Collaborative Data Mining Tool for Education

Enrique García[1], Cristobal Romero[1], Sebastian Ventura[1], Miguel Gea[2], Carlos de Castro[1]

{egsalcines, cromero, sventura, cdecastro}@uco.es, mgea@ugr.es
[1]Department of Computer Science, University of Cordoba
[2]Department of Computer Languages and Systems, University of Granada

Abstract. This paper describes a collaborative educational data mining tool based on association rule mining for the continuous improvement of e-learning courses allowing teachers with similar course's profile sharing and scoring the discovered information. This mining tool is oriented to be used by instructors non experts in data mining such that, its internal operation is transparent to the user and the instructor can be focused in to the analysis of the results and make decisions about how to improve e-learning courses.

Introduction

Nowadays, there are a variety of general data mining tools and frameworks. Some examples of commercial mining tools are DBMiner [1], SPSS Clementine [2], DB2 Intelligent Miner [3], etc. And some examples of public domain mining tools are Weka [4], RapidMiner [5], Keel [6], etc. All these tools are not specifically designed for pedagogical/educational purposes and it is cumbersome for an educator to use these tools which are normally designed more for power and flexibility than for simplicity. However, there are also an increasing number of mining tools specifically oriented to educational data such as: Mining tool [7] for association and pattern mining, MultiStar [8] for association and classification, EPRules [9] for association, KAON [10] for clustering and text mining, Synergo/ColAT [11] for statistics and visualization, GISMO [12] for visualization, Listen tool [13] for visualization and browsing, TADA-Ed [14] for visualizing and mining, O3R [15] for sequential pattern mining, Sequential Mining tool [16] for pattern mining, MINEL [17] for mining learning paths, Simulog [18] for looking for unexpected behavioral pattern. Moodle mining tool [19] for classification, clustering and association rule mining. All these tools are oriented to be used by a single instructor or course administrator in order to discover useful knowledge from their own courses. So, they don't allow a collaborative usage in order to share all the discovered information between other instructors of similar courses (contents, subjects, educational type: elementary and primary education, adult education, higher, tertiary and academic education, special education, etc.). In this way, the information discovered locally by teachers could be joined and stored in a common repository of knowledge available for all instructors for solving similar detected problems.

In this paper, we describe an educational data mining tool based on association rule mining and collaborative filtering for the continuous improvement of e-learning courses and it directed to teachers non experts in data mining. The main objective is to make a mining tool in which the information discovered can be shared and scored between different instructors and experts in education.

Implementation of the collaborative data mining tool

We have developed a data mining tool with two subsystems: client and server application (Figure 1). The client application uses an association rule mining tool for discovering interesting relationships through student's usage data in the form of IF-THEN recommendation rules. The server application uses a collaborative recommender system to share and score the previously obtained rules by instructors of similar courses with other instructors and experts in education.

