Data Warehousing for Improving Web-based Learning Sites

F. Araque[1], A. Salguero[1], L. Martínez[2], E. Navarro[1] and M. Calero[1]
[1] University of Granada, Andalucía, Spain
[2] University of Jaén, Andalucía, Spain

***Abstract*— The use of Data Warehouses (DW) in e-learning applications is very helpful in assessing the students from different points of view. We can take advantage of the use of a computer based system to get information difficult to measure in traditional education. Moreover, a DW system helps us to enhance the personalization and content distribution. Depending on the user behaviour we can modify the content or the appearance of the e-learning platform to achieve the best results. In this paper we present our work related to the use of integrated DW as part of the e-learning application to help teachers and administrator in the decision-making process.**

***Index Terms*—Data Warehousing, Decision Support System, e-learning, e-Business.**

I. INTRODUCTION

In recent years, university education has been undergoing many changes, most notably the increasing demand for higher quality in teaching and for greater flexibility in structure and teaching possibilities. Additionally, competition between university centers is increasing and territorial diversification is taking place on a larger scale. For several years there has been an increase in the number of students who are unable to follow the traditional format of university education, owing to the time and space constraints inherent in regular attendance at traditional classes. At the same time, teaching methods are also undergoing a transformation, with a greater emphasis on student-centered learning, enabling the student to acquire more direct control over his or her learning rather than depending exclusively on the lecturer [10].

In view of recent technological developments, new societal demands and the changes in higher education mentioned above, there is a need to design and apply new computer programs aimed at enhancing the flexibility of teaching with regard to time and place, and at adapting courses to the individual needs of each student [13].

Recently, there has been an increase in the number of e-learning program applications in the field of education. A Learning Management System (LMS) is a software package, usually on a large scale, that enables the management and delivery of learning content and resources to students. Most LMSs are web-based to facilitate "anytime, anywhere" access to learning content and administration. These systems can be organized in:

- Content Management System (CMS) [17]: it is an Internet-based software that manages student enrolment, tracks student performance, and creates and distributes course content; it enables teachers to extend the classroom beyond its traditional boundaries of time and space. Examples of this kind of systems are: WebCT, Virtual-U or Blackboard.
- Enterprise Learning Management System (ELMS): it is a more complex and expensive software that includes a lot of functionalities. It is necessary to customize it to each enterprise and to connect to database system.

Most of the Enterprises and Universities use a CMS. But with the new virtual environments for teaching and learning there are different roles for lecturers and students. Use of the web and interactive programs enables lecturers to enhance the scope and flexibility of the teaching process, and offers a new resource for testing innovative teaching programmers [15]. At the same time, lecturers are becoming aware of the possibilities for changing the ways in which they teach and their students learn [10]. This implies new forms of personalized evaluation for each student. However, actual CMSs do not support tools for data analysis which can allow teachers to assess their students according to their evolution. Teachers really have not control of the evolution of theirs students (how the student navigates inside the e-learning platform), and the students cannot consult their restlessness and problems in an easy way.

This work is continuation of a previous one [3], [5], [8] and presents the use of a DW to analyze the behavior of the users (students) of the e-learning platform to make decisions with respect to their assessment. The DW is fed with the data that are generated by the students when navigating by the e-learning platform. Afterwards, we use On Line Analytical Processing (OLAP) techniques [11] in order to generate valuable reports that will be used to improve the e-learning platform and help in the learning evaluation.

There are some other works related to the same subject, i.e. the use of DW techniques to analyze the student's behavior in an e-learning environment [18]. In [19], the Analysis Services of Microsoft SQL Server is used to analyze the Web server log file. An application called MultiStar was developed in [12] for knowledge discovery. In this application, the user interacts with a diagram representing the star schema – the most used DW schema. The main difference between those projects and ours is that in the formers the analysis of the data in the DW is carried out using tools not accessible from the e-learning platform. This approach has the inconvenient of implying minimal computer skills for the teacher to install and configure the OLAP tool. As a result, the data

analysis is usually performed only few times during the course (only once at the end) because it often requires the help of experts to be performed.

