CIDUI International Congress of University Teaching and Innovation 2010 Barcelona

NEW AREAS OF QUALITY IN HIGHER EDUCATION
A comparative and trend analysis

ASSESSING STUDENT ACTIVITY THROUGH LOG ANALYSIS FROM COMPUTER SUPPORTED LEARNING ASSIGNMENTS

De Pedro Puente, Xavier [1]; Calvo, Miquel [2]; Carnicer, Artur [1]; Cuadros, Jordi [3]; Miñarro, Antoni[1]

[1] Dept. Física Aplicada i Òptica, Universitat de Barcelona

[2] Dept. Estadísica, Universitat de Barcelona

[3] IQS, Universitat Ramon Llull

E-mail: xavier.depedro@ub.edu

1. SUMMARY

The authors of this communication developed tracking-enabled interactive programs to be used in computer-supported learning assignments adapted to the EHEA context in Statistics and Physics courses. Trace files (logs) generated from the student activity can be processed by simple scripts in order to allow teachers to acquire information about the learning process of each single student. This information would comprise work dynamics, activity path followed, learning problems (either conceptual or methodological), resolution attempts with inappropriate procedures or alternative successful resolutions. Hopefully, the methodology will allow us to identify common patterns in student learning paths, typical errors and alternative procedures of resolution in open assignments where several valid solutions are possible.

2. KEYWORDS

Educational Technology, Student Activity, Trace Analysis

3. RESUMEN

Los autores de esta comunicación han desarrollado programas interactivos habilitados para trazar la actividad de los estudiantes, de cara a utilizarlos en ejercicios de aprendizaje asistidos por ordenador en cursos de Estadística y Física, adaptados al contexto del EEES. Los archivos de trazas (logs) generados a partir de la actividad de estudiantes pueden ser procesados por programas sencillos que permiten al profesorado adquirir información sobre el aprendizaje de estudiantes individuales. Esta información incluiría la dinámica de trabajo, el recorrido seguido en la actividad, los problemas de aprendizaje (ya sean conceptuales o metodológicos), los intentos de resolución por procedimientos inapropiados o procedimientos alternativos válidos de resolución. Es de esperar que la metodología nos permitirá identificar los patrones comunes en los caminos de aprendizaje, los errores típicos y los procedimientos alternativos de resolución en ejercicios de tipo abierto en el que son posibles varias soluciones válidas.

4. PALABRAS CLAVE

Tecnología educativa, Actividad del Estudiante, Análisis de Trazas

5. INTRODUCTION

The implementation of the European Higher Education Area (EHEA , or EEES in Spanish or Catalan) asks for more autonomous work of the students but at the same time looks for more guidance and assessment. This research plans to meet this seemingly opposite goals in the case of highly interactive on-line applications. Building upon the ideas on the logging of the common on-line services in many web applications, we are developing a tracing methodology that will allow us to capture students on-line work and automatically analyze it to get useful assessment information.

The EHEA involves deep changes in the teaching methodology at university level (European Commission 2008). For the faculty, some of these changes could mean less face-to-face teaching activities, and more guided and autonomous work for students. In this new scenario, new tools and strategies are needed to provide the faculty with information about students' activity and performance. Some problems involve using computer-supported learning assignments, which enhance student autonomy along their learning process. This could help to move the student to the center of the teaching and learning process in a coherent way with the reform process associated with EHEA. Consequently, tracking the student activity in logs is needed to allow further feedback from students about their work on-line.

Some of the early tools to track student activity are, beyond the web server logs, the ones provided by, or compatible with, the Course Management Systems available at our teaching institutions, like GISMO for Moodle (Mazza & Milani 2004) or others (Hijon & Velazquez 2006).

Those programs often include basic tracking features to write a log of student activity across a few of the activities and resources being used by students, and a few of them specialized in writing more comprehensive action logs adapted to innovative pedagogical scenarios. That's the case of Analytic Toolkit for Knowledge Forum (Burtis 1998, Scardamalia 2004), or the Action Log & Contribution features in Tiki (De Pedro 2006, 2007), while being both of them of general scope to be used in constructionist pedagogical scenarios (Alier & Barceló 2005), and not necessarily attached to any specific assignment or course.

