Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

A Methodology for Mining Data from Computer-Supported Learning Environments

Uma Metodologia para a Mineração de Dados Oriundos de Ambientes de Aprendizagem Apoiados por Computadores

Abstract
Computer-supported learning environments are usually adopted as platforms for distance-based education, but are also used as supporting tools for face-to-face educational settings. However, in such situations educators lose contact with their students and the way they access and use the content made available to them. This paper presents a methodology to process data collected from server logs and from the environments internal databases to provide feedback to authors and tutors about the usage of the content they offer and students about their behavior inside the environment. Two clustering algorithms, K-means and Self-Organizing Maps, were used to analyze the collected users' interaction data and thus establish patterns of content access. An evaluation was performed with data collected from an actual environment used at a Brazilian university. Student usage and document accessing were both clustered and analyzed. A document access summary was constructed to allow tutors to verify interest in the available resources and to allow students to check their resource usage history.
Keywords: Data Mining. Web Mining. Feedback. E-Learning. Learning Environment Evaluation.

Resumo
Ambientes de aprendizagem apoiados por computadores são normalmente adotados como plataformas para a educação a distância, mas são também utilizados no apoio na educação presencial. Entretanto, a intermediação tecnológica faz com que os educadores percam, nesses casos, o contato com os estudantes e o modo pelo qual eles acessam e utilizam os conteúdos disponibilizados. Este artigo apresenta um método para processar dados coletados de registros de servidores e de bancos de dados internos desses ambientes para oferecer retorno a autores e tutores sobre o uso dos conteúdos disponibilizados, bem como a estudantes sobre seu próprio uso dos recursos do ambiente. Dois algoritmos de agrupamento foram usados para analisar os dados de interação coletados e para reconhecer padrões de acesso a conteúdos. O método foi avaliado com dados coletados de um ambiente usado em uma universidade brasileira, considerando o acesso a documentos e uso do ambiente por estudantes. Um sumário dos dados de acesso coletados permite que tutores verifiquem o interesse dos estudantes nos materiais disponibilizados e que estudantes verifiquem seu histórico de acesso a esses recursos.
Palavras-chave: Mineração de dados. Mineração Web. Retorno. Aprendizagem apoiada por computador. Avaliação de ambientes de aprendizagem.

RICARTE, Ivan Luiz Marques; FALCI JUNIOR, Geraldo Ramos. A Methodology for Mining Data from Computer-Supported Learning Environments. *Informática na Educação*: teoria & prática, Porto Alegre, v. 14, n. 2, p. 83-94, jul./dez. 2011.

Ivan Luiz Marques Ricarte
Geraldo Ramos Falci Junior
UNICAMP

1 Introduction

Data mining aims to discover, from a large amount of data, information that is not easily perceived by users. There are several business-related data mining applications, but there are many possibilities for this type of processing also in education. In such settings, records of users' accesses, maintained by application servers, offer significant sources of information.

This article discusses how to mine records from user interactions with a computer supported learning system to provide feedback to actors which are responsible for organizing system contents. For example, information that a resource has not been as accessed by the users as expected may give an indication that the tutor did not make clear that content should be studied (inadequate instructions) or that the content is not being found by users (visibility problems).

Data mining has been applied to data collected from learning environments in many other applications. Romero and Ventura (2007) presented many of these possibilities

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

by examining a wide range of work undertaken in the area from 1995 until 2005. They point out the differences in the processes of data mining applications between commercial and educational applications, differences that encompass mining goals, available data types, and adopted mining techniques. They point out issues such as to whom the mining work is targeted, either to students, to educators, or to learning systems administrators. They also present approaches that have been explored not only in online distance learning environments, but also in other scenarios. Several techniques and approaches for data mining are analyzed, and the authors present an extensive list of works performed by various researchers.

Romero, Ventura and Garcia (2008) provide an overview on the application of several of such data mining techniques to a specific learning environment, moodle, to collect information in order to improve learning outcomes. One of their conclusions is that available data mining tools are too complex for educators, who need easier interfaces to analyze mining results. Dos Santos and Becker (2003) apply web usage mining to evaluate learning sites, whereas Zaiane and Luo (2001) and Sheard *et al*. (2003) used the same type of data source to analyze students' behavior. In general, in such cases all knowledge about the semantics of the environments and their resources is ignored. As modern learning environments combine several independently developed tools, many times with their own accesses logs, there is a great potential to improve the quality of usage analyses when considering this internal data and not only the front-end (web server) interaction.

