Document Version
This is the Accepted Manuscript version.
The version published on the UOC's O2 Repository may differ from the final published version.
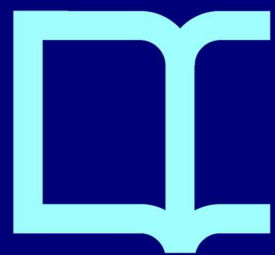
Educational Monitoring Tool Based on Faceted Browsing and Data Portraits

David García-Solórzano
Universitat Oberta de
Catalunya
Rambla del Poblenou, 156
Barcelona, Spain
dgarciaso@uoc.edu

Germán Cobo
Universitat Oberta de
Catalunya
Rambla del Poblenou, 156
Barcelona, Spain
gcobo@uoc.edu

Eugènia Santamaría
Universitat Oberta de
Catalunya
Rambla del Poblenou, 156
Barcelona, Spain
esantamaria@uoc.edu

Jose Antonio Morán
Universitat Oberta de
Catalunya
Rambla del Poblenou, 156
Barcelona, Spain
jmoranmm@uoc.edu

Carlos Monzo
Universitat Oberta de
Catalunya
Rambla del Poblenou, 156
Barcelona, Spain
cmonzo@uoc.edu

Javier Melenchón
Universitat Oberta de
Catalunya
Rambla del Poblenou, 156
Barcelona, Spain
jmelenchonm@uoc.edu

ABSTRACT

Due to the idiosyncrasy of online education, students may become disoriented, frustrated or confused if they do not receive the support, feedback or guidance needed to be successful. To avoid this, the role of teachers is essential. In this regard, instructors should be facilitators who guide students throughout the teaching-learning process and arrange meaningful learner-centered experiences. However, unlike face-to-face teaching-learning, teachers have difficulty in monitoring their learners in an online environment, since a lot of learning management systems provide faculty with student tracking data in a poor tabular format that is difficult to understand. In order to overcome this drawback, this paper presents a novel graphical educational monitoring tool based on faceted browsing that helps instructors to gain an insight into their classrooms' performance. Moreover, this tool depicts data portraits for each learner to show student's individual features. As a result, teachers can use this tool to detect potential problems and take decisions to avoid them.

Categories and Subject Descriptors

J.1 [**Administrative Data Processing**]: Education; K.3 [**Computer Uses in Education**]: Distance Learning; H.5 [**User Interfaces**]: Graphical user interfaces

General Terms

Management, Design, Human Factors

Keywords

Learning analytics, information visualization, student monitoring, instructor support, faceted browsing, data portraits

1. INTRODUCTION

In the last years, distance learning has become very popular, especially, its modality of online education. As far as the online education is concerned, online students have different characteristics, needs and preferences to regular ones who attend face-to-face classes. For instance, an online course often has learners with a wide range of age, most of them are over 30 years old, which means having learners with different professional and educational backgrounds in the same classroom. This heterogeneity also entails helping students with different needs at the same time. Likewise, because most learners are adult, they have a lot of responsibilities, such as: working, taking care of children, etc. Consequently, it is hardly surprising that they prefer studying at their convenience, i.e. anytime, anywhere and at their own pace.

As a result of the aforementioned particularities, online education should avoid using a teacher-centered paradigm, since it treats all learners as if they were the same and forces them to do the same things in the same amount of time [29]. Instead, online education, as any kind of distance learning, involves a student-centered approach in which the instructor is a facilitator and students engage in peer learning [17, 23]. Thus, students are required to take primary responsibility for their learning process [4]. Hence, learners need to have some specific skills, e.g. self-regulation. However, some studies [21, 24] provide evidence that a lot of students need some help to learn these skills, since most of them are not able to achieve these abilities on their own. Consequently, the role of teachers shifts from a masterful figure to a facilitator who guides learners throughout the course and arranges meaningful learner-centered experiences. This role is essential in online education, since learners may become disoriented, frustrated or confused if they do not receive the support, feedback or guidance needed to be successful [7].

In order to carry out a good instruction in an online environment, instructors need appropriate means to set up a

monitoring process so that they can be aware of the students' learning process and provide learners with just-in-time assistance. In addition, monitoring allows teachers to forecast potential problems (e.g. dropouts) and avoid them in time. In this regard, using logs which are stored by Learning Management Systems (LMSs) is a possible way to support instructors in online education. However, a comparative study on tracking functionalities of various LMS [9] concluded that none of them offer much tracking ability. One of the reasons of this devastating conclusion may be the fact that these platforms often provide tracking data in a tabular format which is commonly poorly structured, incomprehensible and difficult to understand [19].

This paper presents a novel graphical educational monitoring tool based on faceted browsing and data portraits. On the one hand, faceted browsing, as an exploratory search technique, helps instructors to narrow the class roster down until finding those learners who meet teachers' requirements. On the other hand, data portraits depict summarized information about an individual student in a single image.

The rest of the paper is structured as follows: a categorization of current educational monitoring tools can be found in the next section. Next, an educational monitoring tool based on faceted browsing and data portraits is proposed in section 3. Finally, conclusions and future work are placed in section 4.

2. CLASSIFICATION OF EDUCATIONAL MONITORING TOOLS

This paper classifies educational monitoring tools into two categories: (1) according to the data processing techniques that they use, and (2) the element which they monitor. They both are described below.