On another side, some initiatives have been recently deployed where tracking systems have been designed to research how learning occurs when using on-line applications (IMMEX 2008, PSLC 2008, Stevens et al 2004, Feng & Heffernan 2005, Feng et al. 2005) grounding on the microgenetic method proposed by Siegler & Crowley (1991). The usage of this type of resources is usually linked to b-learning or e-learning activities (Garrison & Kanuka, 2004), which often makes more difficult the process of following and assessing student activity. Therefore, the faculty work to manage, facilitate and assess the student learning process becomes more difficult. That is why very special care has to be taken when designing such type of computer-supported learning assignments, so that the log files of student activity can provide to the faculty comprehensive information related to the outcomes that we are promoting the students to acquire with such autonomous activities.

6. COMPUTER-SUPPORTED LEARNING ASSIGNMENTS

The groups behind this communication involve faculty from the Statistics and the Optics and Applied Physics departments, both at the University of Barcelona, and faculty from the Institut Quimic de Sarrià, at University Ramon Llull. The experience described here corresponds to courses taught at the University of Barcelona. Two different tracking-enabled computer applications have been designed: (1) Statmedia II (Arcas et al. 2009), already used courses in statistics with large groups of undergraduate students in Biology and (2) an optical tweezers simulation application (Mas et al. 2010, Carnicer at al. in press), to be used with small groups of senior or graduate students in Physics or Photonics.

Learning assignments related to the Statmedia II application comprise a practical case in two web forms that can be interacted through standard Internet browsers. Students have to answer a series of applied questions related to the biological context were the experiment is framed. The assignment requires knowledge of most of the statistical techniques studied along the course, but it does not explicitly indicate at any time which is the right one to use. Moreover, all activity is individualized, so that students have to work on the same experimental scenario, but their data and, possibly, conclusions, will be different for each one.

Figure 1. Application 1 (Stamedia II), showing one of the frames, with the statistics calculator open

The learning assignment corresponding to the optical tweezers application requires the student to set the appropriate values for many variables which comply with particle trapping within the laser beams. Students have the freedom to design the whole experimental system (wavelength, focal distance, system overfilling, etcetera) as well as the values for the environment conditions (temperature, viscosity, and many other). This is an open type of assignment, so many combination of variables and values can be valid solutions. Therefore, an automatic assessment of this type of assignment is not trivial.

