Keeping Tracks of Users' Activities on Discussion Forum: Reflection and Proposal

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: *Forums for discussion are one of asynchronous communication tools that have been widely employed in all platforms of distance learning as important means for communications between learners, and between learners and instructors. Assisting forum users during and after their communications and providing them more useful information on their activities are our research interests. We present in this article our research works, which mainly focus on how to finely track the user's communication activity on the discussion forums. We state the potential elements of problems in keeping tracks of the user's activities while using the discussion forums and in exploiting the collected tracking data. We present next the approach for building the tracking system, applicable on different discussion forums. An example of how the collected tracking data are being analyzed and visualized in order to help the participants in the learning process, is also included.*

KEY WORDS: E-learning, discussion forum, tracking system, tracking learners' activities, exploitation of traces, and visualizing traces

1. Introduction

The tracking data of the learner's communication activities on discussion forum is known as a significant source of information, allowing the instructors to be aware of the activities that were carried through, along with their outputs. Each tracking data collected within learning environments is considered an important element, keeping the instructors informed of different learning aspects of learners, like the progress of the learners' activities [1] and the exchanged communications between the learners [2][3]. 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 [4]. The synthetic information derived from the analysis could as well help learners review their own behavioral aspects and that of others [5]. 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 needs of each user [6]. All of these references affirm the reasons why each trace acquired from learning environments, has obviously become more advantageous to various participants in the learning process, including, the teachers, the researchers and the developers.

Our research is particularly interested in designing and developing a generic system for tracking user's activities while using communication tools in their distance learning activities. The approach we present in this paper is for building tracking system for observing and analyzing users' activities on any type of discussion forum. Later in this section, we will focus on a study of principal elements of challenges in tracking user's activity in forums and exploiting tracking data. In section 2, we describe the different components of our approach. We present in section 3 a Web-based tracking system, followed by an example of visualizing traces.

1.1. Reflections on how tracking data are being captured and exploited

We discuss in this section the lack of functionalities of existing forum tracking systems and the ways traces are being exploited.

- **Log files and their difficulties :**

Log files are often used by forum developers to keep track of user's communications in forum. Data (traces) in log files are in general a series of events, dated, timed, and accompanied by other event attributes, which represent activities of a single or multiple users while interacting within the forum (i.e. display messages, post new messages on forum, etc.).

However, the traces stored in log files have been rarely exploited by the forums users (learner and instructor) either because of the ignorance of their existence, or because the traces do not match the exigency of the users. Moreover, the format of traces in a log file varies from one forum to another, due to the absence of standardization and the fact that each log file depends on how it was generated. This point often causes problems for the reusability and the exploitation of existing traces. Besides, there is a lack of semantic aspects for traces stored in log files (i.e. pure text log file). To manipulate traces in log files without specific tools, some operations such as searching, sorting and updating data, are constrained and laborious for non-computer specialist.

Therefore, creating and storing traces of communication activities should be done with strict specifications in accordance with the standards and rules so that the recorded traces can be accessible by the participants in the learning process, as well as the researchers and developers. As an example, the traces should be described and structured by using only extensible and explicit model of trace.

- **Capturing users' activities on forums :**

In order to track finely the users' communication activities on 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 on the forum. 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 interactions on the client side, which is collected when the user browses the forum without sending request queries to the server (editing messages, moving scrollbars, etc.).

Another remark we made about the existing tracking systems is that the activities of lurkers on forums have never been tracked down and finally none of the tracking data of the lurkers has been recorded. Even lurkers are sort of users who 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 [7]. The research works presented in [8], [9] are dedicated to studies of lurker's behavior on virtual communities (groups of discussion, newsgroup, Usenet,…), but not the traces of their participation in those communities.

One of the particularities of our research is tracking lurker's activities and studying their traces by hoping to address more convenient supports to forum users, especially the instructors so that they can be more aware of other user's activities in the forum, including lurkers'.

- **Storing the collected tracking data :**

We usually encounter in this stage some problems related to the choice of representing the traces and the repository type to store the traces. Since each choice was made just to match each individual need, traces of user's activities stored on existing forums is 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). 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.

For the tracking data that being collected from distinct sources (client and server), the synchronization process must be envisaged. A final trace will be the result of synchronized and structured tracking data. It is also needful to consider at this stage, the means 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 of representation, and reusable for other future uses.

