A Web Analytics and Visualization Tool to Understand Students' Behavior in an Adaptive E-Learning System

Barbara Moissa, Lucas Simões de Carvalho, and Isabela Gasparini

Department of Computer Science
University of Santa Catarina State (UDESC), Joinville, SC, Brazil
{barbara.moissa,lucas.scarvalho}@gmail.com,
isabela.gasparini@udesc.br

Abstract. Web analytics in the learning environment is the use of intelligent data, learner-produced data, and analysis models to discover information and social connections. This paper describes a web analytics and visualization tool which collects, analyses and visually represents the data collected in an e-learning environment. The goal of this tool is allow the teacher to better understand his students' behavior in front of the environment and support the decision making related to the pedagogical content adapted to the students' need.

Keywords: learning analytics, information visualization, adaptive e-learning system.

1 Introduction

Web Analytics is the measurement, collection, analysis and reporting of internet data for purposes of understanding and optimizing web usage [5]. It helps to identify popularity trends, to estimate how traffic to a website changes and also provides information about the numbers of visitors and the number of page views.

In the context of e-learning, it is usually named learning analytics, and it helps the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs. Web analytics in the learning environment is the use of intelligent data, learner-produced data, and analysis models to discover information and social connections. These gathered data can help teachers in discovery and understand students' behavior in their courses, in the collaborative tools provided by the system, and the system could predict and advise people's learning [10].

One problem faced by teachers who use e-learning environments is to understand the students' needs. One way to deal with this issue is to collect data about the students' actions and interactions with the e-learning environment and analyze it to gather useful information and to present the best contents adapted to students' needs and profile.

Although web analytics tools can help teachers discovering students' behaviors and trails, some tools are difficult to understand and to extract information, because their

P. Zaphiris and A. Ioannou (Eds.): LCT 2014, Part I, LNCS 8523, pp. 312–321, 2014.

diversity of the metrics or their poor usability. Visualization information techniques can be applied in these tools, helping user to understand the data generated. Information visualization (InfoVis) produces visual representations of abstract data to reinforce human cognition; thus enabling the viewer to gain knowledge about the internal structure of the data and causal relationships in it [6].

In this work, we describe a tool for mining and visualizing student's interaction to an adaptive web-based educational system that we have developed in order to help teachers/instructors to discover student's behavior.

This paper is structured as follows. In the section 2 we present the related works. Section 3 introduces the adaptive e-learning system where the tool was integrated. Section 4 presents the functionality of the tool to mine and analyze student´s interaction and behavior, how the students' data is collected and the defined metrics. Section 5 describes the visualization tool, how the defined metrics are converted into a graphical and interactive visualization tool and how teachers can benefit from it. In section 6 experimental results obtained are then discussed, showing how the experiment was planned and conducted and also the interpreted results. Finally, in Section 7, conclusions and future work are presented.

2 Related Work

Learning Analytics intends to collect data about and produced by learners in order to use it to enhance education and learners' success. These data can be used in several ways and with several different reasons. The Open Learning Analytics platform [11], for example, is focused in four groups: learners, educators, administrators and researchers. While learners view data regarding their progress in the course, educators view information that can help them to know what has impacted on the learners' engagement; administrators view statistics about the courses and researchers can use this data to discover new information.

While the Open Learning Analytics platform has focus on four distinctive groups, the Student Activity Meter (SAM) [4] focuses only at the learner and the teacher. The tool aids learners to see how they spent their time and also recommend materials to them. In the other hand, teachers can see what and how students are doing, where they are spending more time and which are the most popular materials.

There are also tools as eLAT [1] that focuses only on the teachers and aims to process all the data quickly considering privacy aspects and individual interests. This tool enables teachers to explore and correlate the learning object usage, user character, user behavior and assessment results using a graphical interface and, by doing this, they can reflect and identify opportunities for improvement.

