

# Desarrollo de un almacén de datos para el desempeño académico de estudiantes en cursos b-learning

Sergio Franco Casillas, Sara Catalina Hernández Gallardo

Universidad de Guadalajara, Jalisco,  
México

scasillas@cualtos.udg.mx, shernand@cencar.udg.mx

**Resumen.** El propósito de este artículo es presentar un almacén de datos que permita guardar los elementos provenientes del resultado del desempeño académico de los estudiantes en cursos *b-learning*, con el uso de la plataforma de cursos en línea *Moodle*. La aplicación informática, *Moovisys*, se implementa como interfaz de usuario para la preparación y extracción de los datos. *Moovisys*, aún en desarrollo, se conecta a la base de datos de *Moodle* para conocer el uso de la plataforma, a su vez, se identifican los cursos *b-learning*, y se extrae de ellos los datos correspondientes al desempeño académico en las actividades de aprendizaje, para luego colocarlos en el almacén. El análisis de datos se realizó con *Tableau* para conocer el desempeño académico de los estudiantes en ingeniería en computación. En los resultados se muestra el desempeño académico de los estudiantes, el tiempo anticipado del envío de tareas y el que usaron para realizar un examen.

**Palabras clave:** Almacén de datos, cursos b-learning, actividades de aprendizaje, minería de datos.

## 1. Introducción

En la actualidad las Tecnologías de Información y Comunicación (TIC) desempeñan un papel importante en el ritmo de vida de los seres humanos, entre los que destacan el uso masivo del internet y dispositivos móviles, estos generan una gran cantidad de datos que se almacenan en diversas partes del mundo [1]. Las instituciones educativas utilizan los sistemas gestores de aprendizaje (LMS, por sus siglas en inglés) como medios de interacción; estas plataformas permiten almacenar en su base de datos cada una de las acciones que los profesores y estudiantes realizan entre las actividades de aprendizaje (AA) y el sistema. Los LMS se utilizan como medio de apoyo para la educación *e-learning* (en línea) como para la *b-learning* (semi-presencial) [3]. En un ambiente *b-learning* una parte de los recursos de aprendizaje y la interactividad entre estudiantes de un aula de clases se llevan a una plataforma LMS para que participen los diversos actores [4]. En esta investigación se estudian los cursos *b-learning* porque el centro universitario participante en el estudio (CUP) utiliza la plataforma Moodle como complemento a la actividad presencial de los profesores, sin embargo puede ser adaptada a cursos *e-learning*.

Existen disciplinas que se preocupan por el análisis de datos, tal es el caso de la minería de datos, siendo esta un área que se fortalece de la estadística, las tecnologías de bases de datos, máquina de aprendizaje (machine learning), reconocimiento de patrones, recuperación de información, visualización, algoritmos computacionales, cómputo de alto rendimiento y desarrollo de aplicaciones [5]. Con el almacenamiento de datos provenientes de la educación y el uso de técnicas de análisis se desarrolla un área de investigación en la educación (EDM, por sus siglas en inglés), la cual se preocupa por el desarrollo de métodos de exploración que permite generar conocimiento a partir de los datos provenientes de la educación [2].

La minería de datos (MD) tiene como meta buscar patrones que son previamente desconocidos, válidos, novedosos que se analizan en grandes volúmenes de datos [6, 7, 8]. La MD está implícita en el proceso de extracción de conocimiento en bases de datos (KDD, siglas en inglés) en ocasiones se trata como sinónimo [6]. El KDD es un proceso de varias etapas bien establecidas, que de acuerdo con [6, 9] se utilizan en: 1) Educación, 2) Industria y 3) Modelos híbridos. Estos modelos varían en los diferentes pasos a seguir para llegar al conocimiento. En los modelos propuestos figuran esencialmente seis pasos 1) Entendimiento del problema 2) Entendimiento de los datos; 3) Preparación de los datos; 4) Minería de datos; 5) Evaluación del descubrimiento de conocimiento; 6) Utilización del conocimiento descubierto [6], [7], [10, 11, 12].

La MD cuenta con técnicas y algoritmos para la descripción y predicción del fenómeno que se estudia [7], en esta investigación se utiliza la técnica de visualización que permite mostrar y describir de manera gráfica los datos crudos provenientes del desempeño académico de los estudiantes en un ambiente en línea.