- **Analyzing and visualizing tracking data :**

The tracking data are made up of certain information which is not directly interpretable by the forum users without the assistance of trace analysis and visualization tools. The major problems we are regularly facing in this stage are related to the effectiveness and the quality of the analysis. The treatment and analysis method used must be able to extract from traces the synthetic information that describes explicitly the activities carried through the communication activities on discussion forum (i.e. what exactly happens after this student has logged on to the forum? How to tell if a message has not been entirely displayed by a user? etc.).

Regarding the traces visualization, various visual forms of traces are to be considered: textual, histogram, or graph with multi dimension. The main idea is to allow the forums user to

visualize easily the results returned from the analysis. In order to facilitate users in visualizing traces, the visualization tools must be equipped with friendly Graphic User Interface 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.

1.2. Research objectives and methodology

We are interested in contemplating the possible ways for finely tracking the users' communication while using discussion forms in their distance learning activities. Based on the results of our first Web-based tracking system [10], which was specially developed for CONtextual FORum (CONFOR) [11],[12], we are able to identify more precisely the most fundamental elements of problem related to the collection and exploitation of user's tracking data while using discussion forums, like mentioned in the section above. Thus, the approach presented in this paper is to accomplish the following objectives:

- Keeping tracks of any sort of user, including lurkers, and on both client and server side,

- Define the granularity of traces (compilation of tracking data on server side and tracking data on client side) and handle necessary operations such as synchronizing, structuring, updating traces, etc.

- Represent collected traces in an extensive format so that they can be easily restructured and retransformed into another format of representation.

- Visualize traces with graphical representations.

- Support both instructors and students in their tasks of exploiting traces of users' activities on forum.

We have studied a number of discussion forums, the common points and the particularities of each forum. This allows us to propose tracking system architecture, which is applicable in different discussion forums. We have as well taken into account of AFNOR (French Association for NORmalizations, http://www.afnor.fr), about technical information on collaborative tools and data modeling in discussion forums. The choice of working with this standard allows us to:

    (i)   Identify all common functionalities used for communication within forums,

(ii)   Clearly define which user's activities to be tracked within their forum communications,

(iii)   Build our approach of modeling traces of user's activities, without being slanted to any particular discussion forum.

Regarding the system development, each system component was developed as independent modules, and each module can be integrated in different types of forums with a minimum modification in the existing forum codes. In our user study cases, we have created different use scenarios and tested them with different types of users. The results obtained from the tests were then used for evaluating our system.

2. Suggested approach

The figure 1 presents our suggested approach for building tracking system for forums and the chain of trace exploitation. The approach can be viewed in three levels:
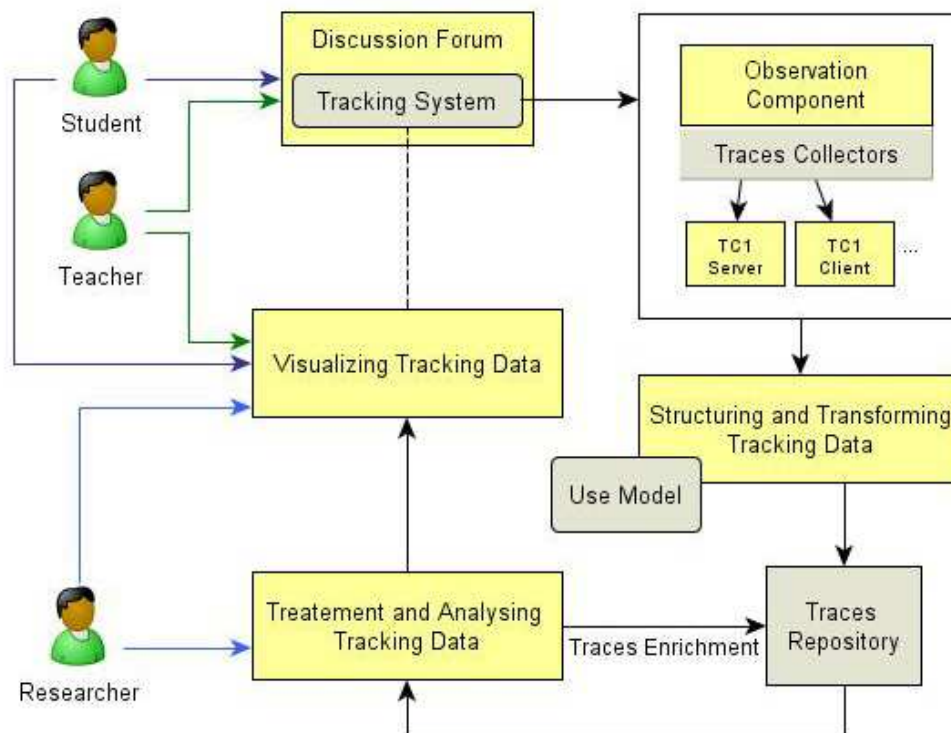