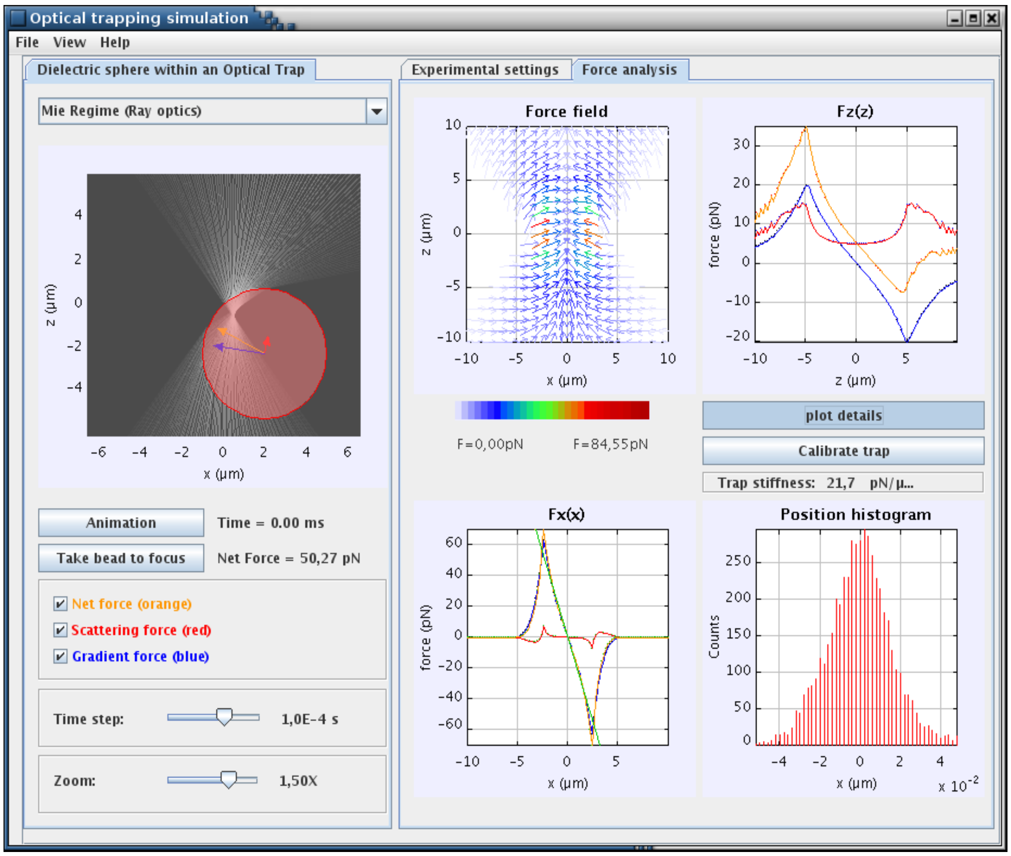


Figure 2: Application 2 (Optical tweezers) showing the analysis interface

7. XML-BASED LOG FILES

To facilitate this assessment, both applications are set to optionally trace the students' activity. The logs recorded by both pieces of software include, among others:

1. Identification of the user and the application being used,

2. the action performed by the student, including the objects that the student interacted with, or modified, and any significant system variables at the interaction time (what is shown on the screen, values of specific parameters of relevance, time stamp, ...)

3. session identification (students may take several visits to complete one assignment)

4. results obtained after the interaction, including the answers to questions or input fields in forms, if any, and the associated right answers, if unique solutions were expected. The goal is that these action logs contain enough information to reproduce the interaction process performed by the student with the program, so that some relevant information about the teaching and learning process can be inferred.