Our proposal has similarities with these works, since we focus on helping teachers to better understand their students' behavior in the system. In the future, we also intend to recommend some action to the teachers (e.g. improve their communication

with the students, changes in his navigational paths, etc.) and to the students (e.g. recommendations about their engagement, works and tasks, etc.) based on the learning styles, patterns and behaviors discovered.

3 AdaptWeb

AdaptWeb (Adaptive Web-based learning Environment) is an adaptive e-learning environment whose purpose is to adapt the content, presentation and navigation in an educational Web course, according to the student's profile. Currently, it is an open source environment in operation in different universities. This adaptation is based in the student's profile, such as his knowledge about a certain concept, his navigational preferences and his background. Another feature of the environment is the ability to offer a discipline to more than one course. For example, the same Calculus discipline can be offered to the courses of Computer Science, Electrical Engineering and Civil Engineering with its content properly adapted.

In AdaptWeb, the student has access to three sections: Class Environment, Messages Board and Discussion Forums. Each one of these three sections is part of an offered discipline in a certain course. In other words, the system make available particular content to each profile, e.g., a Calculus student from Computer Science can see, in all the three sections, only the content available for Computer Science students', while an Electrical Engineering student will see only the content available to Electrical Engineering students'.

The discipline's content can be seen in the Class Environment. From this section, the student can obtain access to the concepts, examples, exercises and complementary materials. The student can access this section by two different ways: Tutorial mode or Free mode.

In the Tutorial mode (guided tour), prerequisites criteria among concepts determine the student's navigation, and navigation adaptation is based on the register of concepts studied: every time the same student accesses the web course, the colors of the menu links are restored. In the Free mode, the student can study any concept available in the navigation menu. The colors used follow the usability rules by Nielsen [9], that is, we adopted blue for links that were not accessed (but they are accessible to click), and purple for the accessed ones. For concepts that the student is not allowed to see (tutorial mode) the color of the link is black, and the mouse cannot click on them [2].

In the Messages Board section, the student can see the messages of his course/discipline from the last 30 days. These messages can be from another student or from the teacher, and they can be public or private. Beyond sending a message to the class (students and teacher), the student can also send a message only to the teacher.

Finally, the Discussion Forum section allows students to interact with each other. They can create discussions (topics) about some subject of their interest and get aid to accomplish the exercises or to understand a concept.

4 Collecting and Analyzing Data

Among several existing web analytics tools, the vast majority use the same mechanism for collection and storage of data, usually done through web logs [7]. The data capture is one of the most important web analytics processes. There are two main ways to collect data of user interactions on the sites. The methods are classified in web logs and page tagging [8].

The web logs files are the source of the data access of the sites. They were designed with the goal of recording errors generated by web servers and later evolved to capture data access enabling an analysis not only of technical failures but also of marketing data and user behavior [8].

The second method that collects data from users is the page tagging. The first techniques were to carry an invisible image along with the requested page of a site to detect whether the page had loaded successfully and thereafter send a message to the server with the user information. This technique is also described as page tagging web beacon [7]. The page tagging technique involves adding scripts on each page of the website to capture and log usage statistic of each page in question. These data are collected during page load and usage of the site by the browser. The page tagging approach provides more accurate and complete data for web analytics applications than web logs [5].

The web analytics tool proposed in this paper intends to complement the analysis of student behavior when using the environment. This proposal was implemented through the combination of different metrics, providing more specific analyzes for the teacher. The development of the tool in AdaptWeb was divided into three modules: i) data collection module, ii) data analysis module, and iii) data visualization tool.

The structure of the data collection module consists in the capture and storage of navigation data, clicks, downloaded files, and other forms of students' interactions with the environment. The method of data collection for this study is based on a hybrid model (as shown in Figure 1) using existing data on the environment in the log table in the database and the page tagging technique to capture the data from the student's browser. We had used Piwik page tagging tool for this technique.