2.1 Data Processing Techniques

The main goal of any monitoring tool, whether this has an educational purpose or not, is to give users insights on the data at which they are looking. To this end, two different data processing techniques can mainly be used. On the one hand, information visualization (IV) techniques and, on the other, data mining algorithms. Next, they both are explained. At the same time, some relevant educational examples related to each method are described briefly.

2.1.1 Information Visualization

One of the most well-known definitions of information visualization was proposed by Card et al [3]. They defined IV as *"the use of computer-supported, interactive, visual representations of abstract data to amplify cognition"*. From this definition, IV can be considered as a set of techniques that use different graphical properties to convey information in an effective way. Thereby, these visual representations generated from LMS data reveal facts and trends which allow users to infer some unknown information by combining the visual inputs with their knowledge of data. In conclusion, IV converts, with minimum processing (basically statistics), tracking data collected by an LMS into visual elements that create an effective graphical representation of the data.

In the educational context, some proposals based on IV techniques have been suggested. Two of the most relevant tools are CourseVis [19] and, its successor, GISMO [20]. The former is a stand-alone visualization tool that obtains track-

ing data from WebCT, transforms them into a form convenient for processing and generates graphical representations that can be explored and manipulated (e.g. zooming) by course instructors. Thereby, teachers can examine social (it uses a 3D visualization of discussion boards), cognitive (it uses a matrix in which each cell is the grade attained by a specific student in a specific quiz), and behavioral (it shows information about access and participation in a timeline) aspects of distance students. CourseVis mainly uses 2D visualization techniques, but it also uses color and shape as a third dimension.

As far as GISMO is concerned, this also uses the LMS tracking data, but in this case from Moodle, to display graphical representations (e.g. bar charts, matrix visualizations, etc.) about overall classroom accesses and detailed information from a specific student.

Similarly, Zhang et al. [32] designed Moodog (Moodle watchdog). This is a visual student tracking data plugin for Moodle which also sends automatic reminder mails to students. This displays information about the course, students, resources and access times. The most relevant difference with GISMO is that Moodog's information is inserted in the main Moodle page, keeping the original Moodle interface as much as possible. On the contrary, both CourseVis and GISMO are separated modules which are not integrated into WebCT and Moodle's webpages, respectively.

Finally, many other studies have also proposed different representations based on IV techniques. For instance, Hardy et al. [8] constructed the route (as a graph) taken through the online course material by one student during a single work session; Hijón-Neira and Velázquez-Iturbide [10] used interactive graph (for students' grades) and data mountain (for access) by using Prefuse API; and Juan et al. [13], in turn, proposed scatter plot and quadrants as well as an evolution graph in order to represent students' performance.

2.1.2 Data Mining

Unlike the previous technique, tools based on data mining process student tracking data by using algorithms from the field of artificial intelligence. According to Romero and Ventura [26], the main data mining categories are: (1) clustering, classification and outlier detection; (2) association rule mining and sequential pattern mining; (3) text mining; and (4) statistics and visualization. The latter is not universally seen as data mining [2]. In fact, this accords with the previous category proposed in this paper, information visualization (IV), due to data processing in IV is minimum compared to the rest of categories. In this regard, a brief discussion about the differences between IV and the rest of categories is presented in the subsection 2.1.3.

The techniques that belong to the first three categories allow to infer underlying patterns, generating some new type of information such as student models or predictions. These new data can be delivered in writing or visually. Next, some educational proposals based on data mining are described.

Kosba et al. [15] developed TADV (Teacher ADVisor). This proposal builds student, group and classroom models by using fuzzy logic. From these models, TADV gives teachers advice on students, group and classroom' performance, e.g. to advise a student to work more with a specific concept. Despite the advice recommended by TADV, instructors are who make the final decision to send an message based on TADV's advice or not.

In 2008, Hung and Zhang [11] used statistical models and machine learning algorithms to analyze patterns of online learning behaviors and, at the same time, to make predictions on learning outcomes.

Zorrilla et al. [34] proposed a decision support system that utilized different techniques. First, two clustering algorithms, Expectation-Maximization (EM) and K-Means, were used to characterize, on the one hand, students and, on the other, sessions. The output of EM, a probability distribution, allowed to determine the number of clusters with which K-Means would be executed. The clusters obtained from K-Means, represented by their centroids, characterized either different student behavior models or sessions patterns, depending on the input data. Secondly, Apriori, an algorithm for finding association rules, was employed to discover the resources that were commonly used together.

There are a lot of works that use data mining for teaching-learning purposes. In this regard, more applications of educational data mining can be found in [2, 26].

2.1.3   Information Visualization vs. Data Mining

One of the main differences between IV and data mining is how they process data. The former calculates statistics with minimum computational cost whereas the latter executes complex artificial intelligence algorithms which are usually time-consuming and often need experts who define rules, parameters, models, etc., to adapt them to each domain.

Due to the complexity of data mining techniques, these are able to suggest some latent information with an explanation based on text, rules, models, clusters, etc., e.g. *"student S is about to drop out because she has not acceded to the classroom for two weeks"*. Nevertheless, the explanation provided often requires users to have some expertise to understand it. Unfortunately, most users neither have this knowledge nor can make time for acquiring it. Likewise, the fact of generating underlying information has the drawback that users must rely on the accuracy of the information suggested by the algorithm. On the contrary, IV techniques show original data and statistics in an effective visual way. Thereby, users infer reliable information by observing graphics and taking advantage of their knowledge of the domain.