Thus, in this paper we present our work related to the use of integrated DW as part of the e-learning application to help teachers and administrator in the decision-making process and the application of OLAP analysis to obtain data that will be used to improve the relationship between teachers and students.

This paper is organized as follows: next section introduces in short the concepts of Data Warehouse and OLAP. Third section presents the functional architecture of the e-learning platform. In fourth section a case study and some examples are presented. Finally, we end up with conclusions and future work.

II. DATA WAREHOUSE

It is obvious that there is no organization running without data. The data can be viewed as tangible assets of an organization just as any physical asset. So, they need to be stored and made available to those who need them in order to be used at any moment. Since the data by themselves are useless, they must be put together to produce useful information. In turn, information becomes the basis for relational decision making. To facilitate the decision-making process, a new piece of technology more sophisticated than a database system was developed and called Data Warehouse (DW). The DW can be generally described as a decision-support tool that collects its data from operational databases and various external sources, transforms them into information and makes that information available to decision-makers (top managers) in a consolidated and consistent manner [9][11]. The persistence of huge amounts of data (possibly distributed and heterogeneous) opens a new perspective for various statistical analysis methods which are essential for strategic decisions in e-learning.

Inmon defined a DW as "a subject-oriented, integrated, time-variant, non-volatile collection of data in support of management's decision-making process" [9]. A DW is a database that stores a copy of operational data which structure is optimized for query and analysis. The scope is one of the DW defining issues: it is the entire enterprise. Related to a more reduced scope, a new concept is defined: a data mart (DM) is a highly focused DW which scope is a single department or subject area.

The DW and data marts are usually implemented using relational databases defining multidimensional structures. The generic architecture of a DW is illustrated in Figure 1 [7]. Data sources include existing operational databases and flat files (i.e., spreadsheets or text files) in combination with external databases. The data are extracted from the sources and then loaded into the DW using various data loaders and ETL tools [2]. ETL stands for extract, transform and load, the processes that enable companies to move data from multiple sources, reformat and cleanse it, and load it into another database or on another operational system to support a business process.

The warehouse is then used to populate the various subject (or process) oriented data marts and OLAP servers. Data marts are subsets of a DW categorized according to functional areas depending on the domain (problem area being addressed) and OLAP servers are software tools that help a user to prepare data for analysis, query processing, reporting and data mining. Thus, a DW coupled with OLAP enables managers to creatively approach, analyze and understand the problems. The OLAP analyzes data using special DW schemas and enables users to view data using any combination of variables.

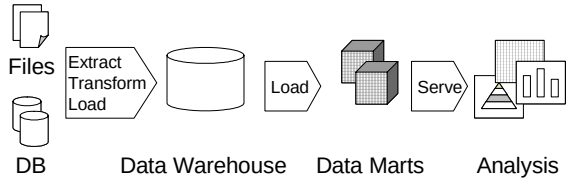


Figure 1. A generic Data Warehouse architecture.

To extract this information from a distributed environment, we would need to query multiple data sources and integrate the information at a particular point before presenting the answers to the user. In a DW environment, such queries find their answers in a central place, thus reducing the processing and management costs. After the initial loading, warehouse data must be regularly refreshed. Modifications of operational data since the last DW refreshment must be propagated into the warehouse such that warehouse data reflect the state of the underlying operational systems [1][2].

III. FUNCIONAL ARCHITECTURE

We can see the functional architecture in figure 2. As we can observe the e-learning platform feeds the DW with the necessary data. After, the Data Mart is loaded with the data useful for the analysis according to the proposed design. The educator will be able to consult the data of Data Mart through an interface inside the e-learning platform (in the professor area). So, he/she will be able to make the appropriate decisions with respect to the evaluation of the student.