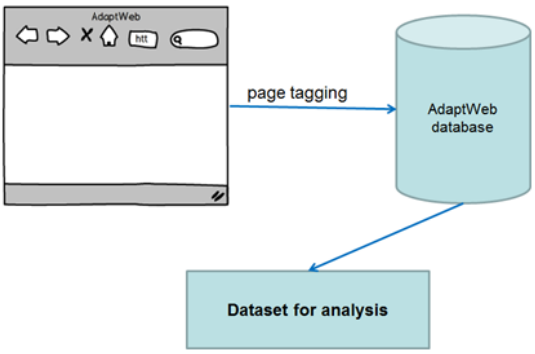

Fig. 1. Hybrid model of data collection

The module collects data from different sources. First it collects data from the AdaptWeb environment, where it reuses the existing logs of the database and then it includes an embodiment of Piwik, an open-source web analytics tool. Piwik helps the capture and storage through page tagging with a JavaScript code embedded in the source code of AdaptWeb' pages. Based on the data collected, the data analysis module was developed considering the metrics defined in Table 1:

Table 1. Metrics of the web analytics tool

Group	*Metric*	*Derived metrics*
General use	Total number of students visiting the e-learning system	Total Visits per student
	Individual access time in each discipline	Students' Average Access time
	Frequency of student access	Frequency of access
	Technological user data	Browsers
		Operation Systems
		Screen Resolutions
	Access time in each section	Total Hits to each section
Class environment	Access time in each section	Navigation Mode
	Concepts	Total Hits to Concepts
	Examples	Total Hits to Examples
	Complementary Materials	Total Hits to Complementary Materials
	Performance of tasks	Total Hits to Exercises
	Internal search	Total Use of The Search Engine
		Keywords searched
Discussion Forum	Participation in the discussion forums	Total Hits to Topics
		Total Topics Created
		Total Answers
Messages Board	Interaction via messages	Total Sent Messages
		Types of Sent Messages
		Total Views

The goal of the data analysis module was to provide metrics for the visualization module and also being flexible and extensible to future work. Its structure can easily be incorporated into other modules of the environment, delivering information in a standardized manner. The metrics are related to data from the environment where they were captured, which can be parameterized by student, teacher, discipline and course.

A total of 20 metrics, according to Table 1, were mapped into the system. These metrics can be queried via a generic function in the web analytics tool library. As a result of this data analysis module, we obtained a set of functions grouped by General Use, Class Environment, Discussion Forum and Messages Board.

5 Visualization Tool

Visualization is the process of transforming data, information and knowledge into visual form making use of humans' natural visual capabilities. In this way, with effective visual interfaces we can interact with large volumes of data rapidly and effectively to discover hidden characteristics, patterns and trends [3].

To provide a way for teachers to find the information they need in a quick and easy way, visualization techniques were also studied and used in the tool to represent the students' data collected according with the metrics established. By using these techniques not only teachers can see the data represented visually, but they can also interact with the tool, and change some aspects of the representation in order to find useful information.

The tool's interface is composed of four areas: (a) filters, (b) metrics selection, (c) chart and (d) detailed information table. In Figure 2 the filter and the metrics selection are shown.