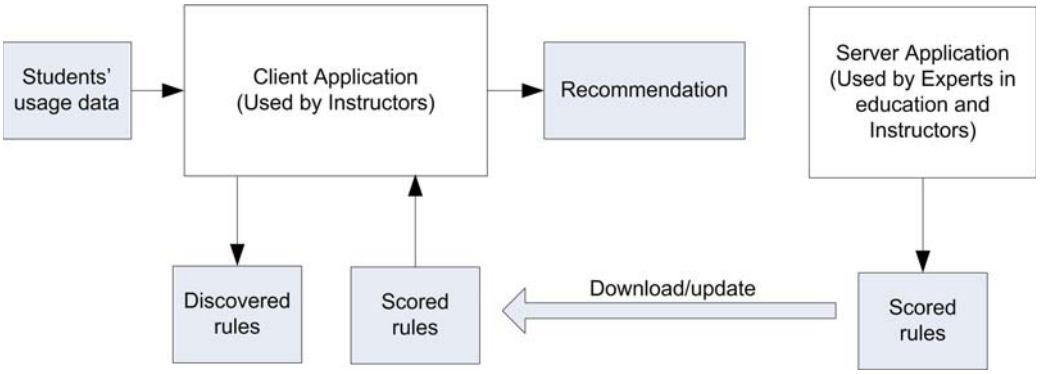

Figure 1. Collaborative data mining tool

As we can see in Figure 1, the system is based on client-server architecture with N clients, which applies an association rule mining algorithm locally on students' usage data. In fact, the client application uses the Predictive Apriori algorithm [20], because it does not require the user to specify parameters such as the minimum support threshold or confidence values. The only parameter is the number of rules to be discovered, which is a more intuitive parameter for a teacher non expert in data mining. The association rules discovered by the client application must be evaluated to decide if they are relevant or not, therefore the client application uses an evaluation measure [21] to classify the rules as being expected or unexpected, comparing them with the scored rules stored in a collaborative rules repository maintained on server side. Also, the expected rules found are then expressed in a more comprehensible format of recommendation about possible solutions to problems detected in the course. The teacher sees the recommendation and can determine if it is relevant or not for him/her in order to apply/use the recommendation. On the other side, the server application allows managing the rules repository using collaborative filtering techniques with knowledge-based techniques [21]. The information in the knowledge base is stored in form of tuples (rule-problem-recommendation-relevance) which are classified according to a specific course profile. The course profile is represented as a three-dimensional vector related with the following characteristic of his/her course: Topic (the area of knowledge, e.g. Computer Science or Biology); Level (level of the course, e.g. Universitary, High School, Elementary or Special Education); and Difficulty (the difficulty of the course, e.g., Low or High). These similarities between courses are available to other teachers to assess in terms of applicability and relevance. A group of experts in online education from University of Córdoba, Spain, propose the first tuples of the rule repository and also vote on those tuples proposed by other experts. On the other hand, teachers could discover new tuples

(in the client application) but these must be validated by the experts (in the sever application) before being inserted in the rule repository.

1.1 Client application

As we mentioned before, the client application is used by instructors in order to find association rules. The main feature of the client application is its specialization in educational environments. Before applying our mining algorithm, the data have to be pre-processed in order to adapt them to our specific data model. First, the teacher has to select the origin of the data to be mined. We have two different formats available for input data: 1) the Moodle relational database, for teachers that work with Moodle as well as the INDESAHC authoring tool [22], so all our attributes are used directly; or 2) a Weka [4] ARFF text file, for teachers that use other LMSs and, therefore, other attributes. Also, the teacher can restrict the search field, we have also added a few parameters related with the analysis depth. Firstly, the teacher must select the level of granularity to carry out the analysis: course, unit, lesson or a specific table of the data base such as course-unit, course-lesson, course-exercise, course-forum, unit-exercise, unit-lesson, lesson-exercise among others.

The rules repository (see Figure 2) is the knowledge database upon which the analysis of the discovered rules is based. Before running the algorithm, the teacher downloads from the server, the current knowledge database, according to his/her course profile.