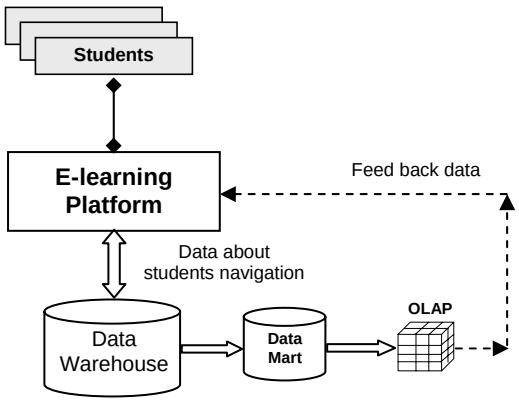


Figure 2. Functional architecture

Based on the structure of the e-learning platform that we have developed in previous works [4], we defined a DW to analyze the behavior of the users in the system. Besides, to provide interesting information when evaluating the result of the students, it allows us to evaluate the usage degree of our system. Using the platform for this purpose has several advantages. We can answer questions like:

- How long the student is connected to the platform (by theme, course, section, day, month, etc)?
- How many sessions does a student finish over time?
- Which is the most accessed page?
- Which is the most popular section?
- Which is the connections percentage from University computers?
- Which is the network distribution traffic over a day, week, etc? And so on.

To answer question like those, we can define a multidimensional scheme applying the concepts presented in section 2. Basically we have five dimensions: activity, student, source, target and time, and one fact table called Navigation. A more detailed explanation can be found in [4].

The multidimensional scheme is centered in the clicks that the users of the system make in their navigation by the platform. It is a simple table, like navigation by hypermedia systems. The rest of dimensions of the scheme characterize this navigation. The dimensions source and target represent the positions of the nodes from and towards which the navigation is realized (between anchors).

We did not define traffic as the volume of information (in bytes) as it is greatly influenced by the size of the files. The analysis is based on the frequency of access by students and by teachers to the platforms in general.

A. E-learning Platform description

We have used ASP.NET to develop the e-learning platform. Microsoft.NET is the new generation of software that combines Technologies of the Information and Communications to transform the use of the Web. It allows the creation and the use of distributed Web Services and supplies of accessible business from multiple platforms through the Internet without taking into account of its technical implementation. ASP.NET is a framework of programming generated in Common Language Runtime that can be used to generate effective Web applications. It offers several important advantages over the previous programming Web models. When a client (a student or a teacher) requests a resource to the Web server, which contains the application, this one sends the request to the ASP.NET runtime along with enough information so that combining it with the stored data in the data base or in the files in XML format can generate a proper answer to the client. XML is no more than a set of rules to define semantic labels that organize a text document in different parts. XML is actually a metalanguage that defines the used syntax to define other languages of structured labels. We use it to deal with documents that are suitable for being exposed as a service to another application, like the register of students, their grades, the description of the lessons, etc. The rest of the data are stored in a Microsoft Access Data Base, which is more efficient than XML files for storing information.

Since the sequences of commands are executed in the server, and not in the client software, the server is responsible to make all the necessary work to generate the pages that are sent to the client. The sequences of commands are hidden to the users, so these only receive the result of the execution in plain HTML format. The problem of the client being able (or not) to execute sequences of commands disappears, since the Web server sends the result in interpretable standard HTML code to the user browser.

The platform has two different user interfaces depending on the kind of user (teacher or student) that use it. Both interfaces have the screen divided in two zones.

On the left side we can see the identification section, the options menu, the surveys that have been defined and the calendar. Each student has an account in the system that is activated only when the student fills up a card with his/her personal information. Once the user has logged in the menu buttons become active and the user can start working. There are different kinds of users that can use the system. Therefore, whenever users access the system, they are offered different services depending on the type of privileges of the different users.