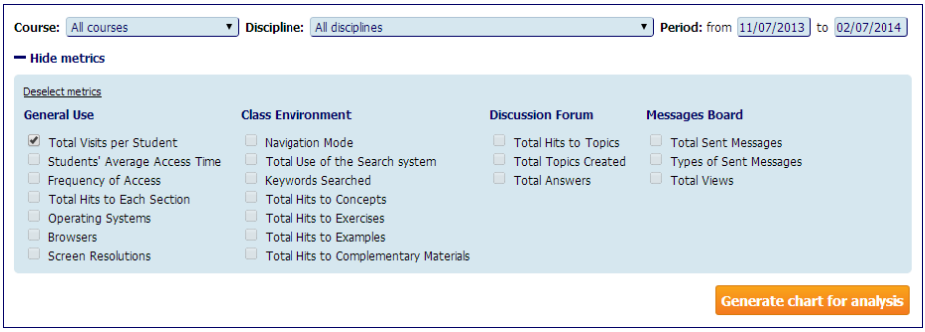

Fig. 2. Filter and metric selection areas

As can be seen in Figure 2, the teacher can filter the data by choosing a course, a discipline and defining a time period (initial and final date - which, if not defined, will be automatically set as the last three months). Another action the teacher can do is choosing the metric(s) he wants to analyze. The metrics available for analysis are listed in the blue area of Figure 2, organized in four categories: General Use, Class Environment, Discussion Forum and Messages Board.

The majority of these metrics can only be analyzed individually, but some metrics can be jointly analyzed: Total Hits to Concepts, Total Hits to Exercises, Total Hits to Examples and Total Hits to Complementary Materials from the Classroom Environment category can be analyzed together; and Total Hits to Topics and Total Answers from the Discussion Forum.

After selecting the metrics, the teacher must press the Generate chart for analysis button and the results will be shown according to Figure 3, below the filter and the metric selection areas.

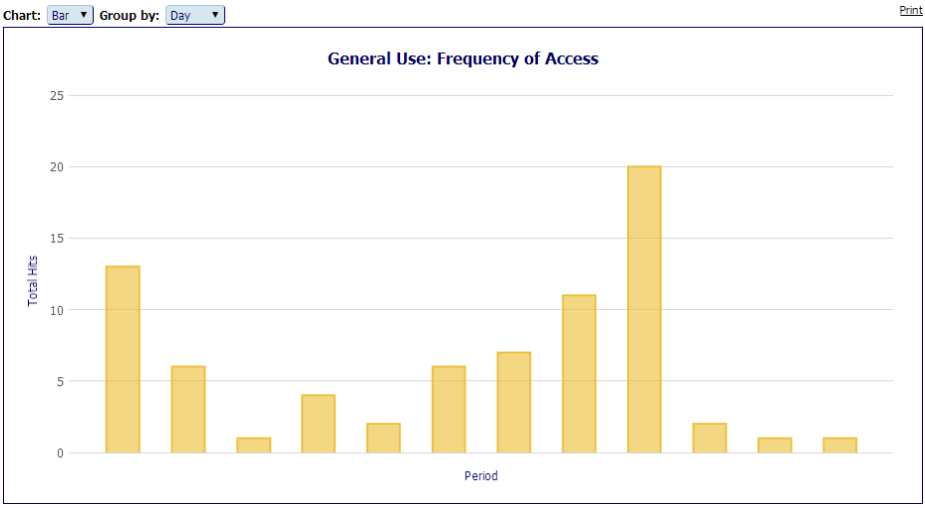

Fig. 3. Graphical results of metric *Frequency of Access*

In the left superior corner of Figure 3 it is possible to see two combo boxes where the teacher can select different graphical patterns for the chart — bar, line or pie — and groups it by period— day, week, month — he wants to analyze the data (available only for the metric Frequency of Access). Other features of the tool are an option to choose between labels and caption to the pie chart (Figure 4) and the tooltip (highlighted in Figure 5) when hovering on a chart element (dot, bar and section).