As far as the user-friendliness is concerned, Merceron and Yacef [22] claim it is essential to use techniques and measurements that are fairly intuitive and easy to interpret. As seen, tools based on IV seem to meet these requirements better, since they use simple data and emphasize visual elements.

Finally, unlike other domains, educational data are very dynamic and vary a lot between samples (i.e. students, courses, activities, etc.). For instance, two courses in Mathematics for 10-year-old children can be taught in different way and, hence, they will probably generate very different values for the same item (or event different type of data). As said, a lot of data mining techniques need be customized, by defining specific rules and models, for each domain in which they are used. Consequently, the fact that a data mining technique is good for one setting, may not necessarily mean being good for others, and vice versa. As a result, it is difficult, or even impossible, to compare different data mining methods and measures, and decide which is the best [22]. However, IV techniques are more independent of the domain, as they do not need rules, models and so on. Therefore, tools based on IV can be generalized to different domains more easily than those based on data mining.

2.2   Monitored element

Although the most popular name is student monitoring tools, what would be more correct is to call them educational monitoring tools. The reason is that these tools can monitor other elements in addition to students. Next, the most common ones are described. Finally, to remark that any of the techniques described in section 2.1 can be used to track any of the following elements.

2.2.1   Classroom

The most common one is the whole classroom, i.e. all students are considered a unique entity. This entity is characterized by the overall information that comes from combining the all students' tracking data (e.g. class's average grade in each assignment). Thereby, instructors are provided with an overview of the processes that are carried out by their students. Thanks to this, teachers can take general decisions to improve the class's learning experience. Most educational monitoring tools show data of the classroom.

2.2.2   Student Group

A particular case of the previous monitoring is that which focuses on groups. This is useful when group activities are proposed, e.g. tasks that belong to a project-based learning. This supervision gives information about how students interact each other, who are the most and least participative members, which group is the best and the worst one, etc. [13, 25] are examples of group monitoring tools.

2.2.3   Individual Student

Teachers often need to take a closer look at a particular student or make comparisons between students. Consequently, they need tools that provide detailed information about an individual learner (e.g. how many times a student has accessed the LMS). These data may help instructors to gain understanding of the reasons why a specific learner has a good/bad behavior. Thereby, teachers can offer each student a better support and a tailored learning experience. Regarding this, [19] and [32] are two tools that display information of a particular student. The first one shows a summary of student's behavior based on a timeline (X-axis) which displays information (Y-axis) about access, participation in forums, studying pace and submissions. As far as [32] is concerned, it represents for each student, by using simple bar charts, a set of information that includes the total number of times that course material has been viewed, total online time, total number of completed sessions, number of viewed resources, number of threads initialized and follow-up messages.

2.2.4   Resources

Information of the resources related to a course (e.g. documents, activities, forums, LMS, etc.) can be analyzed too (e.g. how many times a document has been read). Thereby, it is possible to validate or refute the hypothesis used in the design of the learning process (e.g. to detect a bad design of content pages) and detect potential problems such as misunderstood concepts or dropouts (e.g. by detecting that content pages are not visited quite often). In this regard, [34] is an example of this type of monitoring.

A graphic can sometimes focus on a particular resource or student depending on how this is read. For instance, Mazza and Dimitrova [19] display the student's performance on

quizzes by using a matrix representation. Thereby, if teachers paid attention to a specific column, they would be observing the performance of a particular student in all quizzes of the course. By contrast, if a row were analyzed, then they would focusing on the performance that the classroom has had in a specific quiz.

3. PROPOSAL

From the observations done in the previous section, this paper presents a graphical interactive educational monitoring tool which uses information visualization techniques. This allows instructors to monitor the class and, at the same time, look details of a particular student. The first part, monitoring of the classroom, is based on faceted browsing, a type of exploratory search. As far as the second part is concerned, a novel technique called data portrait is used to depict the information of a specific learner. A detailed explanation of both parts of the tool is exposed below. Nevertheless, before explaining the proposal of this paper, a brief list of the most common graphical characteristics of educational monitoring tools based on IV is given. This will help to better understand the contributions of the present work.

3.1 Graphical features of educational monitoring tools based on Information Visualization

3.1.1 Stand-alone vs. Built-in

There are tools which have their own interface and collect data from an LMS [19, 20], whereas there are others that, due to the success of LMSs (e.g. Moodle), are integrated into their framework as a plug-in [32].

3.1.2 2D vs. 3D

2D graphics are firmly established in monitoring tools. They consist of two dimensions (or axes) in which each of them represents a variable (or attribute) of information (e.g. students, grades, etc.). This representation are generally very intuitive. On the other hand, 3D graphics has also been proposed. An example is the scatter plot for discussion boards proposed in [19]. However, teachers stated that this 3D scatter plot created some confusion [18]. Thus, in general, 2D graphics are easier for instructors.