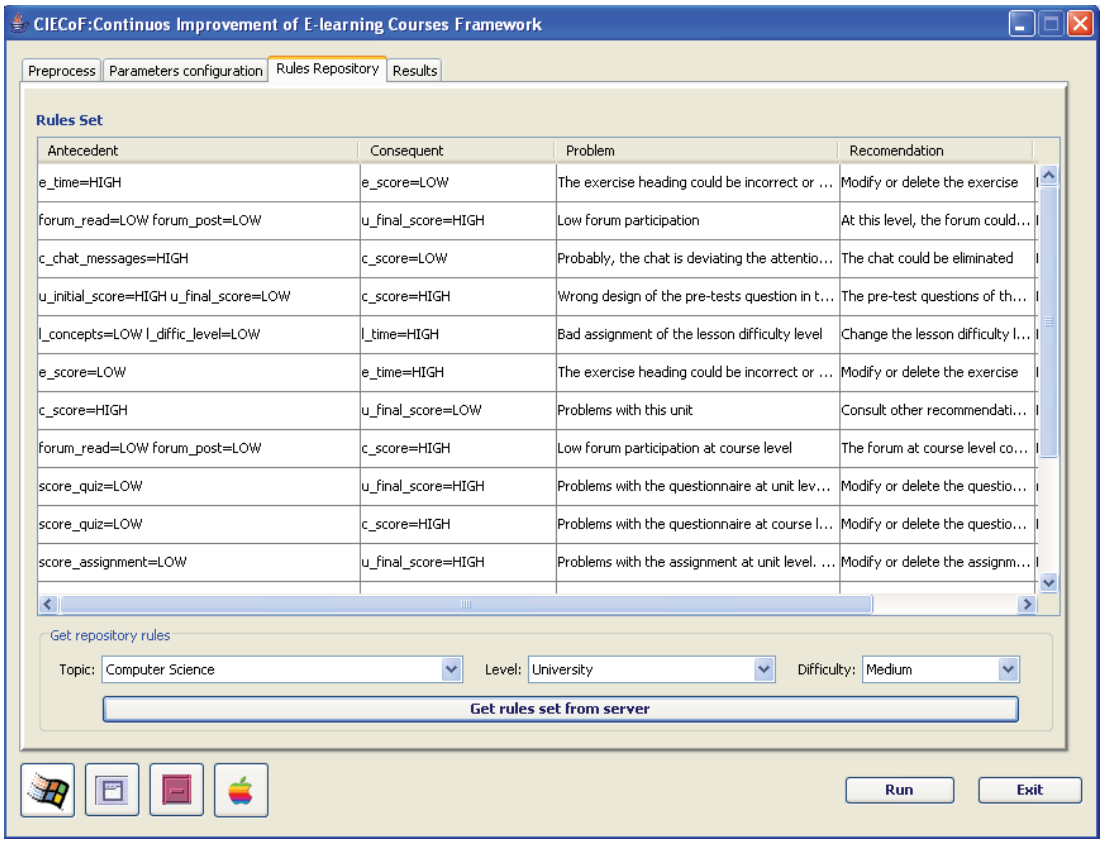

Figure 2. Rules repository panel

Finally, after downloading the rule repository and configuring the application parameters or using default values, the teacher executes the association rule algorithm. Then, client application shows the results obtained in a table (see Figure 3), with the following fields: rule (discovered IF-THEN rule), problem (detected by the rule), recommendation (about how to solve the problem), score (of experts and others instructors have set to the rule) and apply button (to use/apply the recommendation in his/her course).