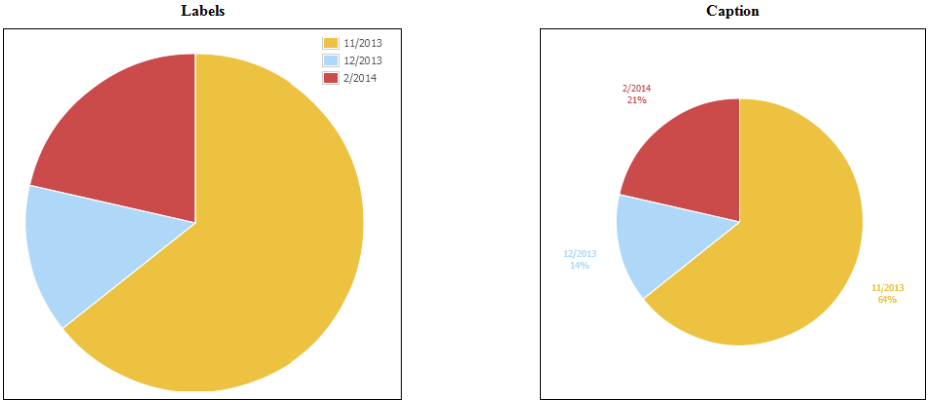

Fig. 4. Pie chart with labels and caption

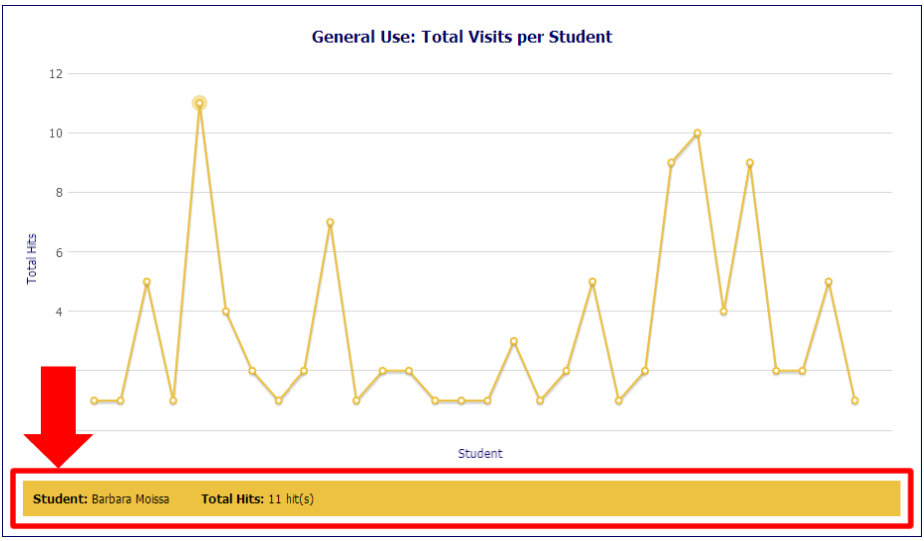

Fig. 5. Tooltip

Finally, the last area of the tool is the detailed information table. This table has a few useful features to help teachers to better understand the values represented such as being manageable and allowing the teacher to sort (ascending and descending) the lines according to the values of a column of his choice and having a filter by keyword to display only the lines that he is really interest. One interesting thing about this table, when analyzing the metric Access Frequency is that the data presented in it changes according to the grouping chosen (i.e. by month, week, or day), this way if it is organized by day, then the table will show the records grouped by day.

This visualization tool has a Help Section and three hyperlinks — located in three different positions of the tool (above the chart, between the chart and the table and below the table) — to redirect the teacher to a print page where he can choose which section he wants to print (i.e. chart, detailed information table or both).

The web analytic tool, composed by the data collection module, the data analysis module, and the data visualization module, was designed and implemented in a real and operational adaptive e-learning environment, namely AdaptWeb.

6 Case Study

We have conducted a case study with the web analytics tool based on a real dataset of undergraduate students of Computer Science and Information Technology. This study involved, at first, the inclusion of our tool in the AdaptWeb's environment placed in the server of University of Santa Catarina State.

After we installed and tested our tool, we invited two teachers to participate in a study case based on their large experience of using e-learning systems. Thus, the data collection tool began monitoring the interaction and behavior of their students.

After a month, we had sufficient data to support different analysis. We invited the teachers to participate in our first usability test. An experiment of observation and analysis with one of the teachers was conducted. In this experiment the developers presented the tool to the teacher and asked him to use it. The teacher could express himself anytime while he was interacting to the tool (beyond the observation, we had used the thinking aloud protocol). After the observation section, the developers discussed the tool with the teacher and understood the challenges and details of the visualization tool.