3.1.3 Multivariate data

Much more variables than the number of dimensions of the graphic (i.e. 2D or 3D) can be displayed thanks to visual elements. However, it is essential to find a balance between showing much information and avoiding overwhelming. To this purpose, different techniques are used. For example, authors in [19] take advantage of single-axis composition method [16] whereby an axis represents a large set of variables (e.g. content topics, number of accesses, etc.). Likewise, visual components, such as color, shape and size, are used as a data dimension. Another technique are the quadrants, which allow to organize data into four groups [13]. Mouse events, such as click and rollover, are also used to provide information (e.g. relationships between elements [10]). Finally, unlike most of the proposals that depict data as points, bars, lines and so on, there are tools that use an only representation to display multivariate observations with an arbitrary number of variables. This is the case of star plots [12], which show each observation (e.g. a learner)

as a star-shaped figure with one ray for each variable (e.g. grades, participation, etc.).

3.1.4 Manipulation

Zoom, rotation (in 3D) and filtering are typical actions that tools allow instructors to do. With regard to filtering, this refers to hide some visual elements so as to emphasize others. Thereby, each value in graphics acts as a layer that can be shown and hidden. Take the example of imagining a bar chart that shows the variables "student" (X-axis), whose values are $\{S_1,S_2,S_3,S_4,S_5\}$, and "number of posts" (Y-axis) with values $\{4,6,7,5,8\}$. In this case, the graphic has 10 layers that can be shown and hidden. Thus, a teacher may indicate that the graphic only shows the bars with length 4 and 5 (students S_1 and S_4) or those that belong to the students S_2 and S_5. However, she cannot indicate a general constraint that asks to show those learners who have written between 3 and 9 messages, since the layers 3 and 9 do not exist, among other things. Similarly, she cannot refine the information space according to other variables that are not in the graphic (e.g. number of accesses). Hence, most tools with filtering options do not allow to freely do an exploratory search to discover new knowledge about the information space that is shown.

3.2 Monitoring of the Classroom

3.2.1 What is Faceted Browsing?

Faceted browsing is becoming a popular method that allows users to interactively search and navigate through complex information collections. This is widely used on a lot of e-commerce websites such as eBay or Wal-Mart.

A faceted browser provides users with key-value metadata that is used for query refinement [14]. Faceted browsing is made up of three stages [31]: opening, middle game and end game. In the opening, the interface shows the whole collection and all facets that can be used. The middle game, in turn, allows users to iteratively narrow down the result set by defining constraints on the values of one or several facets, which refines the search query. Finally, the end game occurs when the user finishes the search by selecting an individual item from the result set and its information is detailed.

3.2.2 Why Is Faceted Browsing Suitable for Monitoring Online Education?

As said in section 1, teachers in online education should be guides. To carry out this role, they require to observe students' behavior by exploring and analyzing any feature associated with learners in order to detect any kind of problem. Unlike the other proposals, which show graphics with predefined attributes or queries, faceted browsing allows teachers to perform an exploratory search strategy by using a wide range of orthogonal variables (i.e. facets). Instructors gain an insight into the behavior of their classrooms by iteratively submitting tentative queries based on more than one facet at the same time. This iterative process is the middle game. This finishes when instructors find relevant information that they did not know or until the result set meets a specific set of requirements that they wanted. Next, teachers can use this information to take decisions (e.g. send an e-mail) or, by contrast, they can select one student in order to obtain detailed information. When instructors click on a specific learner, the faceted browsing is in the end game.
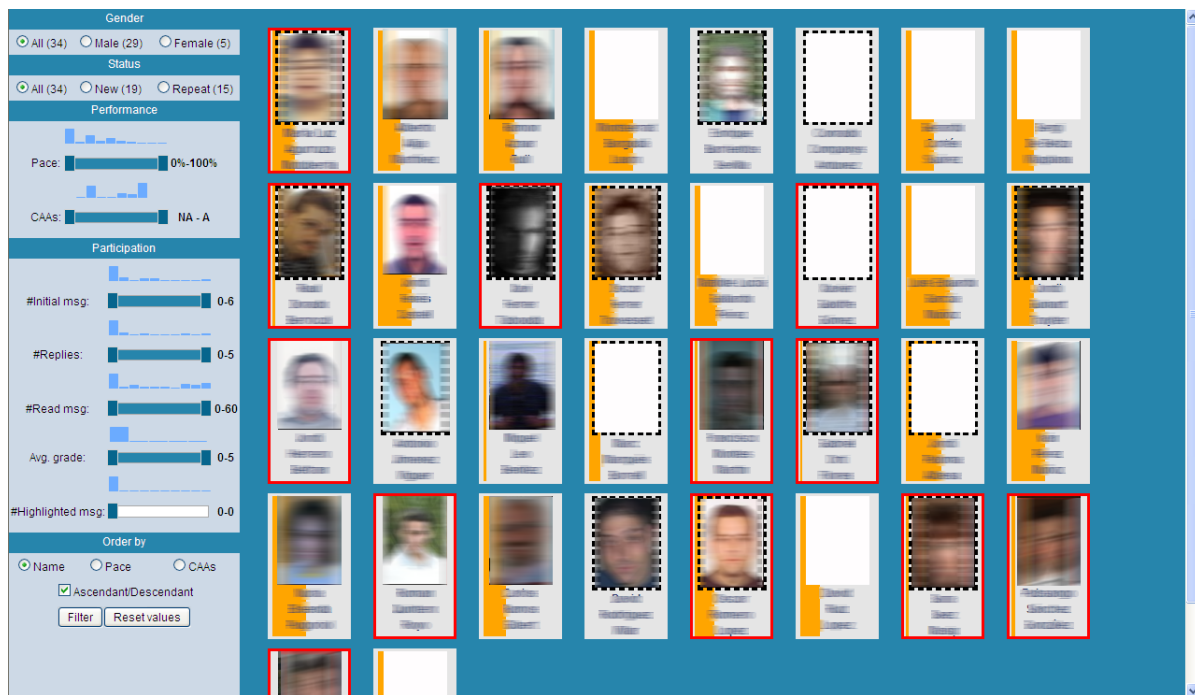
Figure 1: The opening stage of the faceted browser proposal. Students remain anonymous.