This paper presents a methodology to organize data for data mining taking into account the application context and internal structure. As a proof of concept, the technique of data clustering is used to find groups with specific behaviors and to detect individuals with behavior isolated from standard groups. By interpreting the features of such individuals and groups, a tutor may identify problematic patterns and, thus, review the environment organization or content. Two techniques were used to cluster the collected data, K-means and Self-Organizing Maps (SOM); their results are compared also in this paper.

2 Data Preparation Methodology

Based on the traditional steps adopted for data mining procedures, a methodology with five steps is proposed here to guide the process of extracting feedback information from data collected from learning environments.

In the first step, the data analyst must inspect potential sources, including the environment and its tools, which may express access behaviors. In typical systems, an internal database keeps this data, which may be scattered among several tables. Other potential sources are the access log of the application server in which the learning environment was deployed, and records of requests from external educational object repositories.

Another step, strongly related to the previous one, is to define the analysis targets. What educators aim to find out from mining data from the environment records? Typically they will be looking for resource usage and access patterns, either to review their organization for future users or to preview students' behaviors for future accesses.

The third step relates the outcomes from the two previous steps, identifying sources which better capture the data which may describe the desired behavior to be analyzed. In this step, data analysts and educators jointly define hypotheses relating collected data to student behavior.

The fourth step is basically the conventional data cleaning step in data mining procedures. The selected data is read and organized in new structures, more adequate for the data mining process. Potential noise and errors in the data are also eliminated or corrected, when possible, at this time. This step is detached from the data mining process here precisely to emphasize that the concept of noisy data may take into account the application domain. Examples of noisy data in a web-based learning environment are repeated accesses in a short period of time, repeated requests for a given resource by the same student, or extremely long sessions.

The final step occurs after the data mining is concluded. In this step, the goal is to check whether the information obtained in the data mining process actually corresponds to the hypotheses expressed in the third step. Questions likely to be raised by educators include if the groups found out in the clustering process correspond to some pattern of behavior or if the number of groups is significant. If results are satisfactory, educators are ready to review their organization of tools and content in the environment. Otherwise, data analysts and educators must go back to the first steps in the methodology and evaluate how better information can be obtained from the available data.

3 A Case Study

For this study, one of the versions of the TIDIA-Ae environment, developed in the FAPESP project Information Technologies for Advanced Internet Development, was adopted (FAPESP, 2010). The TIDIA-Ae environments were developed on top of the Sakai platform, an international effort to develop a common, open architecture for learning systems (SAKAI, 2010). The version adopted in this study was developed in the e-labora laboratory from UNICAMP, which is the same that was adopted for use in the Virtual University of São Paulo (UNIVESP, 2010), a platform for distance learning supported by the state of São Paulo, Brazil. Nevertheless, results from this study may be easily extended to other Sakai-based installations.

The selected TIDIA-Ae environment is also being evaluated by technicians in the Computing Center of UNICAMP as a platform to support regular undergraduate and graduate courses in the university. In this case, it is used primarily to provide documents to students or links to external materials. Data used in this study was derived from some of these evaluation tests. Although this type of use for an online education environment is not as intense as a distance learning environment, it is not uncommon and is sufficient to allow an analysis of the methodology applicability.

3.1 Information Sources

The first step was to evaluate the potential data sources in the environment. Although there was a possibility of altering tools to collect more data, the adopted approach was not to interfere with the original system, which had the advantage of offering immediate access to existing data collected from previously offered courses.

For this study, the access records for the application server (Tomcat) for all courses in the second period of 2008 were obtained. This data is kept by system analysts essentially for debugging and was not considered adequate for the study. Thus, other data sources of data were sought among the internal structures of the Sakai platform, which has over two hundred tables in its internal database. The TIDIA-Ae implementation adds more ta-

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

bles to these. The selected tables for a first study are presented in Fig. 1. The most basic data related to user accesses is registered in the SAKAI_EVENT table. This table is complemented with data from tables SAKAI_SESSION, which holds details for each session in the system, and SAKAI_USER_ID_MAP, which relates users' internal identifiers used in several tables with the data maintained in the SAKAI_USER table.