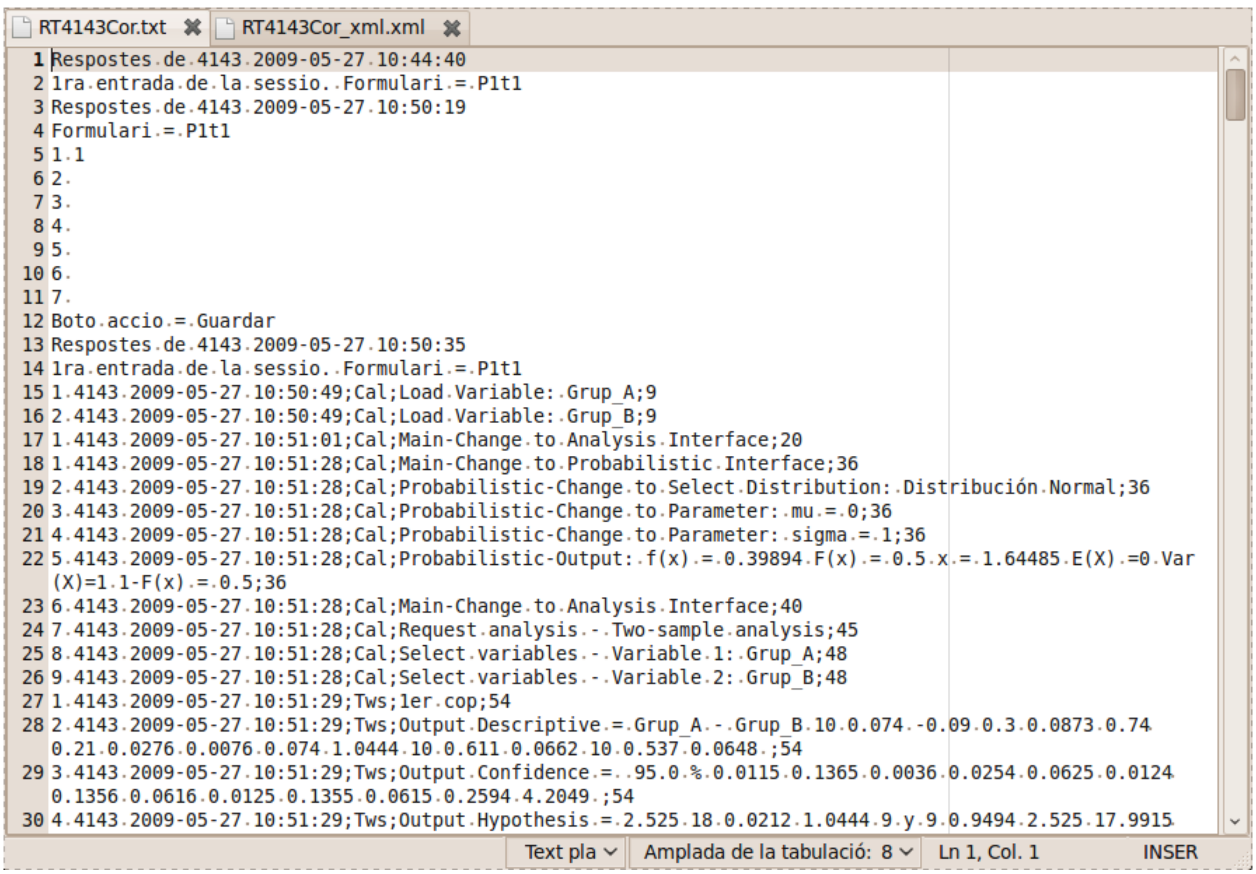
XML format was the chosen format to structure the information in the log files prior to further common processing. This ensures a common format regardless of the application where the information came from. This XML format comprises besides the general XML definition tags (<xml> and <!DOCTYPE>) the following structure:

- A <log> node, which is the root node and encloses the whole data file and refers to the tracing information that has been collected. Its only children nodes are <event>.

- A list of <event> nodes, children of <log>, that captures the events taking place in the interface of the application. For each of these events, the main details are recorded as attributes: application name, action type, student name, session number, date and time, and correlative number of action. Specific information referring to a type of event will be included as children <param> nodes.

- <param> nodes are children of <event> and allows for tracing specific information for specific actions in the application. Since different action will need different additional information to be fully qualified, this information is coded as parameters with a name and a value as unique attributes. For instance, a tag of this type may contents the parameter value of the statistical calculator that the student chose for an analyses, etc.

- Also a <description> node can be included in each <event> tag. This encompasses the basic information of the event action and its parameters but in natural language, so that it can be easily extracted later on by the reporter script to produce meaningful text for humans that want to review the student activity (student himself, professor, ...).