Figure 1. Overview of Suggested Approach for Building Tracking System for Discussion Forums

- First level is "**observing and generating tracking data**". The different **observation components** are specifically designed with a number of **Traces Collectors (TC)**, which take care of observing and capturing the user's activities on the forum. An observation

component describes how an activity on the forum can be performed by a user and how the trace collector generates instantaneously, the tracking data representing that activity. The trace collectors on the server side (**TC*i* server**) observe the activity of the communications occurring on the server. The trace collectors on the client side (**TC*j* client**) observe the interactions on user workstation (client internet browser). The raw tracking data generated by each trace collector will be next synchronized and transformed (**structuring and transforming tracking data**) into a structure, which conforms to a described **use model** (cf. part 2.1) before being stored in the trace repository.

- Second level is "**exploiting and enriching traces**". The **trace repository** presented in this level can be a database server or a collection of files that contain the structured traces submitted from the first level. The exploitation of the traces is defined by operations such as **treatment and analyzing traces,** and a series of **procedures for visualizing traces**. The trace **enrichment** operations we have mentioned earlier, are for restructuring traces by adding more information to their original representations; or transforming traces into another format, which can then be used in other exploitation tools. The synthetic information extracted from the analysis will pass through the pipeline of procedures for visualizing traces.

- Third level is "**visualizing traces**". Each module of this level is built with Graphic User Interface components for visualizing synthetic information provided by the second level. The visualization is also feasible in different visual forms, according to the parameters indicated within the information to be visualized, and independently of the visualization tools.

We give below the explanation of what a **use model** is, and for what purpose it is used. Next, we present how traces are being structured and stored in the **trace repository**.

2.1. Use model for user's activity on forums

A use 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. As shown in figure 2, a use 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.
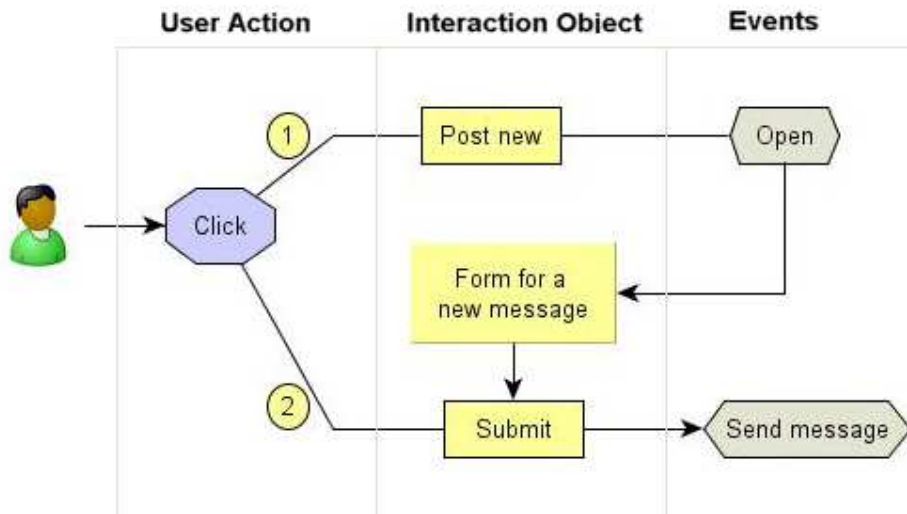
Figure 2. An example of a use model for an activity «*Post a new message*»

The arrow (1) represents a sequence of event 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 comprises several other interaction objects, including 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 interactions with those objects, and to produce simultaneously 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 interaction, and what to generate as tracking data.