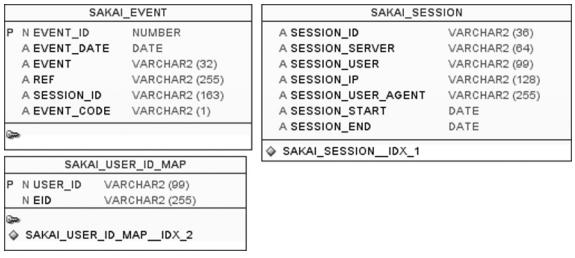


FIGURE 1 – Selected Tables from the SAKAI Environment
SOURCE: Prepared by authors

3.2 Metric Definition and Data Preparation

Given the available data, it is necessary to define what is going to be measured from them. Metrics applied to analyze this type of system are usually related to the time spent by users, mainly students, interacting with the system as proposed in the works of Hsu, Chen, and Tai (2003) and of Castro *et al*. (2007). This type of data is useful to identify patterns of system usage by students and thus to locate students with erratic behavior, or to identify individuals who ignore the system almost completely, motivating the tutor to take preventive actions.

In this study, such information is obtained from the SAKAI_SESSION table which keeps, among other data, information about number of sessions and total time of system access for each user. An analytical database, DATA_SUMMARY, was created to keep data summary. The USER_SUMMARY table keeps the summarized data about sessions.

To ease the implementation of clustering algorithms, these data were organized as two-dimensional vectors. The first dimension keeps the total number of sessions for each user, and the second, his total time of system use, calculated from the initial and final times of each session. The goal is to find out whether are groups of students with similar behavior, thus defining usage patterns and identifying users with unexpected behavior.

3.3 Summary of Access to Resources

In this study, the environment was adopted as a support to regular classes, being used mainly as a repository of documents and media. Thus, it focused on availability and access to resources of the most common types in the system, which were images (whose files had extensions jpg, gif, or bmp), text (pdf, doc, odt), audio (mp3, wav), and presentations (ppt, odp, pdf).

Two tables were constructed using data extracted from the environment event table (Fig. 2). The table DOCUMENT_ UPLOAD_ SUMMARY contains data related to the upload of new resources, associated with events of type *content.new.* The table DOCUMENT_ READ_SUMMARY describes events related to reading, revision, and removal of resources.

Two types of vectors were prepared for the clustering algorithms using data from this database. The first contains the time differences between upload time and each subsequent access of the same resource. It is a vector of variable size, depending on the maximum amount of hits recorded in the system. For clustering algorithms that need to work with vectors of the same size, missing values were

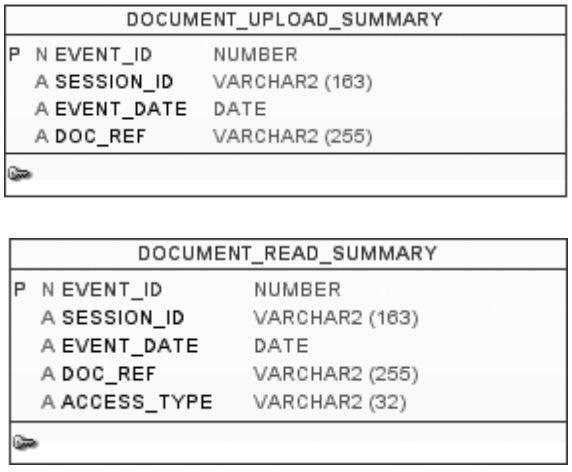


FIGURE 2 – Interaction Summary Tables
SOURCE: Prepared by authors

filled with nulls. The second type, smaller, records three elements, for each document: The total amount of hits, the difference between the upload time and first access, and the time difference between the first and last access.

When users have connection failures when trying to access a resource, it is common that they try to refresh the browser page consecutive times until the page loads. This attitude results in several requests to the system, and a significant proportion of registered read accesses are originated by this behavior. During clean-up, these requests are replaced a single read access to the same document in that session, thus reducing the influence of this problem.

This information reflects, for the tutor, which groups of resources were actually considered more important by students. With this information, tutors can verify how the available material was used by students along the course and whether this practice corresponds to the expected behavior.