Like many other e-learning applications, ours is based on learning modules. Each module contains a set of didactic materials and an evaluating method. Instructors can offer and learners can sign up for whole courses that are scheduled to run over a fixed duration of time. Courses are structured by combining a number of didactic modules. These modules can be selected to be visible for students from a repository at any time. Students can select the modules that wish to work on. In order to develop these modules the collaboration of several students may be necessary. The result of these must be sent to the professor through this tool. This mechanism saves the professor the work of collecting the students assignments, by creating an automatic control of dates of delivery. After the predefined code of the module, the teacher can specify its characteristics. There are common characteristics to all kind of didactic modules (evaluating method, name, and description) as well as characteristics that depend on the evaluating method (category, date of presentation, etc).

Modules are organized into categories. Each student can choose only an assignment belonging to a specific category. Teachers can define sequences of didactic modules that have to be accomplished in a certain order using the "after category" parameter. If this parameter is set, the student cannot select the module after accomplishing a module pertaining to the category indicated. Each category of assignment has been given an associated percentage that affects the final grade.

Our application allows the use of multiple-choice questions (MCQ) as the basic evaluating method of a module, together with a more traditional one that allows the student the submission of complete projects or case studies. After evaluating these projects or case studies the teachers can make observations to each student individually.

The e-learning system also includes a set of tools to improve the communication between the users: forum, private messaging, Microsoft NetMeeting, surveys, news board. There is also a section that allows the complete modification of the visual aspect. Using this tool the teachers can customize the visual style of their course without the need to have web programming skills. It is based on the programmatically modification of the Cascade Style Sheet (CSS) files that define the appearance of the HTML elements of the pages.

IV. DATA ANALYSIS

The e-learning platform design and development, and content delivery are important keys as they are directly related to the nature of the online activities students are engaged in, and hence also to the level of engagement they have in the activities. A key factor of good design and content delivery involves the concept of interactivity [20], i.e. how students interact with learning materials. Online interactivity involves interactions with either the content which might be text, audio visual resources and quizzes or with people via asynchronous online communication or synchronous communication. Students can interact with the teacher and with peer learners.

In the present paper we shall describe the platform we have developed and present some results relating to its application over two consecutive academic years in two courses which form part of the BA Honours Degree in Psychology at the University of Granada, Spain.

In psychology, application of educational resources via e-learning platforms is especially interesting since evaluation and treatment resources are increasingly computer-based. However, while the number of published programs is rising exponentially year by year, present use of information technology by psychologists remains at a level similar to that of 1987 [14][16]. In the view of some authors, this phenomenon is due to the lack of adequate training and instruction in information technology for practicing psychologists [16]. With this in mind, the motivation for the present experiment has been to incorporate an e-learning platform as part of the practical teaching of psychology. At present, the e-learning platform is being used in Faculty of Psychology at the University of Granada. A total of 1103 students used the platform. Of these, 72.2% were students following a compulsory third year course (Psychological Evaluation) and 27.8% were students following an optional fourth-year course (Clinical Child Psychology). 83.7% of students were female and 16.3% were male. Only 0.7% had jobs in addition to their studies. 3.1% were over thirty years of age.

The interface has been carefully designed to fit the Data Mart schema described in the previous section. Initially, the dimension hierarchy is presented to the user in form of tables (see figure 3). Each table corresponds to a different dimension. The dimension elements are distributed through each dimension table: the most detailed items are placed at the bottom of the table, whether the most aggregated items are placed at the top. At this point, the user has to select the elements that contain the information he or she wants to include in the report and the elements that contains the restrictions to filter the data. The user can define a list of filters for each dimension element selecting the desired option and clicking on "Manage groups" button. For each element selected its attributes are listed in the next step (figure 4). Each attribute has a verification box. The box allows users to decide whether the corresponding attribute value must appear as a column in the output report. It is possible to filter the result by creating groups of restrictions. For each group the members of a dimension are shown along with one or two the text fields (one for alphanumeric parameters and two for numeric ones) which allow users to specify conditions in order to filter the output report. In this way, it is really easy to make complex reports, where many parameters and restrictions are involved.