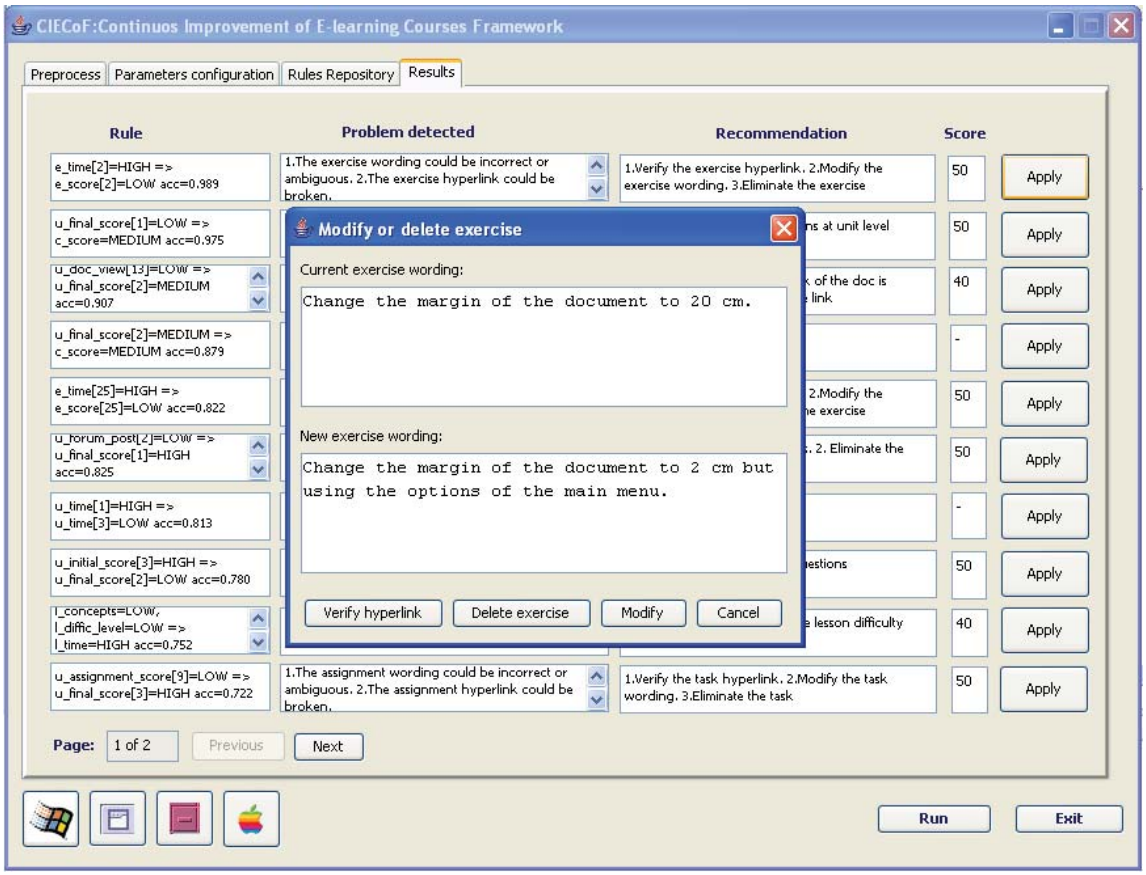

Figure 3. Results panel

We have distinguished between two types of recommendations: 1) Active, if it implies a direct modification of the course content or structure; or 2) Passive, if it detects a more general problem in the course or unit and it advices the teacher to consult more specific recommendations related with these didactic resources. Active recommendations can be linked to: modifications in the formulation of the questions (see Figure 3) or the practical exercises/tasks assigned to the students; changes in previously assigned parameters such as course duration or the level of lesson difficulty; or the elimination of a resource such as a forum or a chat room.

1.2 Server application

The server application is used by experts and instructors. The experts in education insert the tuples and they explicitly vote for them by indicating degrees of preference (see Figure 4). The teachers vote implicitly when they push the "Apply" button, in order to side-step one of the main problems for collaborative filtering systems, that is how to encourage teachers to vote or evaluate. In this case, if teachers apply one of the recommendations to their course, they are implicitly voting for this specific tuple.

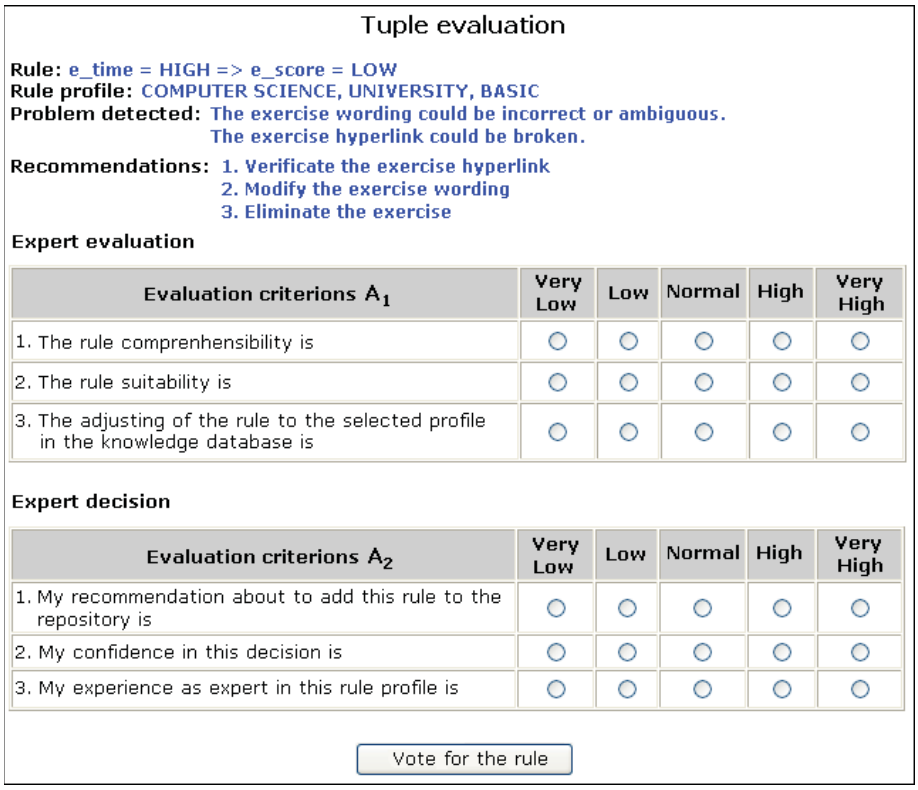

Figure 4. Vote panel

The server application is a web-based application for managing the knowledge database or tuple repository (see Figure 5). In order to access easily to all the editing options for the repository, a general course profile was created which is the profile used by the experts in educational domain. These experts have permission to introduce new tuples into the rule repository and vote explicitly for existing ones (see Figure 4). In order to allow information exchange (tuples) between client and server, we have developed a web service for downloading/uploading the repository. Each time a client application updates its repository, all the tuples are reordered in the repository.