3.4 Data Clustering

The application of a data mining technique enables to evaluate whether the adopted metrics are appropriate descriptors for students behavior. This step may require fine tuning by educators, who know which features are relevant in their courses and may identify group behaviors. Even though it would be possible, from the available data, to identify individually each student, this resource was not used here.

Two clustering algorithms were analyzed in this study. The first, K-means, was selected due to its simplicity and efficiency. Another algorithm was the self-organizing map (SOM), available with the MATLAB SOM Toolbox (ADAPTIVE INFORMATICS RESEARCH CENTER, 2005).

K-means applies a simple calculation of the Euclidean distance between each element in a cluster and the possible cluster centers assigning them to the closest centers. At the beginning of each new iteration it adjusts the clusters centers, moving them closer to their elements. As these centers move, elements are reassigned accordingly. The algorithm stops when centers do not change between consecutive iterations, or when a limit of iterations is reached. Its input parameters are the vectors with normalized data, the expected maximum number of groups, the maximum number of iterations, and a base value for random selection of coordinates as initial centers for each group.

Self-organizing maps are simple to use and do not require many adjustments. Experiments were performed encompassing various sizes of maps. Training in each run is done with a different random portion of the input data set and cells are clustered by applying a sequence of executions of a K-Means algorithm internal to the tool, keeping only the

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

best result. The maximum number of groups for this algorithm is defined by $\sqrt{dlen}\sqrt{dlen}$, being *dlen* the quantity of input vectors.

3.5 Data Consolidation

The goal of the data processing step was not only cluster the usage data, but to build information structures that would enable educators to find out relevant information. This final step in the study had thus the main goal of providing summary presentations. In this study, two types of summary were generated. The first summary presented the general behavior of all users' access characteristics. The second was produced from resources access patterns by each single user, with detailed description for these accesses.

At first, the information used to build these summaries was obtained from the summary tables described above. Another approach was also developed which extracted the data directly from the system databases. For the first summary report, starting from each user identification, queries are submitted to the database requesting all resources uploaded by the user, when upload occurred, total number of requests to the resource, how many distinct users requested the resource, and time of last access. Similarly, the second summary report queries the database, for each user, obtaining requested resources, quantity of accesses to each resource, and times for the first and last accesses to it.

In this study, two report formats were presented for each summary. The first format presented the data using an HTML (Hypertext Markup Language) file, for visualization with Web browsers. The other format organized the data in an XML (eXtensible Markup Language) file, to enable integration with other applications.

4 Results

It is clear that the conditions under which this study was performed limit somehow the quality of the collected data, since the collected data refer only to the use of the environment in support to regular classes. As students were not obliged to use the system in most cases, derived conclusions in this case may not be associated with absent students, but to someone that is collecting the material from a colleague, for example. Although there is some prejudice from the pedagogical point of view, it is still possible to validate the proposal. The following sections present the results obtained from the experiments conducted with the available data, focusing on the fourth step.

4.1 Qualifying Clustering Results

The first analyzed results refer to clustering algorithms, both for users' access as for document requests. In both cases, different configurations were used for K-means, with the results stored in XML files for further analysis. Comma delimited values (CSV) files were used to import data in the SOM toolbox, both to train and to process, with the results recorded both as images and as a text files with the list of groups, cells, and members, which was used to compare the results with those obtained by K-means.

One difficulty in the K-means configuration is to estimate the proper number of groups. Plotting the raw data on a Cartesian plane (Fig. 3), it is possible to visualize that there are several groups, but the goal is to find the best number of cohesive groups, associated to reasonable behaviors, without getting groups that are too small, with one or two users only. Thus, the algorithm was run seve-

ral times observing the maximum number of different clusters of 6, 8, 10 or 12.

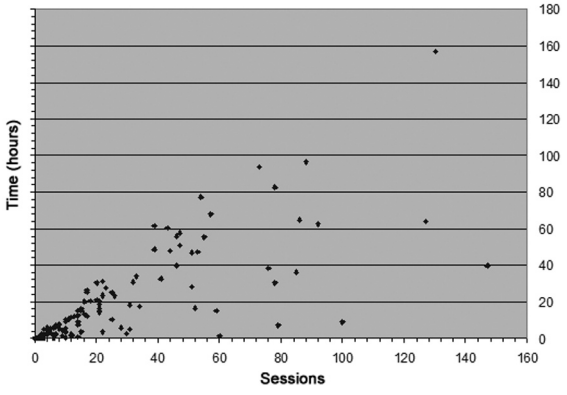


FIGURE 3 – Complete Data Set For Users
SOURCE: Prepared by authors

The clusters centers were arbitrarily set with the initial value 10000. Experimentation has shown that variations in this value had little to no influence in the results. Should small enough values be used it would lead to the identification of a number of groups less than the maximum proposed.