As seen, faceted browsing seems to be good at monitoring a classroom, since it allows instructors to define any tailored query and find relevant information that they did not know while they explore/refine the result set.

3.2.3 Proposal of Faceted Browser

This paper presents a built-in faceted browser (see Figure 1) that collects data from an ad hoc learning management system [1]. This uses information visualization techniques in order to effectively display data, leaving knowledge inference in teachers' hands. Therefore, data mining techniques are not used to obtain underlying knowledge about students.

With regard to its interface, 2D graphics are used. Its design is divided into two areas. The main one shows the class roster as a set of cards in which each card includes student's name and photo. The use of a photo is better than identifying students by looking at points, squares, bars, etc. Actually, Zhao et al. [33] point out that there is evidence for the existence of a dedicated face processing system in human's brain. On the other hand, on the left, there is a menu that has different facets whereby teachers can narrow down or sort out the class. Table 1 gathers the set of facets that have been included in this tool along with their possible values. Facets have also been grouped into three categories (or metafacets) to make the exploratory search easier: basic information, performance and participation in forums.

As seen in Table 1, facets can have nominal or numeric values. Those facets that allow gradation are represented with sliders. To indicate the number of students in each facet value, numbers in parentheses and histograms are used. Their values are updated depending on the data set that is shown in the main area every time. This helps teachers to set more meaningful queries during the iterative process (i.e. the middle game) that they carry out through the facet browser. In addition to filtering the data set, this can be sorted by name, studying pace or the average grade of assignments in ascendant or descendant order.

Finally, different visual elements are used to represent more variables in the main area. The background color of each card indicates the student's studying pace. Thereby, the more orange the background is, the more advanced the student's studying pace is. Besides the background color, the card can have a red border that means that the average grade of assignments is C- or less and, hence, this student would not pass the course if the term finished at that moment. Moreover, the border of the photo can be drawn with black dashes. This means that the student is repeating the subject. Likewise and, because results can be ordered, the position of cards is another visual element that gives extra information. Thereby, if a teacher sorted learners by the average grade of assignments in ascendant order, then students with lower grades would be on the top positions.

3.2.4 An Example of Using the Faceted Browser

From the opening stage shown in Figure 1, an instructor may introduce any query. For instance, she could indicate the following query: *"To retrieve, ordered by studying pace in a descendant way, those students who, regardless of their participation in forums, are not repeating the course and, at the same time, have achieved 24% of activities as well as they have an average grade of assignments equal or greater than C+"*. If the previous query was executed, then the result would be that shown in Figure 2b. In this regard, Figure 2a shows the transition between the opening (see Figure 1) and the middle game (see Figure 2b). As seen, instructors can see how students change their positions or even they dissapear. Thanks to the possibility of seeing the transition, teachers can gain extra understanding of their learners.

Finally, it is worth emphasizing that the result of the Figure 2b may be either an intermediate step of the middle

Table 1: Set of Metafacets and Facets with their Values

| Category / metafacet	Facet	Values	Type
Basic information	Gender	male, female or all	Nominal
	Status	new (i.e. first time in the course), repeat or all	
Performance	Studying pace	[0%, 100%]	Numerical
	Assignments grade	[NA (Not Assessed),NP (Not Presented),D,C-,C+,B,A]	Nominal
Forums participation	Num. initial posts	[0, max. num. messages posted by a student]	Numerical
	Num. replies	[0, max. num. replies written by a student]	
	Num. messages read	[0, num. messages in the forum]	
	Average grade of messages	[0, 5]	
	Num. highlighted messages	[0, num. messages highlighted by the teacher]	

(a) Transition

(b) Middle game
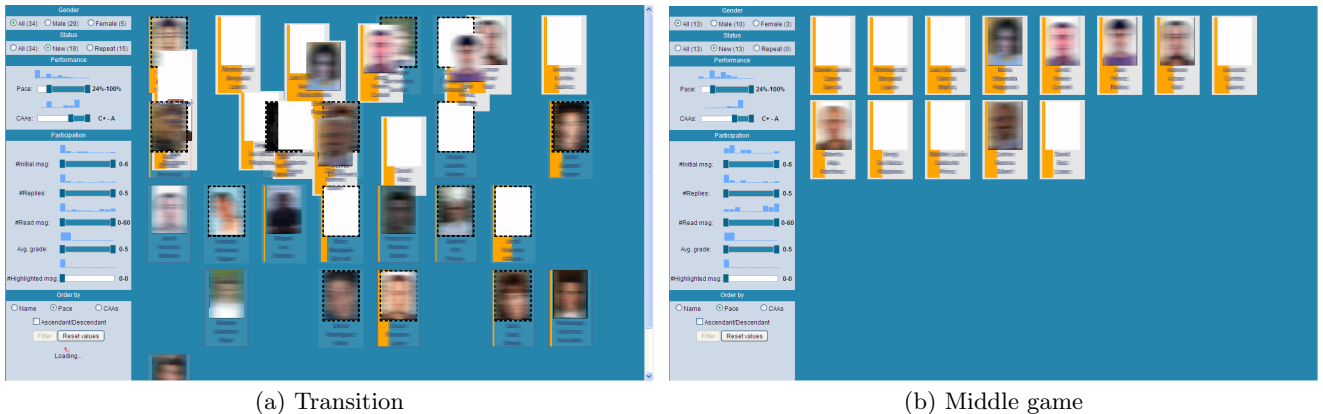
Figure 2: Execution of a query in the proposed faceted browser.