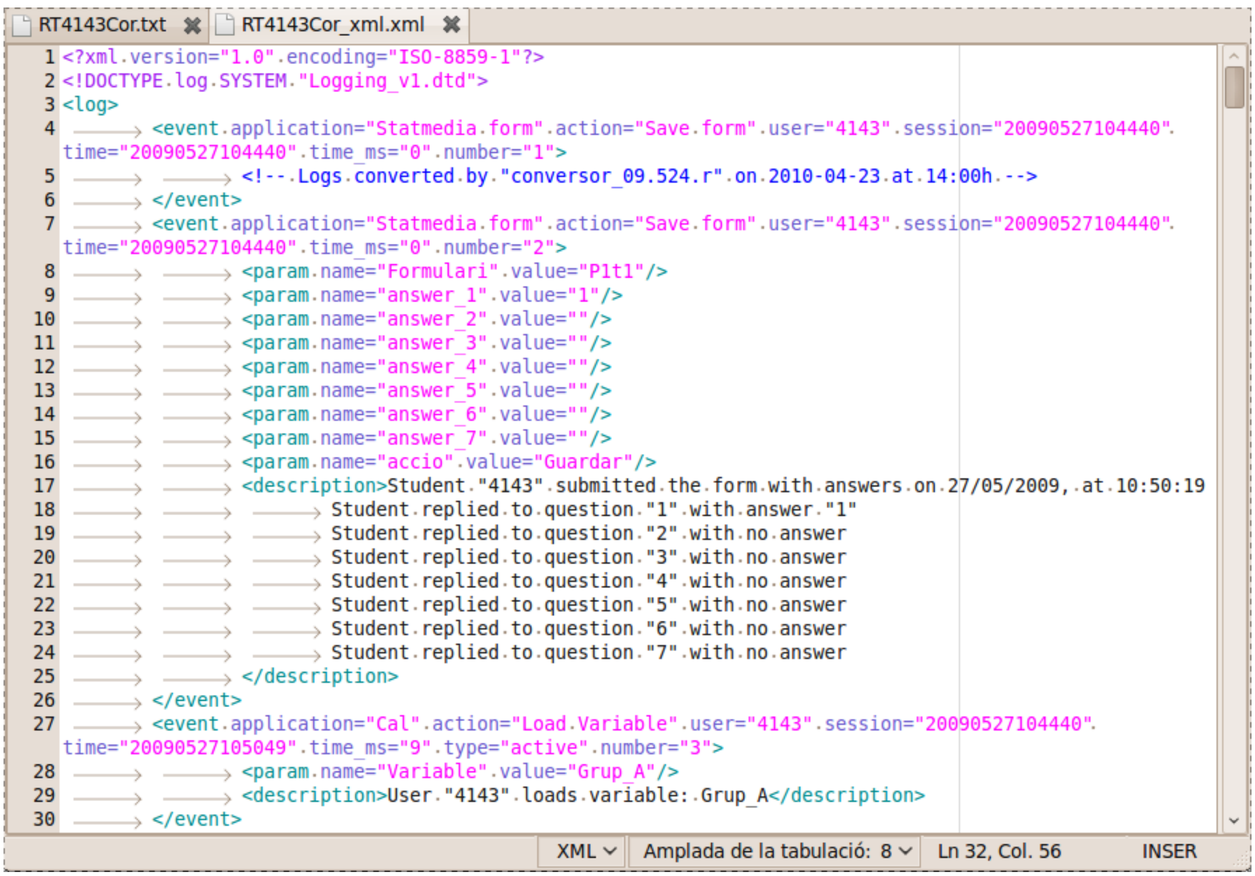

RT4143Cor.txt
RT4143Cor_xml.xml

```
Respostes de 4143 2009-05-27 10:44:40
1ra entrada de la sessio  Formulari = P1t1
Respostes de 4143 2009-05-27 10:50:19
Formulari = P1t1
1 1
2 
3 
4 
5 
6 
7 
Boto accio = Guardar
Respostes de 4143 2009-05-27 10:50:35
1ra entrada de la sessio  Formulari = P1t1
1 4143 2009-05-27 10:50:49;Cal;Load Variable: Grup_A;9
2 4143 2009-05-27 10:50:49;Cal;Load Variable: Grup_B;9
1 4143 2009-05-27 10:51:01;Cal;Main-Change to Analysis Interface;20
1 4143 2009-05-27 10:51:28;Cal;Main-Change to Probabilistic Interface;36
2 4143 2009-05-27 10:51:28;Cal;Probabilistic-Change to Select Distribution: Distribución Normal;36
3 4143 2009-05-27 10:51:28;Cal;Probabilistic-Change to Parameter: mu = 0;36
4 4143 2009-05-27 10:51:28;Cal;Probabilistic-Change to Parameter: sigma = 1;36
5 4143 2009-05-27 10:51:28;Cal;Probabilistic-Output: f(x) = 0.39894 F(x) = 0.5 x = 1.64485 E(X) =0 Var
(X)=1 1-F(x) = 0.5;36
6 4143 2009-05-27 10:51:28;Cal;Main-Change to Analysis Interface;40
7 4143 2009-05-27 10:51:28;Cal;Request analysis - Two-sample analysis;45
8 4143 2009-05-27 10:51:28;Cal;Select variables - Variable 1: Grup_A;48
9 4143 2009-05-27 10:51:28;Cal;Select variables - Variable 2: Grup_B;48
1 4143 2009-05-27 10:51:29;Tws;1er cop;54
2 4143 2009-05-27 10:51:29;Tws;Output Descriptive = Grup_A - Grup_B 10 0.074 -0.09 0.3 0.0873 0.74
0.21 0.0276 0.0076 0.074 1.0444 10 0.611 0.0662 10 0.537 0.0648 ;54
3 4143 2009-05-27 10:51:29;Tws;Output Confidence =  95.0 % 0.0115 0.1365 0.0036 0.0254 0.0625 0.0124
0.1356 0.0616 0.0125 0.1355 0.0615 0.2594 4.2049 ;54
4 4143 2009-05-27 10:51:29;Tws;Output Hypothesis = 2.525 18 0.0212 1.0444 9 y 9 0.9494 2.525 17.9915
```