2.2. Structuring Traces of User's Activity on Forum

Figure 3 presents the structure of traces being stored in the trace repository. We define a general structure of trace by an alternate sequence of **states** and **transitions**.
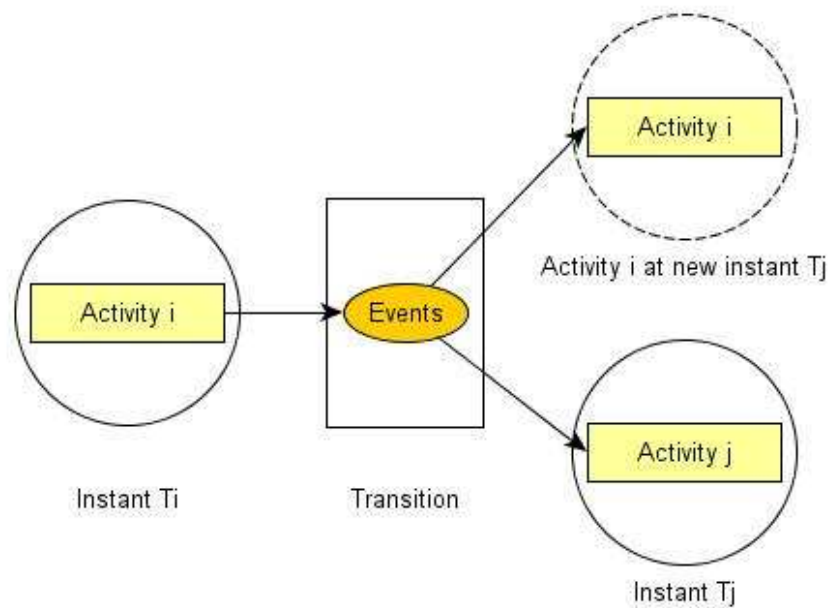
Figure 3. Structure of Traces of User's Activity on Forum

A **state** represents the instantaneous state of an activity on the forum. A **transition** results from one or many events produced by users' interactions during their activities on forum, which particularly make changes to the states of an activity from instant *Ti* to instant *Tj*. According to figure 3, at *instant Ti*, a user performs an *Activity i*, in which more than one event can be occurred. Within this activity, an event makes a transition and changes the current state of the *Activity i* to another state (at *instant Tj*), which can be either a new state of the same *Activity i* or a new state of another new *Activity j*.

Let us give an example of how we structure the traces of the activity «*Post a new message*». The event that represents the user's interactions on their browser, such as typing a 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. Such event does not make changes to the state of the current activity. But, if the user clicks on the «*Submit*» button, there is a transition: the written message is being submitted to the server. After the submission of the message, user is no longer at his/her current activity «*Post a new message*», but he/she is now at another state of another new activity, as «*Display posted message*» or «*Return to forum index*».

At each transition, 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.

2.3. Trace Repository

We chose the relational database MySQL to implement the trace repository. The choice of using a relational database for storing traces of user's activities on the forum has several advantages such as:

- Traces are structured in a rich format.
- Traces can be easily restructured and transformed into another format, such as TXT or XML.
- The operations for traces manipulation such as insertion, modification, etc., can be easily performed with simple SQL queries.

We present briefly below the five tables of the trace repository that contain the relational information used for structuring traces.

- Table **User**: In general, each forum database possesses a table that stores only the information of the forum subscribed members. In our case, the table **User** is used to store the information of every other user, such as guests without password or non identified users.
- Table **UserType** contains the information that represents user's profiles on forum (i.e. learners, instructors, etc.).
- Table **Activity** stores the information of the different activities of communication on forum (i.e. browse forum, exchange messages on forum, exchange private messages, etc.).
- Table **ActivityCategory** is for storing the information of the different categories of traces (i.e. traces of users' activities on forum, traces of users' interactions on remote workstation, etc.).
- Table **Transition** contains the final result of traces representing the user's activities in forum.

3. A Web-based tracking for contextual forum

3.1. Recorded tracking data of a use scenario

The system we have developed is a Web-based system offering support for the task of tracking a user's activity on a contextual forum. A number of tests were conducted with different use scenarios, which mainly represent communication activities of both active members and lurkers. We present below one among other use scenarios that consists of user's activities on server side as well as user's interactions on client side.