Fig. 4 shows the clustering identified by K-means, adjusted for eight clusters. This figure relates user sessions with the duration of each session. Thus, points close to the origin represent a group of short sessions. It is possible to identify different patterns of user behavior, as the excessively large session times for some users in Cluster 3.

Fig. 5 shows the clustering obtained with self-organizing maps. After training the map and distributing data among its cells, the SOM algorithm applies several K-means on the map. Several values for the number of groups are used and the one with best results, based on squared errors and Davies-Bouldin index, is returned (ADAPTIVE INFORMATICS RESEARCH CENTER, 2005).

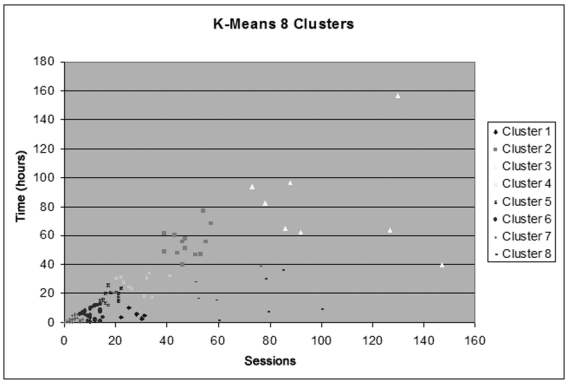


FIGURE 4 – K-means for Users Accesses, Adjusted for Eight Clusters
SOURCE: Prepared by authors

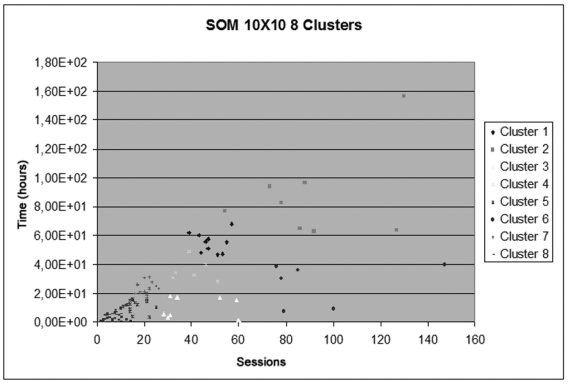


FIGURE 5 – Self-Organizing Map Clustering for Users Acesses, Eight Clusters
SOURCE: Prepared by authors

For each run over the data set, a new training of the map was performed. Thus, different results were obtained for the several runs, ranging from three to twelve clusters. In general, the resulting number of clusters was close to the expected, e.g., 12 clusters were obtained when the number of input vectors *dlen* was 140. Fig. 5 presents the result for eight clusters for purpose of comparison with results obtained for K-means, shown in Fig. 4.

It is interesting to observe that access data was obtained from several courses and make no distinction between tutors and students.

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

It was observed that users with few accesses to the system were from courses less related to computers or technology. Thus, the lack of familiarity with computers would explain the presence of these clusters in the results. Were the data collected from a more uniform sample, this dispersion would probably be smaller. But the presence of such clusters could yet be explained by a lack of interest from the student or even a miscomprehension about how to properly use the tools and resources of the system.

In an ideal scenario, educators could use these results to compare them with the students expected behavior, as well as to find out unexpected features. It would also be possible, with this interaction with educators, to find out which number of clusters would be more appropriated, based on the distinct types of expected behavior. Of course, this would make sense only when data from a single course is analyzed, which was not the case in this study.

To analyze results related to access to resources, the same configurations for the clustering algorithms were used. Besides adjusting the configuration to interpret the different data format, there was no need for additional adjustments.

Two approaches were previously presented to summarize and organize the data collected from access to the environment resources. Both were evaluated and the second option, with simpler structure, was adopted for this data set. The first approach yield to data sets with very high dimensionality with a large quantity of null values inserted to adjust these dimensions, due to few resources with a significant number of accesses, and a large amount of documents that were seldom or never accessed. Consequently, results derived from these sets contained little information, masked under the excessive number of artificially inserted values.