Text pla | Amplada de la tabulació: 8 | Ln 1, Col. 1 | INSER

Figure 3: Sample log file from application 1 (Statmedia II), showing the first portion of it from the student activity directly recorded from the program in plain ASCII file.

The main difference between both applications regarding the logs is that the first one (Statmedia II) records student activity data in a plain ASCII file (Figure 3), and from there it is converted into an XML file (Figure 4) by means of an R script. Instead, the second application (Optical Tweezers applet) saves directly in an XML file with the designed common structure (the same type as shown in Figure 4). This Optical Tweezers applet was originaly designed to generate XML traces whereas the Statmedia II applications, developed some time ago, have adapted their tracing files to that common XML structure.




```
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE log SYSTEM "Logging_v1.dtd">
<log>
	<event application="Statmedia form" action="Save form" user="4143" session="20090527104440" time="20090527104440" time_ms="0" number="1">
		<!-- Logs converted by "conversor_09.524.r" on 2010-04-23 at 14:00h -->
	</event>
	<event application="Statmedia form" action="Save form" user="4143" session="20090527104440" time="20090527104440" time_ms="0" number="2">
		<param name="Formulari" value="P1t1"/>
		<param name="answer_1" value="1"/>
		<param name="answer_2" value=""/>
		<param name="answer_3" value=""/>
		<param name="answer_4" value=""/>
		<param name="answer_5" value=""/>
		<param name="answer_6" value=""/>
		<param name="answer_7" value=""/>
		<param name="accio" value="Guardar"/>
		<description>Student "4143" submitted the form with answers on 27/05/2009, at 10:50:19
			Student replied to question "1" with answer "1"
			Student replied to question "2" with no answer
			Student replied to question "3" with no answer
			Student replied to question "4" with no answer
			Student replied to question "5" with no answer
			Student replied to question "6" with no answer
			Student replied to question "7" with no answer
		</description>
	</event>
	<event application="Cal" action="Load Variable" user="4143" session="20090527104440" time="20090527105049" time_ms="9" type="active" number="3">
		<param name="Variable" value="Grup_A"/>
		<description>User "4143" loads variable: Grup_A</description>
	</event>
```


Figure 4: Converted XML file obtained form the Statmedia II plain-text log file shown in Figure 3.

These XML log files are then processed to report basic data of interest in a human readable format by means of some other small scripts in "R" language. "R" is a free software statistics program (http://www.r-project.org) which makes relatively easy to manage any type of data files, including plain text or XML files, on the most common operating systems for servers and desktops (GNU/Linux, Mac OSX, MS Windows, ...). It works based on scripts than can be easily created or modified to produce tasks, and there are a few Graphical User Interfaces for some of the most common tasks with it (De Pedro 2009). Moreover, the scripting syntax allows a similar powerful environment for programing custom programs with minimum code (Bronken 2006), combined with the power of a large community of researchers and developers worldwide improving the software and highly specialized custom scripts month after month with free/libre open source licenses shared with the community.