As the first results, the graphical representations of the metrics were easily understood by the teacher; however he had some difficulties while selecting the metrics because he could not visually distinguish which metrics could be analyzed together and which could not. The teacher needed to be guided to try features such as changing the type of frequency of the *Frequency of Access* (e.g. grouped by month, by week, by day) and the features of the detailed information table, however, after he used it for the first time, he considered it very useful.

Concerning to the mapped metrics, the teacher missed the option to analyze the metric *Frequency of Access* with others such as *Total Hits to Concepts*, *Total Hits to Examples*, *Total Hits to Exercises* and *Total Hits to Complementary Materials*. He also mentioned that a new metric, *Student's Path*, could also be interesting, to visually identify students trails and, perhaps, to change some of his discipline structure.

7 Conclusion and Future Works

In this paper we presented a web analytics and visualization tool to help teachers to understand student´s behavior in an adaptive e-learning environment. This tool consists of three modules: i) data collection module, ii) data analysis module, and iii) data visualization tool, which enables the teacher to visualize and interact to the generated analyzes. Our tool was integrated in the AdaptWeb environment, an adaptive e-learning system used in our university.

We have conducted a case study, tested with data collected from two courses — Computer Science and Information Technology, aiming to identify the prior usability problems related to the visualization tool, and to discover whether the metrics satisfy teacher's needs. The preliminary results demonstrated the importance of well described metrics and filters. Currently we are conducting other experiments, in order to improve our tool.

Future work will include the integration of our web analytic tool with more disciplines from different courses and areas of knowledge. We also plan to enhance our tool with intelligent recommendation support and evaluate its usefulness and usability.

References

1. Dyckhoff, A.L., Zielke, D., Bültmann, M., Chatti, M.A., Schroeder, U.: Design and Implementation of a Learning Analytics Toolkit for Teachers. Educational Technology & Society 15(3), 58–76 (2012)

2. Gasparini, I., Eyharabide, V., Schiaffino, S., Pimenta, M.S., Amandi, A., Palazzo, M., de Oliveira, J.: Improving user profiling for a richer personalization: modeling context in e-learning. In: Graf, S. (ed.), ch. 12, pp. 182–197. IGI Global (2012)
3. Gershon, N., Eick, S.G., Card, S.: Information visualization. Interactions 5(2), 9–15 (1998), http://doi.acm.org/10.1145/274430.274432
4. Govaerts, S., Verbert, K., Duval, E.: Evaluating the student activity meter: Two case studies. In: Leung, H., Popescu, E., Cao, Y., Lau, R.W.H., Nejdl, W. (eds.) ICWL 2011. LNCS, vol. 7048, pp. 188–197. Springer, Heidelberg (2011)
5. Hasan, L., Morris, A., Probets, S.: Using Google Analytics to evaluate the usability of e-commerce sites. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 697–706. Springer, Heidelberg (2009)
6. InfoVisWiki - Information Visualization community platform (2014), http://www.infovis-wiki.net/index.php/Information_Visualization
7. Jansen, B.J.: Understanding User–Web Interactions via Web Analytics. Morgan & Claypool (2009)
8. Kaushik, A.: Web Analytics: An Hour a Day. Wiley Publishing Inc. (2007)
9. Nielsen, J.: Designing Web Usability: The Practice of Simplicity. New Riders Publishing, Indianapolis (2000)
10. Romero, C., Santos, S.G., Freire, M., Ventura, S.: Mining and Visualizing Visited Trails in Web-Based Educational Systems. In: EDM, pp. 182–196 (2008)
11. Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S., Shum, S.B., Ferguson, R., Duval, E., Verbert, K., Baker, R.S.: Open Learning Analytics: an integrated & modularized platform. Proposal to design, implement and evaluate an open platform to integrate heterogeneous learning analytics techniques (2011)