing Computer Mediated Communications", *Proceeding of the IEEE/WIC/ACM International Conference on Web Intelligence*, Hong Kong, 18-22 December 2006, pp. 983-986.
- [8] Smith A., "Invisible Crowds in Cyberspace : Measuring and Mapping the Social Structure of USENET", *In Communities in Cyberspaces : Perspectives on New Forms of Social Organization*, London, Routledge Press, 1999.
- [9] Takahashi M., Fujimoto M., Yamasaki N., "The active lurker: influence of an in-house online community on its outside environment", *Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work*, 2003.
- [10] Pozzi F., Manca S., Persico D., Sarti L., "A general framework for tracking and analysing learning processes in computer-supported collaborative learning environments", *Innovations in Education and Teaching International*, Volume 44, Issue 2, May 2007, pp. 169-179
- [11] De Wever, B., T. Schellens, M. Valcke, and H. Van Keer. "Content Analysis Schemes to Analyze Transcripts of Online Asynchronous Discussion Groups: A Review.", *Computers & Education* 46, no. 1, 2006, pp. 6-28.
- [12] May M., George S., Prévôt P., "Tracking, Analyzing, and Visualizing Learners' Activities on Discussion Forums", *proceedings of the 6th IASTED International Conference on Web-based Education (WBE 2007)*, Chamonix, France, 14-16 March 2007, pp. 649-656