1. A user *Gregory*, logs on to the server, where the forum is hosted
2. He displays the *PhotoShop* forum
3. He clicks on the *Help* hypertext to read the help messages of forum
4. He clicks on a title of the message in the forum *PhotoShop* to read the message contents
5. He moves the vertical scrollbar downward to read a part of the message
6. He continues moving the vertical scrollbar downward for reading till the end of the message
7. He clicks on the *Reply* button to answer this message

IDTran ▲	User	Activity	Attribute	Date	Time	Delay	RefTran	Comments
1904	Gregory	Connection	login=Gregory	2006-06-12	02:05:10	NULL	0	NULL
1905	Gregory	Browse a course/forum structure	IDForum=4586	2006-06-12	02:05:22	NULL	0	NULL
1906	Gregory	Visit a hypertext	Help	2006-06-12	02:05:27	NULL	0	NULL
1907	Gregory	Display the content of message in forum	IDForum=4586,IDMsg=68	2006-06-12	02:06:07	00:00:41	0	NULL
1908	Gregory	Move scrollbar downward	IDForum=4586,IDMsg=68,VScrollbar	2006-06-12	02:06:18	NULL	1907	NULL
1909	Gregory	Move scrollbar downward (reaches the bottom)	IDForum=4586,IDMsg=68,VScrollbar	2006-06-12	02:06:49	NULL	1907	NULL
1910	Gregory	Reply to a message in forum	IDForum=4586,IDMsg=68	2006-06-12	02:07:02	NULL	0	NULL

Figure 4. Traces of Communication Activities Performed by a User *Gregory*

The traces of the given use scenario are stored in the table **Transition** of the trace repository, as pictured in figure 4. The "**Attribute**" and "**Delay**" properties are the two major compositions of traces. The "**Attribute**" property describes the attributes of traces associated to each individual communication activity. The "**Delay**" property represents the total time that the user spent on an activity. As an example, the trace *IDTran = 1907* explains that user *Gregory* displayed a message *IDMsg = 68*, in the forum *IDForum = 4586*. This user spent 41 seconds to display (or probably read) this message before performing another activity.

From the trace *IDTran = 1909*, we are able to know that the user moved the vertical scrollbar till the bottom of the message. This information can be very substantive for the automatic analysis and to help the instructors make better assumptions on user's activity while reading a message in forum. For example, after displaying rapidly the message and if user has not touched or moved the vertical scrollbar downward the page, we conclude that the displayed message has not been entirely read. Nonetheless, this assumption can not be made in the opposite way. Even after displaying the message for sometimes long enough and the vertical scrollbar has been moved downward (and to the bottom of the page), we can not justify that the user has read the entire message.

3.2. Visualizing Traces

Getting started with the real needs of the instructors, the tools for visualizing traces are developed with friendly Graphic User Interface, allowing both instructors and learners to easily interrogate the trace repository and to view the traces in graphical representations.

« How do we know whether or not a displayed message is read? ». This question has been frequently asked, particularly by the instructors who regularly use discussion forums in their teaching activities. Although we can not really prove if a message was read by the user who displayed it, we can give some indicators to show that a message has never been entirely read. The given example below aims to offer better ways for visualizing traces and necessary elements for making hypothesis on «*Reading a message in forum*».

Figure 5 shows an example of visualizing trace in graphical representation. It presents the traces of activity « *Reading a message in forum* » of every users who have read the message *IDMsg = 25* on *May 31, 2006*. The main objective of this visualization is to find if the users have displayed the message contents, if they have done action on it, and for how long. Each sphere shown in figure 5 represents an activity of reading a message and the diameter of the sphere is proportional to the time spent by each user for reading the displayed message. The distance between two spheres represents the time gap between two different readings. A sphere can be in one of the three colors: orange, blue, or green.

The green sphere notifies that the user read the message by having moved the vertical scrollbar downward, and to the bottom of the page (reading till the end of the message). The orange sphere expresses the fact that the user has only displayed the message contents without touching the scrollbar. The blue sphere signifies that the user has displayed the message contents, by moving the vertical scrollbar downward, but not moved to the bottom of the page (partial reading).