Finally, we must mention that an evaluation of this collaborative data mining tool can be found in [21].

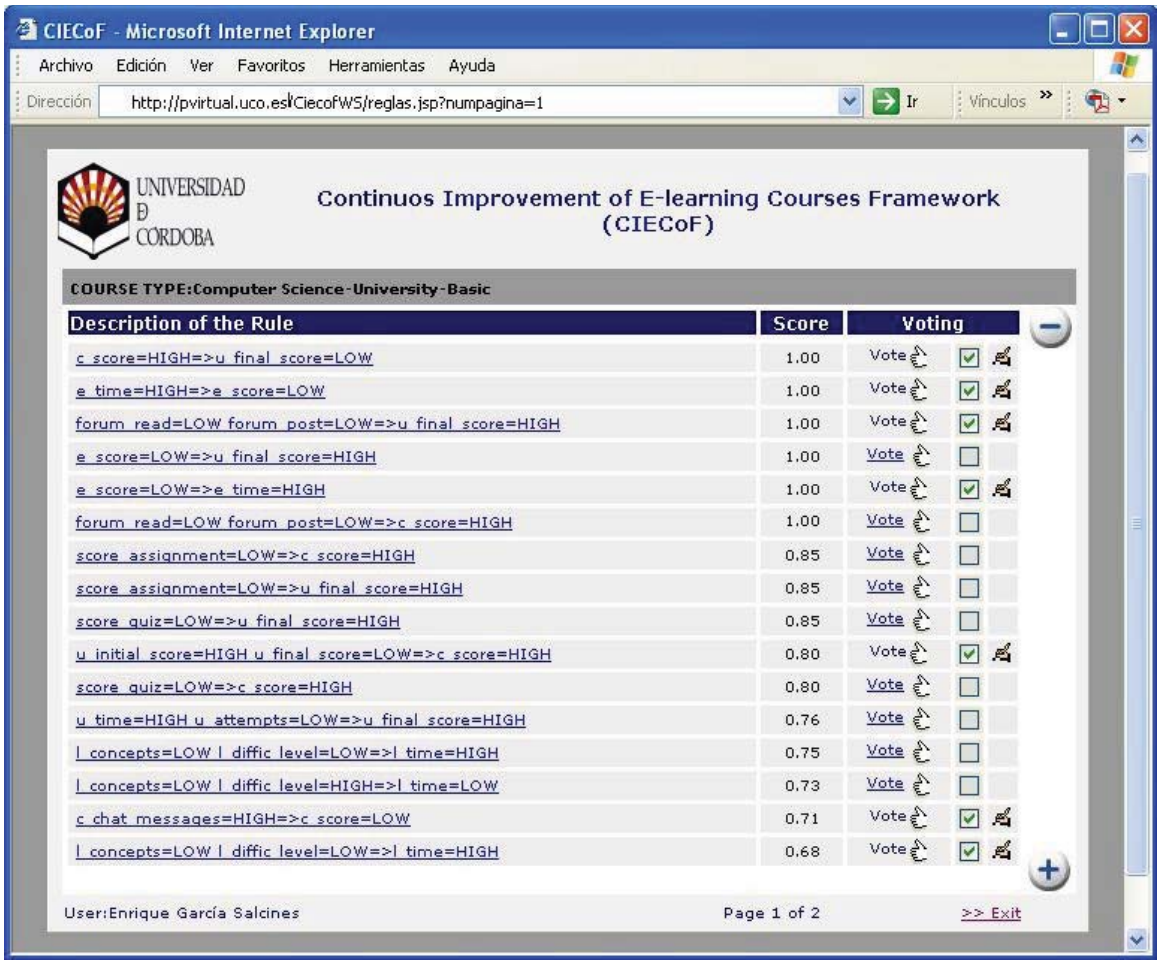

Figure 5. Server application interface

2 Conclusions

In this paper we have shown a data mining tool that uses association rule mining and collaborative filtering in order to make recommendation to instructors about how to improve e-learning courses. This tool enables to share and score the discovered rules by other teachers of similar courses. Currently, the mining tool has been only used by a group of instructors and expert involved in the development of the own tool. So, in the future we want to test the tool with several groups of external instructors and experts in order to can test the usability of the tool with external users.