Let suppose a teacher who wants to know the first section which students visit after being identified. Furthermore, he or she wants to analyze the evolution of this value over the second and third quarter.

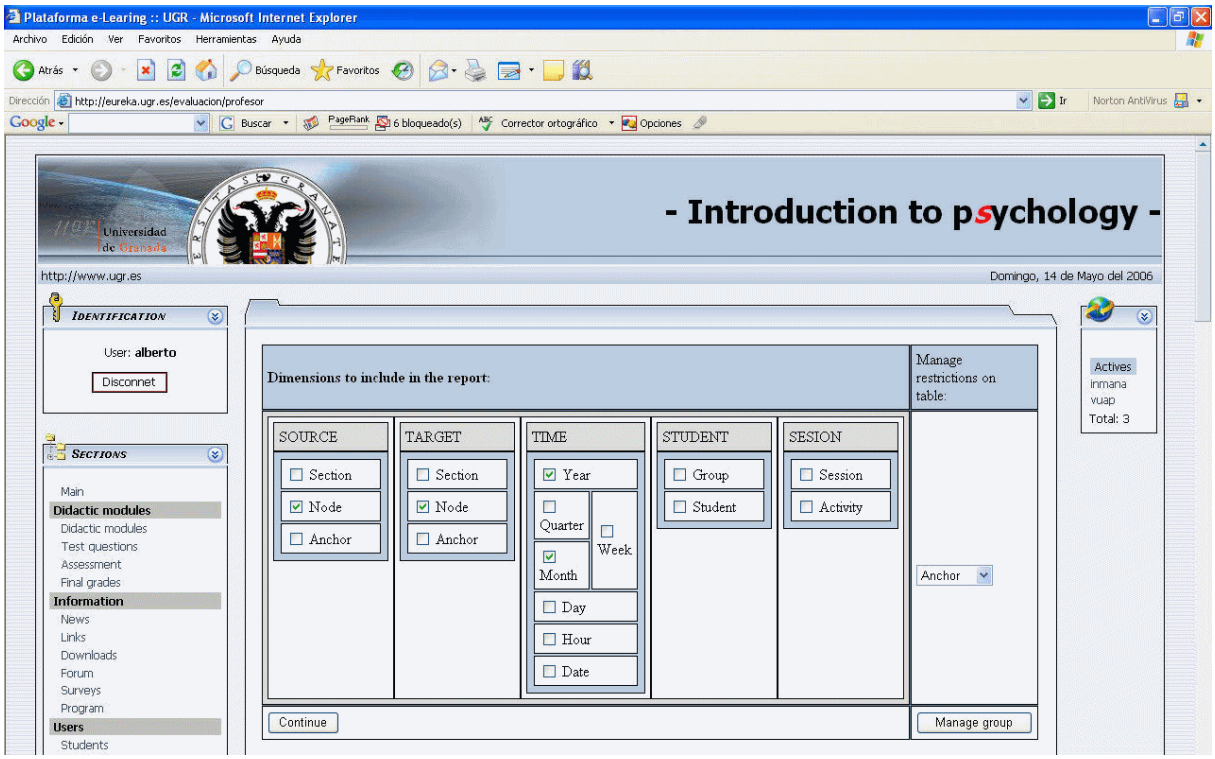


Figure 3. OLAP interface. Dimension selection.