In order to process XML files and generate a report, a couple of extra R packages have to be installed to standard installations of the R program. They comprise the packages: "XML" (from http://www.omegahat.org/R) and "R4X" (from http://R-Forge.R-project.org).

8. PROCESSING LOG FILES: CONVERSION TO NATURAL LANGUAGE.

In a later stage, the description fields (the text between the <description> and </description> tags shown in Figure 4) are extracted from the XML file generated and they can be used as a report in natural language potentially for both the student and the professor. This way students will be able to review the process that he/she followed, and learn later on more easily from that process, and professors, will have some additional material to assess student progression and performance with the autonomous learning activities by the students.

The report contains basic information on top (before the description of all actions in natural language), such as the number and type of interactions performed in each session, a summary of the data used, the number of objects that the student has interacted with, the time invested, among others (Figure 5). So far it is stored also in a CSV file format (comma separated values), so that any part of the report can be imported in common Office suite programs in order to allow quoting specific parts of the report as tables, if desired. This report can be extended progressively as needed fairly easy, through adding more queries based on regular expressions to the R script that generates it.

```
"Basic report on the xml file"
"-----------------------------"
"nodeCounts.param", 70
"nodeCounts.event", 30
"nodeCounts.log", 1
"nodeCounts.description", 1
"attributes.event.application", 30
"attributes.event.action", 30
"attributes.event.user", 30
"attributes.event.session", 30
"attributes.event.time", 30
"attributes.event.time_ms", 1
"attributes.event.number", 30
"attributes.param.name", 70
"attributes.param.value", 70

"Basic report on student activity"
"--------------------------------"
"Number of Calibrations", 1
"Number of Animations" ,2

"Calibrations"
"------------"
"Parameters at the Calibration Number", 1
        "power", 0.05959999999999999
    "Wavelength", 5.550000000000001E-7
```

Figure 5: A section of the report generated from a log file.

The tracking system in the Statistics courses has been used twice in a real classroom context: 160 enrolled students, from which 120 regularly assisted to face-to-face classes. Those students were offered the chance to voluntarily participate, provided that results would be kept anonymous, and that they could improve their grade if they successfully participated in the experience. The system generated around 1,7 MB of data in plain text log files regarding 52 students, which has been used later to fine tune the common log file markup in XML among all faculty involved.

The tracking system from optics-related courses also comprised a brief collection of problems that were used in October 2009 with last year students of Physics. Although the participation has been smaller (20 students), this data set will be enough to start a systematic log analysis, and new data sets from graduate students (masters in Nanoscience and Photonics) are expected.

9. CONCLUDING REMARKS

Two tracking systems were adapted to record student activity in logs using a common XML markup. The initial analysis of some preliminary logs provides a simple report to the faculty in a human readable format.

Moreover, logs from student activity in real classroom contexts have been collected from several courses at university level. They are going to be processed in the near future, in order to shed light about the learning paths followed by the students, common problems, alternative procedures of resolution and ways to improve the computer-supported learning assignments.

10. BIBLIOGRAPHICAL REFERENCES

Alier, M.; Barceló, M. (2005). An application of social constructionism and open source, in Mendez Villas, B.; González Pereira, J.; Mesa González, J.A. Recent research developments in tearning technologies, pp. 1170-1175.

Arcas, A; Miñarro, A.; Calvo, M. (2009). "Statmedia Projects in Statistical Education". Boletín de Estadística e Investigación Operativa.Vol. 25, No. 1,pp. 42-56.
http://www.seio.es/BEIO/files/BEIOv25n1_HE_A.Arcas+A.Minarro+M.Calvo.pdf

Bronken Gundersen, V. (2006) MATLAB commands in numerical Python (NumPy). pp. 17. http://mathesaurus.sourceforge.net/matlab-python-xref.pdf

Burtis, J. (1998). Analytic Toolkit for Knowledge Forum. Centre for Applied Cognitive Science, The Ontario Institute for Studies in Education/University of Toronto.