It would be possible, with the collected data, to separate the resources by the courses to which they belong. For that, it would suffice to consider the name of resource that was stored in the system table of events, which contained the full path to the resource. A single text filter would perform the separation by courses, and could also be used to filter out private documents, which were not shared to other users and thus would affect the clustering results, yielding too many resources with few accesses. In this initial study, such option was not adopted, but with larger data sets could be useful.

Fig. 6 show the data clustering obtained for the resource access data using K-means. As for user accesses, the same data was also clustered using self-organizing maps, and the result for this clustering is presented in Fig. 7. In these figures, the graphics relate, for each resource, the number of accesses, the interval between the resource upload and its first access, and an estimated lifetime for the resource, given by the interval between its first and last access. Thus, points close to the origin represent resources with few accesses, that were accessed shortly after publication, but that were not accessed anymore after a short period.

As with the results shown in Fig. 4 and 5, K-means and SOM differ mainly in the definition of clusters where there are most elements, in this case, representing resources. It can be observed that there are very distinct patterns of access to these resources, mostly documents, what can be associated to the difference of relevance attributed by students to each resource. It was also observed that the amount of accesses is related to the resource function – audio files have access patterns distinct from reading materials, which are distinct from answers to questions and from eva-

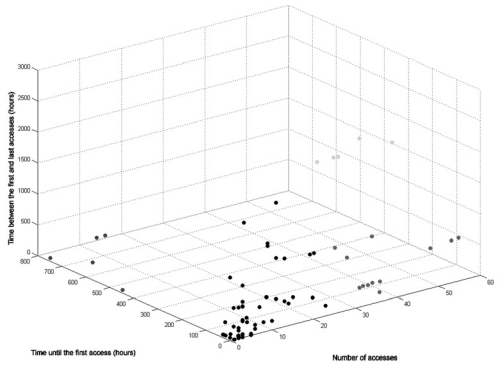


FIGURE 6 – K-means Clustering for Access to Resources
SOURCE: Prepared by authors

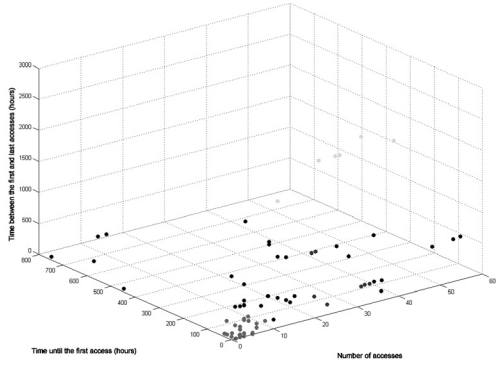
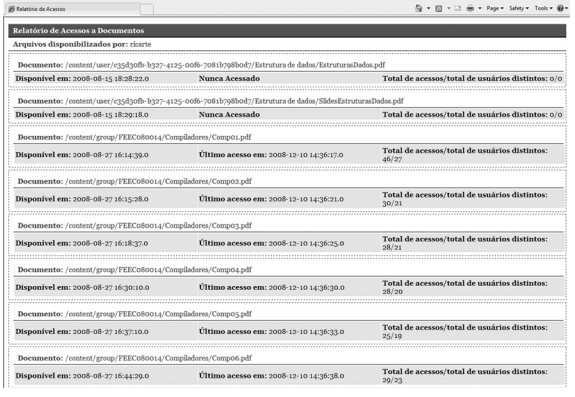


FIGURE 7 – SOM Clustering for Access to Resources
SOURCE: Prepared by authors

luation results. Thus, the result can help the tutor to identify documents that do not have the expected access pattern, locating resources that do not receive the expected attention by the students. In this way, the tutor may investigate potential causes for this neglect.

Textual reports generated from this collected data enable both tutors and students to identify access information for each resource in their courses. Fig. 8 presents a sample of a report generated for the tutor. Entries in this report describe, for each resource, the resource identification (full path within the environment), date when the resource became available, date of the last access, total number of accesses, and the number of distinct users that accessed the resource. The tutor may, with this report, see which students accessed each resource, identifying those that may have difficult to interact with the environment. Another possibility for the tutor is to identify resources with least accesses, which may indicate lack of adequate instructions or lack of visibility for the resource. For the students, their individual reports serve both as a history of access to resources in the environment and as a pointer to call attention to eventual resources that they have not accessed in the system.