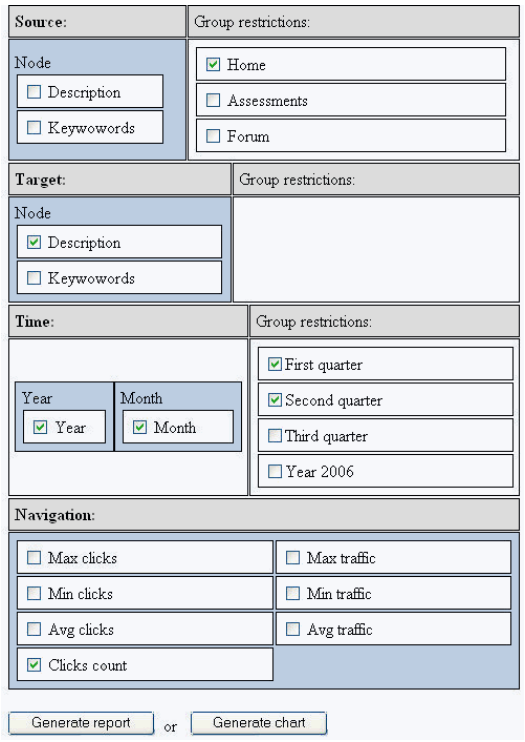


Figure 4. OLAP Interface. Attribute selection and filtering.

This information can be used by the teacher to find out the information of interest of the students with respect to the course progress. To get this information, the teacher has to select the member Node in both Source and Target hierarchy, Year and Month. After selecting them, he or she has to select the attributes or filters associated to each of them that are of interest.

Selecting the Target→Node→Description attribute a column in the report will be included to specify the destination. To consider only the displacements that are performed from the initial page, the filter Source→Home are checked. In the same way, for restrict the report to the first and second quarter, Time→"First quarter" and Time→"Second quarter" attributes are checked. Obviously, it is necessary to select the Time→Month attribute to identify the results by months. Attribute Time→Year is selected for data not being aggregated through years. The selection of Time→Year attribute allows us to find out if it has been any difference between courses. Each selected member of the dimensions turns into a column in the report.

Once we have selected the desired dimension attributes it is necessary to select at least one measurement from the fact table. In this case, to get the total clicks performed by the students, it is needed to select the attribute Navigation→"*Clicks count*". The result of this query can be seen in figure 5.

When teachers have to assess the work of the students they can easily make detailed reports about their behaviour. For example, for setting the final grade of a work group, the global participation of all component of the group can be considered, without considering the individual participation of each student.

To get a report with this information, the teacher has to select the member *Activity* in *Session* hierarchy and *Date* member in *Time* hierarchy. Then, he or she has to check the attributes Session→Activity→Description and Time→Date. The filter Session→"Issue27 - Work group interval" is checked to consider only the activity during the interval between the start and the end of the specific work group. To get the global clicks performed by the students in the work group the teacher has to select the attribute Navigation→"Clicks count". The result of this query can be seen in figure 6.

V. CONCLUSIONS

In this paper we have presented our work related to the utilization of DW as an integrated module in the e-learning platform to help teachers, students and administrator in the decision-making process. Besides, we have exposed how the application of OLAP analysis is useful to obtain data that later will be used to improve the relationship between teachers and students.

In the near future, we have planned to incorporate to the platform concepts of adaptive hypermedia, trying to cause that the e-learning platform adapts automatically to each one of the students who use it. For achieve this goal, we are based on the ideas on active DW [6], that combine mechanisms based on ECA rules (event-condition-action) with the ability of analysis of the DW in order to make react to the system when certain situations take place.

Result:

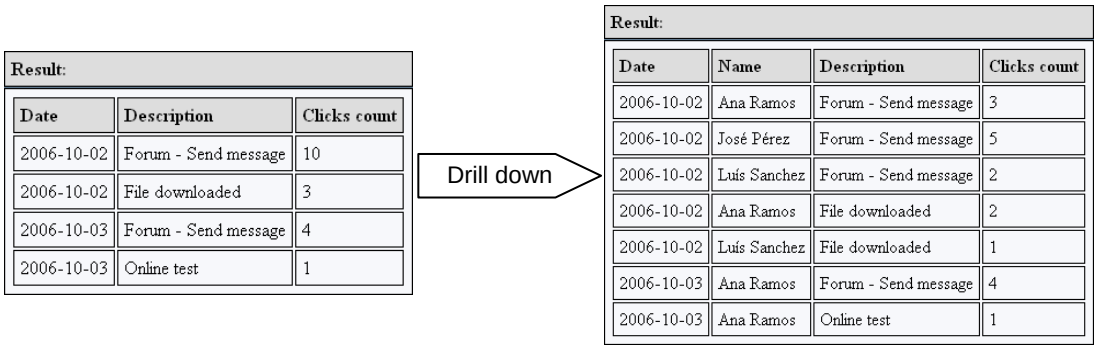
Year	Month	Description	Clicks count
2006	1	Schedule	1245
2006	1	News	1357
2006	1	Downloads	1800
2006	2	Schedule	903
2006	2	News	1235
2006	2	Downloads	505
2006	3	Schedule	503
2006	3	News	1465
2006	3	Downloads	457

Roll up

Result:

Year	Description	Clicks count
2006	Schedule	2651
2006	News	4057
2006	Downloads	2762

Figure 5. Left: Report about evolution over quarter. Right: Roll-up by Month. In OLAP, Roll-up is defined like the computation of aggregates that are coarser than existing precomputed aggregates.



Result:

Date	Description	Clicks count
2006-10-02	Forum - Send message	10
2006-10-02	File downloaded	3
2006-10-03	Forum - Send message	4
2006-10-03	Online test	1

Drill down

Result:

Date	Name	Description	Clicks count
2006-10-02	Ana Ramos	Forum - Send message	3
2006-10-02	José Pérez	Forum - Send message	5
2006-10-02	Luís Sanchez	Forum - Send message	2
2006-10-02	Ana Ramos	File downloaded	2
2006-10-02	Luís Sanchez	File downloaded	1
2006-10-03	Ana Ramos	Forum - Send message	4
2006-10-03	Ana Ramos	Online test	1

Figure 6. Left: Report about work group global participation. Right: Drill-down by Name. In OLAP, Drill-down operation is defined like the act of moving from the general to the specific.

REFERENCES

[1] Araque, F. (2003). Integrating heterogeneous data sources with temporal constraints using wrappers. The 15th Conference On Advanced Information Systems Engineering. Klagenfurt/Velden. Austria.

[2] Araque, F. (2003). Real-time Data Warehousing with Temporal Requirements. Decision Systems Engineering, DSE'03 (the CAISE'03 conference). Klagenfurt/Velden, Austria.

[3] Araque, F. Salguero, A., Hlavickova, V. (2004). Plataforma de E-learning para apoyo a la docencia. 6º Simposio Internacional de Informática Educativa (SIIE04). 16-18, Cáceres, Spain.

[4] Araque, F., Salguero, A., Calero, M.D., Fernández-Parra, A., Jiménez Morales, M.I., Vives, M.C., Navarro (2007). E. E-learning platform as a teaching support in psychology. EUROCAST, Las Palmas de Gran Canaria, Canary Islands, Spain.

[5] Araque, F., Gaitán, J.J., Vlasta Hlavickova (2003). La informática en los estudios de Gestión y Administración Pública s. IV Jornadas de Enseñanza Universitaria de la Informática (JENUI'03). Cádiz, Spain.

[6] Brobst, S. (2002). Active data warehousing: a new breed of decision support. Database and Expert Systems Applications, 2002. Proceedings. 13th International Workshop on , vol., no.pp. 769-, 2-6 Sept.

[7] Chaudhuri, S., & Dayal, U. (1997). OLAP technology and data warehousing, ACM SIGMOD Records, Hewlett-Packard Labs.

[8] Gaitán, J.J, Araque, F. (2002). La plataforma de e-learning como soporte a las prácticas de E-Business. International Conference on ITC´s in Education.

[9] Inmon W.H. (2002). Building the Data Warehouse. John Wiley.