game or the last one. This depends on whether the teacher is satisfied with the result and wants to click on some student or, by contrast, prefers to narrow down the result by changing some facets.

3.3 Monitoring of an Individual Student

So far, only information about the overall classroom is represented. However, instructors need to know details of a specific learner quite often. In this regard, the great majority of monitoring tools usually show the same kind of graphic for both the classroom and the learner. Instead of this, the present paper suggests using a novel technique, called data portraits, to display the individual student's information.

3.3.1 What Is a Data Portrait?

According to Donath [5], *"data portraits depict their subjects' accumulated data rather than their faces. They can be visualizations of discussion contributions, browsing histories, social networks, travel patterns, etc. (...). Data portraits depict a person through their digital archive"*. In short, the idea behind data portraits is to convey a large amount of information from an individual compactly.

An example of data portrait used for discussion forums is PeopleGarden [30]. This is a metaphor in which each user is a flower and, hence, the forum is a garden. Each petal symbolizes a message written by the user. Thereby, the number of petals indicates the user's posting frequency. Thus, the more petals a flower has, the more active the user is. Likewise, the petal's color represents if the message is an initial post (in magenta) or a reply (in blue). Moreover, pistil-like circles are used on top of the petals to show how many responses each message has received. To display how

old a message is, petals, like in the real life, fade over time.

Another similar proposal is called daisy maps [12]. A daisy map is a star-shaped glyph that displays the scores received on different parts of an assignment (e.g. reading, writing, etc.). Thereby, each student is a daisy map and each petal has a color depending on score attained.

Lexigraphs [6] is, in turn, a group of data portraits in which each user is represented as a face-like outline. Each one is drawn by the words written in the user's Twitter account. Thus, silhouettes are updated with each new tweet.

Finally, Authorlines [28] is an horizontal timeline with vertical monthly dividers that represents the user's yearly posting behavior in a set of newgroups. Each month is divided into weeks and each week is shown as a vertical line of circles. Each circle represents a conversation whose size indicates the number of author's messages in that thread. Authorlines places threads that were initiated by the author above the timeline, whereas those threads to which the author has contributed but which were not initiated by her are placed underneath the timeline.

3.3.2 Proposal of Student Data Portrait

The data portrait proposed in this paper creates a snapshot of an individual student from her learning process's data. Each portrait appears when the instructors moves the mouse over a card from the proposed faceted browser. The items represented in the portrait are the aforementioned facets except gender.

The data portrait (see Figure 3) is a bar that is divided into five squares: number of initial posts (in dark blue), number of replies (light blue), number of read messages (pink),

(a) Active   (b) Threads initiator   (c) Replies a lot   (d) Lurker

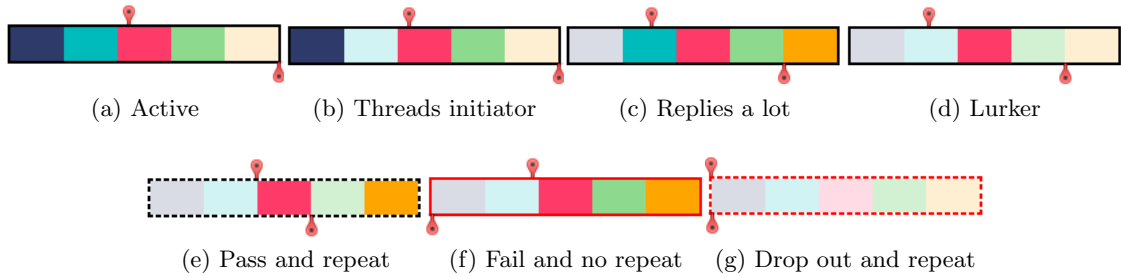(e) Pass and repeat   (f) Fail and no repeat   (g) Drop out and repeat

Figure 3: Different examples of student data portrait.

the average grade of the messages (green) and the number of messages highlighted (orange). Therefore, these five areas represent the student's participation data. Each one of them change its opacity in order to indicate the level of achievement. The lighter the color of the square is, the lower student's performance in that item is, and vice versa.

Likewise, the border of the bar can be red or black. Red means that the student would not pass if the course finished at that moment. Moreover, this can be drawn with dashes. This indicates that the learner is repeating the course.