El desempeño académico (DA) es una variable importante de abordar en la educación ya que se asocia a otros problemas como la reprobación, deserción o fracaso escolar y que preocupa a los gobiernos como a las instancias académicas [13]. El DA es por excelencia un indicador que mide la calidad de la educación y que a su vez se expresa por el conjunto de calificaciones que se obtienen de las evaluaciones y de las AA donde el profesor designa una calificación [14, 15]. Para la elaboración de esta investigación utiliza como DA el conjunto de calificaciones que el estudiante obtiene en un curso *b-learning* en Moodle. Existen trabajos relacionados como el elaborado por [16] que es una aplicación de visualización y monitoreo de estudiantes, GISMO (*Graphical Interactive Student Monitoring Tool for Moodle* - <http://gismo.sourceforge.net>), que se agrega internamente a Moodle y se trata como un bloque dentro de él; el administrador de la plataforma agrega a GISMO y solo puede observar los datos de cada curso de manera separada, en cambio Moovisys es independiente de Moodle y puede ayudar a la toma de decisiones de los responsables académicos o administradores de la plataforma para mejorar las practicas docentes y de evaluación dentro de este ambiente.

De acuerdo con [2, 6] la extracción y preparación de los datos es una tarea complicada ya que consume la mayor parte en el proceso de KDD y porque se integran datos de diversas fuentes, y sin ella no se podrá avanzar a la siguiente fase: minería de datos; lo anterior comparándolo con el pre-procesamiento manual que permite Moodle en su interfaz en la extracción de calificaciones de cada curso. En este artículo se presenta Moovisys (*Moodle Visualization System*) que es una aplicación informática, en desarrollo, que facilita la conexión con la base de datos de Moodle de manera externa. Moovisys posibilita, a través de su interfaz, la recolección de datos

provenientes del DA de los estudiantes de cursos *b-learning*, transformarlos y almacenarlos en una nueva base de datos relacional (almacén de datos). Con un almacén de datos se podrá dar seguimiento sobre el DA de los estudiantes tanto en un semestre como a lo largo de su carrera. Para el análisis de datos se utilizó *Tableau* (<http://www.tableau.com>) que permite la conexión al almacén y utiliza la técnica de visualización para mostrar de manera gráfica los datos. Al final del artículo se contestan las siguientes preguntas de investigación, primer paso del KDD: a) ¿Qué actividades de aprendizaje utilizan los profesores?; b) ¿Qué actividades de aprendizaje tienen un mejor desempeño académico los estudiantes?; c) ¿Cuánto tiempo le lleva al estudiante realizar un examen?; d) ¿Con qué anticipación envía una tarea los estudiantes?

## 2. Materiales y Métodos

El centro universitario participante (CUP), de la Universidad de Guadalajara, proporcionó una copia de la base de datos de Moodle (versión 1.9+) en un archivo .sql con 2.4 Gb de información, con un total de 201 tablas, y utiliza MySQL como sistema gestor. El periodo de la base de datos es de septiembre de 2007 a junio de 2013. La base de datos se colocó en un servidor con tecnología Apache, PHP y MySQL que permitieron conectar y desarrollar a Moovisys con la base de datos de Moodle. El tratamiento de los datos se realizó con la categoría correspondiente a la carrera de ingeniería en computación del CUP.

### 2.1. Selección de los datos

Para extraer la información de los participante en los cursos de Moodle, se analizaron cada una de las 201 tablas y se seleccionaron aquellas que tenían datos de interés sobre los cursos, el DA de los estudiantes y la participación (Tabla 1).

**Tabla 1.** Nombre de las tablas y la información de interés.

Nombre de tabla	Información de interés relacionada a:
<i>mdl_user</i>	Los participantes
<i>mdl_course</i>	Los cursos
<i>mdl_course_category</i>	Categorías o carreras a la que pertenece el curso
<i>mdl_role</i>	Rol de cada participante
<i>mdl_role_assignment</i>	Rol del participante en el curso
<i>mdl_context</i>	Contexto del curso y los participantes
<i>mdl_assignment</i>	Cada tarea que se proporciona en el curso
<i>mdl_assignment_submission</i>	Participación en cada tarea
<i>mdl_quiz</i>	Cada examen que se proporciona en el curso
<i>mdl_quiz_attempt</i>	Participación en cada examen

Para la extracción en particular de cada una de las AA y su respectiva participación, se utilizó una combinación de las tablas (Tabla 1) para crear el almacén de datos, por ejemplo, para los exámenes se utilizaron las tablas *mdl\_quiz* y *mdl\_quiz\_attempt*.

Para que un curso sea catalogado como *b-learning* se consideran las siguientes reglas: 1) Tener más de un participante (el profesor cuenta como participante y se asigna cuando se crea el curso); 2) tener al menos un recurso de aprendizaje (como lecturas, páginas web, etc.); 3) tener más de un foro de discusión (Moodle coloca uno por default); 4) tener más de una AA (tanto en tareas como en exámenes).