[10] Jelfs, A., & Colbourn, C. (2002). Virtual seminars and their impact on the role of the teaching staff. Computers & Education, 38, 127- 136.

[11] Kimball, R., & Ross, M. (2002). The Data Warehouse Toolkit: The Complete Guide To Dimensional Modeling, 2nd Edition. John Wiley.

[12] Lam, P., Keing, C., McNaught, C., & Cheng, K. F. (2006). Monitoring eLearning environments through analyzing web logs of institution-wide eLearning platforms. In Who¡¦s learning? Whose technology?. Proceedings of the 23rd annual Australian Society for Computers in Learning in Tertiary Education conference, University of Sydney, Australia.

[13] Masiello, I., Ramberg, R., & Lonka, K. (2005). Attitudes to the application of a web based learning system in a microbiology course. Computers & Education, 45, 171-185.

[14] McMinn, M.R., Buchanam, T., Ellens, B.M., & Ryan, M.K. (1999) Technology, professional practice and ethics: survey findings and implications. Professional Psychology: Research and Practice, 30, 165-192.

[15] Neo, M. (2005). Web enhanced learning: engaging students in constructivist learning. Campus-Wide Information Systems 22, 4-14

[16] Olson, K.R. (2001) Computerized psychological test usage in APA-accredited Training Programs. Journal of Clinical Psychology, 57 (6) 727-736.

[17] Pastore, S. (2006). Web Content Management Systems: using Plone open source software to build a website for research institute needs. Digital Telecommunications, 2006. ICDT '06. International Conference on , vol., no.pp. 24- 24, 29-31.

[18] Silva, D. R., & Vieira, M. T. P. (2002). Using data warehouse and data mining resources for ongoing assessment of distance learning. In P. Kommers, V. Petrushin, X. Kinshuk & I. Galeev (Eds.), ICALT 2002 (pp. 40–45). Proceedings of the IEEE International Conference on Advanced LearningTechnologies, Kazan, Tatarstan, Russia.

[19] Zorrilla, M., Menasalvas, E., Marín, D., Mora, E., Segovia, J. (2005). Web Usage Mining Project for Improving Web-Based Learning Sites. Lecture Notes in Computer Science 3643 Springer, ISBN 3-540-29002-8.

[20] Swan, K. (2003). Learning effectiveness online: What the research tells us. In J. Bourne and J. C. Moore (Eds), Elements of quality online education: Practice and direction (pp. 13–46). Vol. 4, Sloan-C series. Olin Way, Needham, MA: The Sloan Consortium.

AUTHORS

F. Araque was with the Department of Computer Science, University of Jaén. He is now with the Department of Software Engineering, University of Granada, C/ Periodista Daniel Saucedo Aranda s/n, E-18071, Granada (Andalucía), Spain. (e-mail: faraque@ugr, faraque@ujaen.es).

A. Salguero is with the Department of Software Engineering, University of Granada, C/Periodista Daniel Saucedo Aranda s/n, E-18071, Granada (Andalucía), Spain. (e-mail: agsh@ugr).

L. Martínez is with Department of Computer Science, University of Jaén, Campus Las Lagunillas s/n, 23071, Jaén (Andalucía), Spain. (email: martin@ujaen.es).

E. Navarro is with Faculty of Psychology, University of Granada, Campus U. de Cartuja, E-18071, Granada (Andalucía), Spain. (e-mail: enavarro@ugr.es).

M. Calero is with Faculty of Psychology, University of Granada, Campus U. de Cartuja, E-18071, Granada (Andalucía), Spain. (e-mail: enavarro@ugr.es).

Manuscript received 05 February 2007. This work has been supported by the Spanish Research Program under project TIN2005-09098-C05-03, by The Ministry of Education and Science of Spain through Project of Teaching Innovation, and by Andalucía Research Program under project 2006/282916.

Published as submitted by the authors.