Finally, there are two red markers above and underneath the bar. The former represents the studying pace (in percentage), i.e. how many activities she has already finished. Thereby, the top of the bar works as a continuous axis that goes from 0% to 100%. As far as the second marker is concerned, this indicates the average grade of the assignments. In this case, the bottom of the bar works as a discrete axis whose values are: NA/NP, D, C-, C+, B and A. Each value coincides with the ends of each square.

3.3.3   Examples of Using the Data Portrait Proposals

Figure 3 shows different student behaviors by using the proposed data portrait. Learner in Figure 3a participates actively in forums, since she writes (opaque dark blue) and replies (opaque light blue) a lot of messages and, at the same time, she reads most of messages (opaque pink).

On the other hand, student in Figure 3b initiates a lot of conversations (opaque dark blue), but she does not reply to other messages (transparent light blue). Therefore, she only writes when she initiates the thread. By contrast, learner in Figure 3c helps classmates by participating in threads initiated by others (opaque light blue). However, she does not initiate any conversation. Moreover, the teacher has highlighted some of her messages (opaque orange). As seen, the students in Figure 3b and Figure 3c participate in discussion forums, but in an opposite way.

According to Taylor [27], the learner in Figure 3d would be a lurker. A lurker can be defined as a user that writes occasionally or not at all (transparent dark and light blues), but she regularly participates as a reader (opaque pink). Likewise, lurker student usually obtain similar grades to more active ones. For instance, in Figure 3d, this student has a B as the average grade of her assignments.

The last three data portraits show: Figure 3e) a student that is repeating the course (dashed border) and, at that moment, she would pass the course (she has a C+); Figure 3f) a lurker who does activities, but she fails the subject (red border and bottom marker in the beginning); and Figure 3g) a student who has clearly dropped out (all squares are transparent, her studying pace is low and she has not handed in any assignment).

In short, thanks to this data portrait, it is possible to compact all student's information in a single picture. Thereby, instructors can obtain an overall idea of the student's learning process at a glance.

4.   CONCLUSIONS AND FUTURE WORK

Online education involves a student-centered approach in which the instructor is a facilitator. Consequently, teachers should be guides who help their students to achieve learning goals successfully. To carry out this role properly, instructors should be provided with the maximum amount of data about both the overall classroom and a particular student.

Regarding this necessity, this paper introduces a novel approach of educational monitoring tool whose contribution is twofold. On the one hand, the use of facet browsing to monitor a classroom is proposed. As seen, facet browsers provide a user-friendly way to navigate through data collections. With regard to its use in online education, faceted browsing may help instructors gain understanding of the behavior of their classrooms, since it allows instructors to explore and analyze their classrooms by requesting iterative tailored queries based on different combinations of facets.

On the other hand, this paper promotes to employ a novel technique called data portrait that allows to compact the user's information in an only picture. With regard to its use in online education, a data portrait can be a good way to summarize the student's data visually, allowing teachers to gain understanding of student's behavior at first glance.

This paper also proposes a classification of educational monitoring tools from two points of view. The first one focuses on the data processing technique that is used, i.e. information visualization and data mining, whereas the second one is based on the element that is monitored. Likewise, this paper presents a study on the most relevant graphical features that tools based on IV usually have.

As future work, an experiment with teachers throughout a term must be conducted so as to check the effectiveness and usefulness of both the faceted browser and the data portrait. After the experiment, a survey could be sent in order to detect strong and weak points of the proposal. Likewise, a focus group or semi-structured interviews with 4-6 people may also be useful to know the instructors' opinion in depth. From the results of the survey and focus group/interviews, valuable conclusions may be drawn. Thereby, new features and improvements may be defined and added to the tool.

5. REFERENCES

[1] A. Author and A. Author. Paper blinded because this is related to this work. In *International Conference on*, pages 1–3, 2011.

[2] R. S. Baker and K. Yacef. The State of Educational Data Mining in 2009: A Review and Future Visions. *JEDM - Journal of Educational Data Mining (ISSN 2157-2100)*, 1(1):3–17, Oct. 2009.

[3] S. Card, J. Mackinlay, and B. Shneiderman. *Readings in information visualization: using vision to think.* The Morgan Kaufmann series in interactive technologies. Morgan Kaufmann Publishers, 1999.

[4] N. Dabbagh and A. Kitsantas. Supporting self-regulation in student-centered web-based learning environments. *International Journal on E-Learning*, 3(1):40–47, 2004.

[5] J. Donath, A. Dragulescu, A. Zinman, F. Viégas, and R. Xiong. Data portraits. In *SIGGRAPH '10: ACM SIGGRAPH 2010 Art Gallery*, pages 375–383, New York, NY, USA, 2010. ACM.

[6] A. C. Dragulescu. *Data Portraits: Aesthetics and Algorithms.* PhD thesis, Massachusetts Institute of Technology, 2009.

[7] N. Hara and R. Kling. Students' distress with a web-based distance education course. *Information, Communication and Society*, 3(4):557–579, 2000.

[8] J. Hardy, S. Bates, J. Hill, and M. Antonioletti. *Tracking and Visualisation of Student Use of Online Learning Materials in a Large Undergraduate Course*, volume 4823 of *Lecture Notes in Computer Science*, pages 464–474. Springer Berlin / Heidelberg, 2008.

[9] R. Hijón and Ángel Velázquez. E-learning platforms analysis and development of students tracking functionality. In E. Pearson and P. Bohman, editors, *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2006*, pages 2823–2828, Chesapeake, VA, June 2006. AACE.