Carnicer, A. (Coord.) (2006). Curso de Óptica en Java (JavaOptics). http://www.ub.es/javaoptics/index.html

Carnicer, A.; Mas, J.; Cuadros, J.; Juvells, I.; de Pedro, X. (In press, 2010). Seguimiento del aprendizaje mediante análisis de trazas en una simulación de trampas ópticas, Opt. Pura Appl.

De Pedro, X. (2006). Estimulación y evaluación del aprendizaje 'experiencial-reflexivo' del alumnado mediante la formulación explícita del tipo de contribuciones. In Grupo de trabajo sobre Aprendizaje Reflexivo, del Instituto de Ciencias de la Educación, Universidad de Barcelona (pp. 1 - 100). http://www.ub.edu/gclub/dl55

De Pedro, X. (2007). New method using Wikis and forums to evaluate individual contributions in cooperative work while promoting experiential learning: results from preliminary experience. In Wikis at Work in the World: Open, Organic, Participatory Media for the 21st Century (pp. 87 - 92). ISBN: 978-1-59593-861-9. http://www.wikisym.org/ws2007/_publish/dePetro_WikiSym2007_ContributionsInCooperativeWork.pdf

De Pedro, X. (2009). Colaboración entre PDI (1): Estadística y gráficos con R. VIII Jornadas de Software Libre, Universidad Politécnica de Cataluña. Barcelona, SPAIN. http://www.ub.edu/gclub/jornadespl2009

European Comission (2008). Education and Training.
http://ec.europa.eu/education/index_en.html .

Feng, M.; Heffernan, N. (2005). Informing Teachers Live about Student Learning: Reporting in Assistment System. 12th Annual Conference on Artificial Intelligence in Education.
http://www.learnlab.org/uploads/mypslc/publications/reportingaied.pdf

Feng, M.; Heffernan, N.; Koedinger, K. (2005). Looking for Sources of Error in Predicting Student's Knowledge. Proceedings of AAAI 2005 workshop on Educational Data Mining.
http://www.learnlab.org/uploads/mypslc/publications/reportingaaai.pdf

Garrison, R.; Kanuka, H. (2004). Blended learning: uncovering its transformative potential in higher education. The Internet and Higher Education, 7, pp. 95-105.

Hijon, R.; Velazquez, A. (2006). E-learning Platforms Analysis and Development of Students Tracking Functionality. In E. Pearson & P. Bohman (Eds.), Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications (pp. 2823-2828). Chesapeake, VA: AACE.

IMMEX (2008). IMMEX. http://www.immex.ucla.edu/

Mas, J.; Farré, A.; Cuadros, J.; Juvells, I.; Carnicer, A. (2010). Understanding Optical Trapping Phenomena: a Simulation for Undergraduates, IEEE TRANSACTIONS ON EDUCATION.
http://dx.doi.org/10.1109/TE.2010.2047107

Mazza, R.; Milani, C. (2004). Gismo: a graphical interactive student monitoring tool for course management systems. In T.E.L. Technology Enhanced Learning 04 International Conference.

PSLC (2008). Welcome : Pittsburgh Science of Learning Center. http://www.learnlab.org/

Scardamalia, M. (2004) CSILE/Knowledge Forum, in Education and technology: An Encyclopedia, pp. 183-192.
http://www-personal.umich.edu/~jaylemke/courses/ED750/CSILE_KF_illus.pdf

Siegler, R.S.; Crowley, K. (1991). The microgenetic method. A direct means for studying cognitive development. Am Psychol. 1991 Jun;46(6):606-20.

Stevens, R., Soller, A., Cooper, M., and Sprang, M. (2004). Modeling the Development of Problem Solving Skills in Chemistry with a Web-Based Tutor. Intelligent Tutoring Systems. Lester, Vicari, & Paraguaca (Eds). Springer-Verlag Berlin Heidelberg, Germany. 7th International Conference Proceedings (pp. 580-591).
http://www.immex.ucla.edu/docs/publications/pdf/its_paper.pdf