## 2.2. Desarrollo del almacén de datos (AD)

Una vez que se conoce en qué tablas de la base de datos de Moodle se almacena la información de interés, se procede a realizar un esquema de bases de datos relacional para el AD. Con él se evita redundancia de datos en: categorías, cursos, actividades y tipos de actividad, además de, participación en las AA que se almacenan en la tabla de hechos (*eparticipacion*), como se muestra en la Fig. 1.

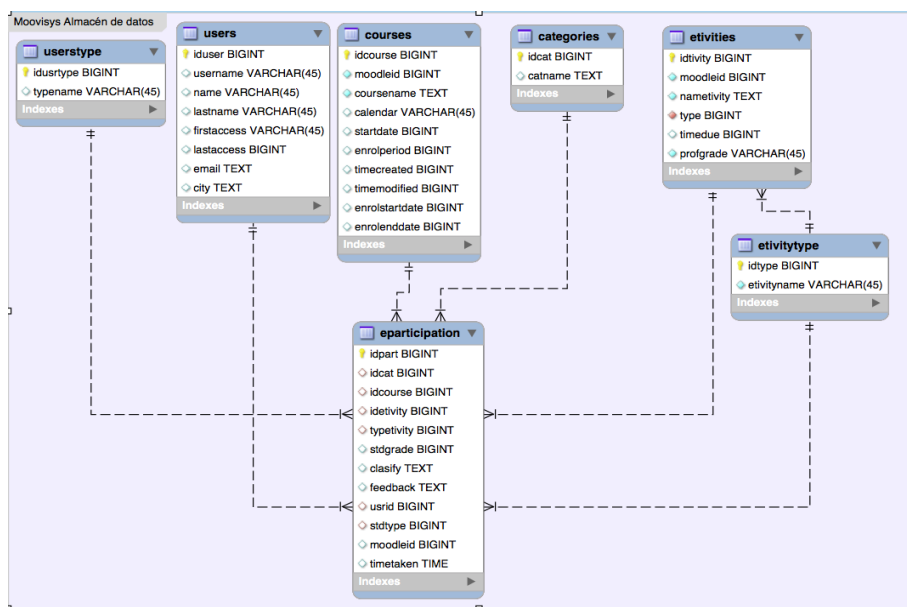


Fig. 1. Almacén de datos de Moovisys.

El AD tiene una configuración de copo de nieve (*snowflake*, en inglés) de acuerdo con [5]. En la tabla 2 se muestra los datos de interés que guarda cada tabla del almacén.

Tabla 2. Tablas utilizadas de Moovisys y los datos que almacenan.

Nombre de tabla	Información de interés relacionada con:
<i>users</i>	Los participantes
<i>userstype</i>	Sí son estudiantes o profesores
<i>Courses</i>	Los cursos
<i>Categories</i>	Categorías o carreras dentro de Moodle
<i>etivities</i>	Actividades de aprendizaje en línea

Nombre de tabla	Información de interés relacionada con:
<i>ativitytype</i>	Los tipos de actividad (tareas, exámenes, etc).
<i>epartitipation</i>	La tabla de hechos o de participación de los estudiantes.

Una vez que se diseñó el modelo del AD para Moovisys con MySQL Workbench (<http://dev.mysql.com/downloads/workbench>) se procedió a sincronizarlo con el sistema gestor, es decir MySQL.

### 2.3. Transformación de los datos

Los datos correspondientes a las calificaciones de cada actividad de aprendizaje se transformaron porque el profesor designa para cada una de ellas una puntuación diferente, por lo tanto no pueden tratarse como iguales. Para esta investigación los datos se almacenaron en la tabla de hechos de una forma normalizada y entendible, estos se convirtieron a escala de 100, donde se utilizó la siguiente fórmula 1:

$$C_R = \frac{C_O \cdot 100}{C_D} \quad (1)$$

Donde  $C_R$  = calificación resultante;  $C_O$  = calificación obtenida por el estudiante;  $C_D$  = calificación designada a la actividad de aprendizaje.  $C_R$  se almacena en las calificaciones de la tabla de hecho.

### 2.4. Diagrama de flujo para almacenar los datos

Moovisys añade la programación como se observa en el diagrama de flujo correspondiente a la Fig. 2, en él se integran los pasos para desarrollar el algoritmo que permite consultar, extraer y guardar en el AD lo correspondiente a las categorías, cursos, actividades, tipo de actividad, usuarios, tipo de usuario y la participación en cada actividad.