Acknowledgments

The authors gratefully acknowledge the financial support provided by the Spanish department of Research under TIN2008-06681-C06-03 and P08-TIC-3720 Projects.

References

[1] DBMiner. Available at http://www.dbminer.com. 2009.

[2] Clementine. Available at http://www.spss.com/clementine/. 2009.

[3] Miner. Available at http://www-306.ibm.com/software/data/iminer/. 2009

[4] Weka. Available at http://www.cs.waikato.ac.nz/ml/weka/. 2009

[5] RapidMiner. Available at http://rapid-i.com/. 2009.

[6] Keel. Available at http://www.keel.es/. 2009.

[7] Zaïane, O., Luo, J. Web usage mining for a better web-based learning environment. *In Proceedings of conference on advanced technology for education*, 2001. Banff, Alberta , pp. 60–64.

[8] Silva, D., Vieira, M., 2002. Using data warehouse and data mining resources for ongoing assessment in distance learning. *In IEEE International Conference on Advanced Learning Technologies*, 2002. Kazan, Russia, pp. 40–45.

[9] Romero, C., Ventura, S., and Bra, P. D. Knowledge discovery with genetic programming for providing feedback to courseware author. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research*, 2004, 14(5), pp. 425–464.

[10] Tane, J., Schmitz, C., Stumme, G. Semantic resource management for the web: An elearning application. *In Proceedings of the WWW Conference*, 2004 New York, USA, pp. 1–10.

[11] Avouris, N., Komis, V., Fiotakis, G., Margaritis, M., Voyiatzaki, E. Why logging of fingertip actions is not enough for analysis of learning activities. *In Workshop on Usage analysis in learning systems at the 12th International Conference on Artificial Intelligence in Education*, 2005. Amsterdam, Netherland, pp. 1-8.

[12] Mazza, R., Milani, C. Exploring usage analysis in learning systems: Gaining insights from visualisations. *In Workshop on Usage analysis in learning systems at 12th International Conference on Artificial Intelligence in Education*, 2005. New York, USA, pp. 1-6.

[13] Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., Heiner, C. An educational data mining tool to browse tutor-student interactions: Time will tell! In *Proceedings of the Workshop on Educational Data Mining*, 2005. Pittsburgh, USA, pp. 15–22.

[14] Merceron, A., Yacef, K. TADA-Ed for educational data mining. *Interactive multimedia electronic journal of computer-enhanced learning*, 2005, 7(1), pp. 267-287.

[15]Becker, K., Vanzin, M., Ruiz, D. A. Ontology-based filtering mechanisms for web usage patterns retrieval. *In International Conference on E-Commerce and Web Technologies*, 2005. München, Germany, pp. 267–277.

[16]Romero, C., Porras, A.R., Ventura, S., Hervás, C., Zafra, A. Using sequential pattern mining for links recommendation in adaptive hipermedia educational systems. *In International Conference Current Developments in Technology-Assisted Education*, 2006. Sevilla, Spain, pp.1016-1020.

[17]Bellaachia, A., Vommina, E., Berrada, B. Minel: a framework for mining e-learning logs. *In Proceedings of the 5th IASTED international conference on Web-based education*, 2006. Mexico, pp. 259-263.

[18]Bravo, J., Ortigosa, A. Validating the Evaluation of Adaptive Systems by User Profile Simulation. *In Proc. Workshop on User-Centred Design and Evaluation of Adaptive Systems*, 2006, Dublin, Germany, pp. 52-56.

[19]Data mining algorithms to classify students. C. Romero, S. Ventura, P. Espejo, C. Hervas. *Educational Data Mining Conference EDM 2008*, 2008, 20-21 Junio, Montreal, pp. 8-17.

[20]Scheffer, T. Finding Association Rules That Trade Support Optimally against Confidence. *Intelligent Data Analysis*, 2005, 9(4), pp. 381-395.

[21]García, E., Romero, C., Ventura, S., de Castro, C. An architecture for making recommendations to courseware authors through association rule mining and collaborative filtering. *UMUAI: User Modelling and User Adapted Interaction*, 2009, 19(1-2), pp. 99-132.

[22]De Castro, C., García, E., Romero, C., and Ventura, S. Herramienta autor INDESAHC para la creación de cursos hipermedia adaptativos. *Revista latinoamericana de tecnología educativa* 3, 2004, pp. 349-367.