Relatório de Acessos a Documentos

Arquivos disponibilizados por: ricarte

Documento: /content/user/c35d30fb-b327-4125-00f6-7081b798b0d7/Estrutura de dados/EstruturasDados.pdf
Disponível em: 2008-08-15 18:28:22.0 | **Nunca Acessado** | **Total de acessos/total de usuários distintos:** 0/0

Documento: /content/user/c35d30fb-b327-4125-00f6-7081b798b0d7/Estrutura de dados/SlidesEstruturasDados.pdf
Disponível em: 2008-08-15 18:29:18.0 | **Nunca Acessado** | **Total de acessos/total de usuários distintos:** 0/0

Documento: /content/group/FEEC08014/Compiladores/Comp01.pdf
Disponível em: 2008-08-27 16:14:39.0 | **Último acesso em:** 2008-12-10 14:36:17.0 | **Total de acessos/total de usuários distintos:** 46/27

Documento: /content/group/FEEC08014/Compiladores/Comp02.pdf
Disponível em: 2008-08-27 16:15:28.0 | **Último acesso em:** 2008-12-10 14:36:21.0 | **Total de acessos/total de usuários distintos:** 30/21

Documento: /content/group/FEEC08014/Compiladores/Comp03.pdf
Disponível em: 2008-08-27 16:18:37.0 | **Último acesso em:** 2008-12-10 14:36:25.0 | **Total de acessos/total de usuários distintos:** 28/21

Documento: /content/group/FEEC08014/Compiladores/Comp04.pdf
Disponível em: 2008-08-27 16:30:10.0 | **Último acesso em:** 2008-12-10 14:36:30.0 | **Total de acessos/total de usuários distintos:** 28/29

Documento: /content/group/FEEC08014/Compiladores/Comp05.pdf
Disponível em: 2008-08-27 16:37:10.0 | **Último acesso em:** 2008-12-10 14:36:33.0 | **Total de acessos/total de usuários distintos:** 25/19

Documento: /content/group/FEEC08014/Compiladores/Comp06.pdf
Disponível em: 2008-08-27 16:44:29.0 | **Último acesso em:** 2008-12-10 14:36:38.0 | **Total de acessos/total de usuários distintos:** 29/23

FIGURE 8 – Textual report of access to resources, tutor view. For each document, the report presents the upload date, date of last access, and the number of accesses, both total and by distinct users
SOURCE: Prepared by authors

These tools, both for visualizing clusters of access (either by K-means or self-organizing maps) and for generating reports, used data from the original implementation of the Sakai databases, and thus can be easily integrated to any learning environment based on

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

the same architecture. It is expected that this tool would be very useful for distance learning courses, in which all interaction occurs through the environment.

5 Conclusion

Computer-based learning environments, mainly when used in distance learning settings, generate a huge amount of data concerning user interactions and accesses to available tools and resources. On the other hand, this data is seldom used provide feedback to the users of these virtual learning environments. This yields a feeling of not knowing what is happening with the effort of producing and making these resources available to students.

This paper presented a methodology to extract data about resource usage directly from learning environment databases and how to prepare them for data mining processes. A case study using a learning platform based on Sakai, an internationally developed and adopted open architecture for learning systems, was presented, with results presenting clusters of access reports, both for user accesses and for access to resources or documents in the environment. This study, developed with data extracted from actual courses, demonstrated the viability of performing useful data mining from the environment internal databases.

The characteristics of the sample used for the case study, based on the environment usage for supporting regular courses and not for a distance learning setting, may affect the interpretation of the results, since on-campus students may have access to the same materials by other means rather than the environment – going to a library or borrowing from a colleague, for example. Thus, in such settings, a student not accessing the system is not necessarily an absent student. This would not be the case in a distance learning course, in which all interaction should happen through the environment and be registered in the environment databases.

In distance learning courses, which have in general a large number of students and tutors involved in an offer, the presented methodology would be an essential tool to provide feedback for both tutors and students. As presented in the paper, is quite simple to isolate the information by course in an environment supporting several simultaneous offers.