The label « ID Message: 25 | Users: All | Date: 2006-05-31 » located at the top of the graph means that the trace being visualized refers to all the traces of every user who read the message number 25 on May 31, 2006. The labels « Start » and « End » indicate that on May 31st, 2006, the message number 25 has been read only between 11:12:51 and 23:22:44. With this view, when the user moves the mouse cursor over the sphere, an information tag fades in with brief information corresponding to the reading activity. If user clicks on the sphere, a pop-up window is displayed to detail the information of the reading.
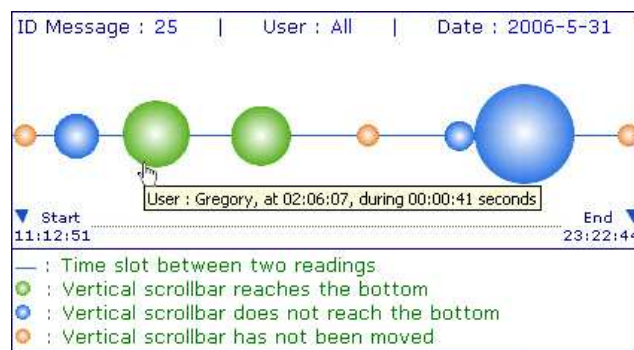

Figure 5. Visualizing Traces of *« Reading a message in forum »*
in Graphical Representation 1

The given example is only one of the possibilities in visualizing traces in graphical representation. The same traces can be of course visualized in other forms as well as in different scale. Figure 6 shows another visual form of the same traces, except here the vertical line presents the reading activity of a user and the time that he/she spent on it.
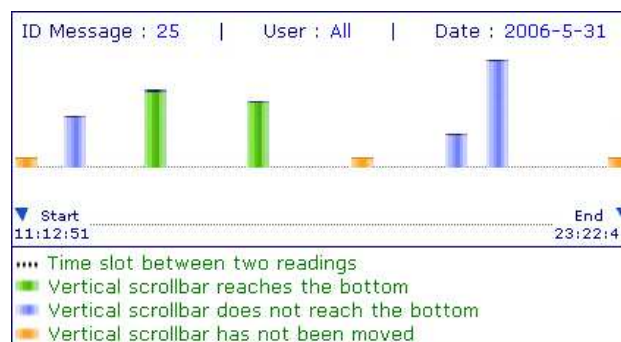

Figure 6. Visualizing Traces of *« Reading a message in forum »*
in Graphical Representation 2

4. Conclusion and future work

There is a variety of discussion forums used in Web-based learning environments, thus building a tracking system for every single forum is not a wise solution. That is why we have studied the common points and the particularities of each tool and to propose tracking system architecture, which is applicable in different discussion forums. The use model we presented in section 2.1 permit us to identify what to observe and how, as well as what to generate as tracking data to represent the communication activity on discussion forums. We have stated as well the potential elements of problems in tracking user's activities on discussion forums, which can be used as starting points for other research works, particularly those that are involved in exploiting the tracking data to help the participants in the distance learning process. Besides, our Web-base system that we present in the section 3, affirms the feasibility of a system that is capable to track finely any kind of user and their interactions on both client and server side.

In the forthcoming work, we wish to apply the contents analysis methods in order to analyze the contents of users' communications in the forum. We aim to make our system more competent in extracting from traces both quantitative and qualitative information, which reflects more behavioral, social and cognitive aspects of learners. Regarding the future tracking system, we expect to develop it in the form of a package of functions, which can be integrated in the existing distance learning platforms with limited technical skills.

5. References

[1] 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.

[2] 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.

[3] Donath J., Karahalios K., Viegas F.B., "Visualizing Conversation", *Proceedings of the 32nd Annual Hawaii International Conference on System Sciences, Vol. 4(4)*, 1999.

[4] 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.

[5] 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.

[6] Avouris N., Dimitracopoulou A., Komis V., Margaritis M., "Participatory Analysis of Synchronous Collaborative Problem Solving Using the OCAF Methodology and Tools", *CSCL Interactive Event Proposal*, 2003. pp. 232-234.

[7] 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,

[8] Whittaker, S., Terveen, L., Hill, W., Cherny, L., "The dynamics of mass interaction", *Proceedings of the Conference on Computer Supported Cooperative Work (CSCW 98)*, Seattle, WA, 1998, pp. 257-264.

[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] May M., George S., Prévôt P., "A Web-based System for Observing and Analyzing Computer Mediated Communications", *proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence,* Hong Kong, 18-22 December 2006.

[11] George S., "Contextualizing Discussions in Distance Learning Systems", *Proceedings of the 4th IEEE International Conference on Advanced Learning Technologies (ICALT 2004)*, Joensuu, Finland, August 30-September 1, 2004, pp. 226-230.

[12] George S., Labas H., "Fostering discussions with forum tools based on e-learning standards", *IEEE and IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2005)*, Porto, Portugal, December 14-16, 2005, pp. 60-68.