[10] R. Hijón-Neira and J. A. Velázquez-Iturbide. How to improve assessment of learning and performance through interactive visualization. In *ICALT '08: Proceedings of the 2008 Eighth IEEE International Conference on Advanced Learning Technologies*, pages 472–476, Washington, DC, USA, 2008. IEEE Computer Society.

[11] J. L. Hung and K. Zhang. Revealing Online Learning Behaviors and Activity Patterns and Making Predictions with Data Mining Techniques in Online Teaching. *Journal of Online Learning and Teaching*, 4(4):428–437, 2008.

[12] I. Icke and E. Sklar. A visualization tool for student assessments data. In *From Theory to Practice: Design, Vision and Visualization Workshop. IEEE VisWeek (2008)*, 2008.

[13] A. A. Juan, T. Daradoumis, J. Faulin, and F. Xhafa. Developing an information system for monitoring student's activity in online collaborative learning. In *CISIS '08: Proceedings of the 2008 International Conference on Complex, Intelligent and Software Intensive Systems*, pages 270–275, Washington, DC, USA, 2008. IEEE Computer Society.

[14] J. Koren, Y. Zhang, and X. Liu. Personalized interactive faceted search. In *Proceeding of the 17th international conference on World Wide Web*, WWW '08, pages 477–486, New York, NY, USA, 2008. ACM.

[15] E. Kosba, V. Dimitrova, and R. D. Boyle. Using Student and Group Models to Support Teachers in Web-Based Distance Education. In L. Ardissono, P. Brna, and A. Mitrovic, editors, *UM2005 User Modeling: Proceedings of the 10th International Conference*, pages 124–133, Edinburgh, Scotland, 2005. Springer Verlag.

[16] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Trans. Graph.*, 5(2):110–141, 1986.

[17] D. Maor. The teacher's role in developing interaction and reflection in an online learning community. *Educational Media International*, 40(1/2):127–138, 2003.

[18] R. Mazza. Evaluating information visualization applications with focus groups: the coursevis experience. In *Proceedings of the AVI workshop on Beyond time and errors: novel evaluation methods for information visualization*, BELIV '06, pages 1–6, 2006.

[19] R. Mazza and V. Dimitrova. Visualising student tracking data to support instructors in web-based distance education. In *13th International World Wide Web Conference - Educational Track*, pages 154–161. ACM Press, 2004.

[20] R. Mazza and C. Milani. Gismo: a graphical interactive student monitoring tool for course management systems. In *T.E.L.'04 Technology Enhanced Learning '04 International Conference. Milan*, pages 18–19, 2004.

[21] C. McInnis, R. James, and C. McNaught. First year on campus. a commissioned project of the committee for the advancement of university teaching. Technical report, Canberra: Australian Government Publishing Service, 1995.

[22] A. Merceron and K. Yacef. Interestingness measures for associations rules in educational data. In *EDM'08*, pages 57–66, 2008.

[23] T. J. F. Mitchell, S. Y. Chen, and R. D. Macredie. The relationship between web enjoyment and student perceptions and learning using a webâĂŘbased tutorial. *Learning, Media and Technology*, 30(1):27–40, 2005.

[24] OECD. Learning for tomorrows world | first results from pisa 2003. Technical report, Organisation for Economic Co-operation and Development (OECD), 2004.

[25] R. Pirrone, V. Cannella, and G. Russo. A map-based visualization tool to support tutors in e-learning 2.0. In *HSI'09: Proceedings of the 2nd conference on Human System Interactions*, pages 482–487, Piscataway, NJ, USA, 2009. IEEE Press.

[26] C. Romero and S. Ventura. Educational data mining: A survey from 1995 to 2005. *Expert Syst. Appl.*, 33(1):135–146, 2007.

[27] J. Taylor. Teaching and learning online: the workers, the lurkers and the shirkers. In *2nd Conference on Research in Distance and Adult Learning in Asia, CRIDALA 2002*, 5–7 June.

[28] F. B. Viégas. Newsgroup crowds and authorlines:

Visualizing the activity of individuals in conversational cyberspaces. In *In: Proceedings of the 37th Hawaii International Conference on System Sciences, IEEE*, 2004.

[29] W. R. Watson and S. L. Watson. An Argument for Clarity: What are Learning Management Systems, What are They Not, and What Should They Become? *TechTrends*, 51(2):28–34, Mar. 2007.

[30] R. Xiong and J. Donath. Peoplegarden: creating data portraits for users. In *UIST '99: Proceedings of the 12th annual ACM symposium on User interface software and technology*, pages 37–44, New York, NY, USA, 1999. ACM.

[31] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst. Faceted metadata for image search and browsing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, CHI '03, pages 401–408, 2003.

[32] H. Zhang, K. Almeroth, A. Knight, M. Bulger, and R. Mayer. Moodog: Tracking students' online learning activities. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007*, pages 4415–4422, 2007.

[33] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Comput. Surv.*, 35:399–458, December 2003.

[34] M. Zorrilla, D. García, and E. Álvarez. A decision support system to improve e-learning environments. In *Proceedings of the 2010 EDBT/ICDT Workshops*, EDBT '10, pages 11:1–11:8, 2010.