It was also observed the important role that educators have in the process of tuning the data mining process. In the definition of the initial parameters for supervised clustering algorithms, as K-means, educators could provide an estimate of initial number of clusters corresponding to expected patterns of behavior to access the environment. On the other hand, unsupervised clustering algorithms, as self-organizing maps, may help educators to identify such patterns and validate their theoretical models.

Future works could try to correlate behavior patterns with outcomes achieved by learners in their courses. Results from such studies could provide feedback both for students, as a set of guidelines to achieve better results in distance learning courses, and for tutors, which could receive suggestions on the best way to organize content to reach their students.

Acknowledgments

The TIDIA-Ae learning environment was developed as a collaborative project among several institutions in the State of Sao Paulo, Brazil, and supported by FAPESP, the state agency for research support. Particularly, the version used in this case study was developed by the UNICAMP e-learning Laboratory (e-labora), supported by FAPESP Grant Number 05/60572-1. The authors wish to thank the collaboration of the e-labora members and from the UNICAMP Computing Center analysts, in which the environment was deployed and used to support UNICAMP courses. Mr. Falci-Junior received a grant from the Brazilian Internal Revenue Service (Receita Federal), during his participation in the Harpia Project.

References

ADAPTIVE INFORMATICS RESEARCH CENTER. *SOM Toolbox 2.0*. [S.l.], 2005. Available at: <http://www.cis.hut.fi/projects/somtoolbox/>

CASTRO, F. *et al*. Applying data mining techniques to e-learning problems. In: EVOLUTION of Teaching and Learning Paradigms in Intelligent Environment. [S.l.]: Springer, 2007. P. 183-221.

FAPESP. *TIDIA-Ae portal*. São Paulo, 2010. Available at: <http://tidia-ae.iv.org.br/>

HSU, H.H.; CHEN, C.J.; TAI, W.P. Towards error-free and personalized Web-based courses. In: INTERNATIONAL CONFERENCE ON ADVANCED INFORMATION NETWORKING AND APPLICATIONS, 17., 2003, Xi'an, China. *Proceedings*. [S.l.: s.n.], 2003. P. 99.

ROMERO, C.; VENTURA, S. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, New York, v. 33, n. 1, p. 135-146, 2007.

ROMERO, C.; VENTURA, S.; GARCIA, E. Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, New York, v. 51, n. 1, p. 368-384, 2008.

SAKAI. *Sakai Project*: An open source suite of learning, portfolio, library and project tools. [S.l.], 2010. Available at: <http://sakaiproject.org/>

SANTOS, L. dos; BECKER, K. Distance education: a Web usage mining case study for the evaluation of learning sites. In: INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, 3., 2003, Athens, Greece. *Proceedings*. Los Alamitos, CA: IEEE, 2003. P. 360-361.

SHEARD, J. *et al*. Inferring student learning behaviour from website interactions: A usage analysis. *Education and Information Technologies*, New York, NY: Springer, v. 8, p. 245-266, 2003.

UNIVESP. *Virtual University of the State of São Paulo*: learning platform. São Paulo, 2010. Original in Portuguese. Available at: <http://www.univesp.ensinosuperior.sp.gov.br/9/plataforma-de-aprendizagem>

Porto Alegre, v.14, n.2, jul./dez. 2011.
ISSN impresso 1516-084X ISSN digital 1982-1654

ZAIANE, O.R.; LUO, J. Towards evaluating learners' behaviour in a Web-based distance learning environment. In: INTERNATIONAL CONFERENCE ON ADVANCED LEARNING TECHNOLOGIES, 1., 2001, Madison, USA. *Proceedings*. Los Alamitos, CA: IEEE, 2001. P. 357-360.

Recebido em 29 de maio de 2010.
Aprovado para publicação em 02 de outubro de 2010.

Ivan Luiz Marques Ricarte
Professor Associado junto ao Departamento de Engenharia de Computação e Automação Industrial da Universidade Estadual de Campinas, UNICAMP, Campinas/SP – Brasil. E-mail: ricarte@fee.unicamp.br

Geraldo Ramos Falci Junior
Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica da Universidade Estadual de Campinas, UNICAMP, Campinas/SP – Brasil. E-mail: geraldofalci@gmail.com