## 3. Resultados

Para el tratamiento de los datos, del CUP, se utilizó Moovisys y conectó Tableau al AD para obtener los resultados expresados a continuación:

- 3.1 Tipos de actividades: Moovisys, como parte de la visualización de datos, proporciona un módulo donde se observan las actividades utilizadas en cada curso. En la Fig. 3 se muestran 17 cursos (eje Y), y la cantidad de los tipos de actividad en el eje X. Las AA utilizadas con mayor frecuencia son tareas y exámenes.
- 3.2 Cursos *b-learning*: de los 17 cursos que se muestran en la Fig. 3 (eje Y), solo uno de ellos cumplía con las reglas del apartado 2.1, Fig. 4.
- 3.3 Calificaciones: la Fig. 5 muestra las calificaciones obtenidas con el tratamiento de los datos en *Tableau*. En ella se aprecia que los exámenes tienen bajo DA comparados con las tareas, además de tener mayor participación ya que en las

tareas, el profesor restringe el tiempo de entrega; las calificaciones se muestran en rectángulos color azul, entre más intenso, mejor DA.

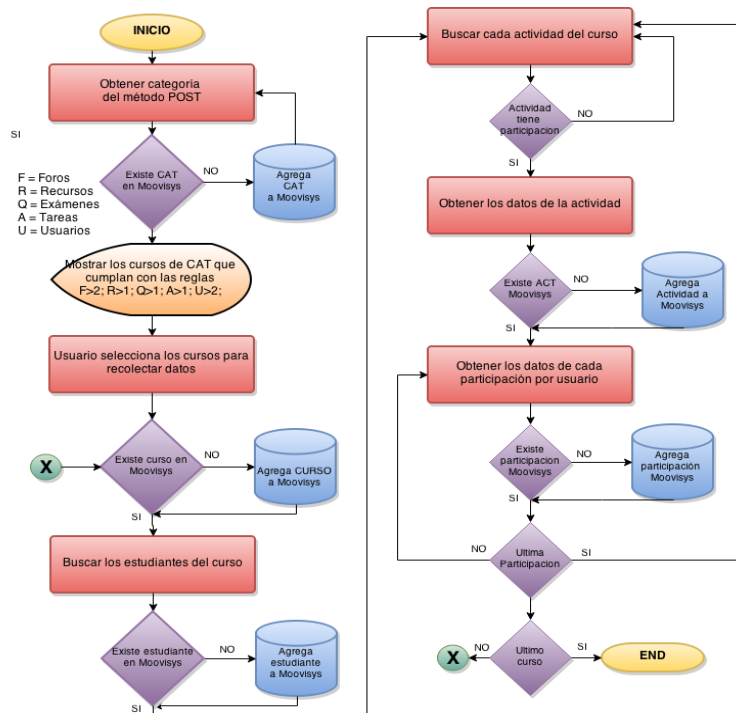


Fig. 2. Diagrama de flujo para extracción de datos.

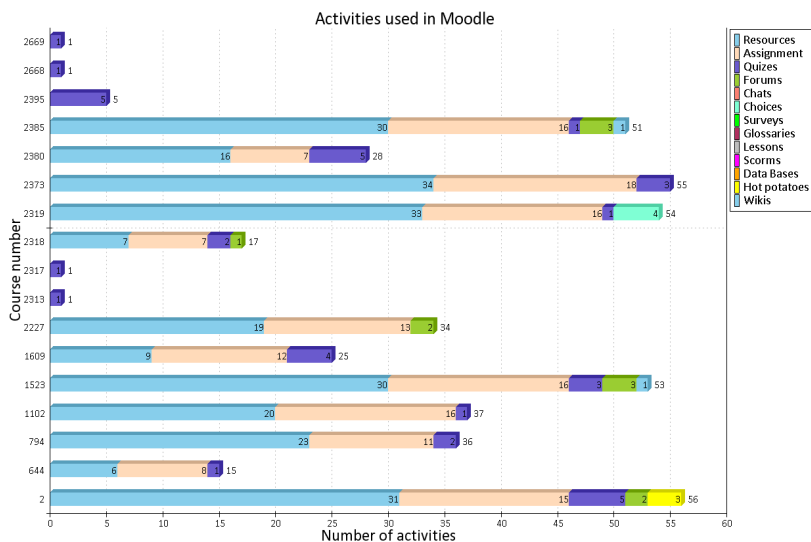


Fig. 3. Tipos de actividad utilizadas en los cursos. Tomada de Moovisys.

### Show B-Learning Courses

Course id	Short course name	Create year	Last modification	Number of Sections	Number resources used	Assignments	Quiz	Forum	Chat	Choice	Survey	Glossary	Lessons	Scorm	Data bases	Hot potato	Wikis	Users	Data collected
2	Lenguajes	2007	2013	7	31	15	5	2	0	0	0	0	0	0	0	3	0	36	
Total courses found			1	There is not course data to be gathered															

Fig. 4. Cursos b-learning. Tomada de Moovisys.

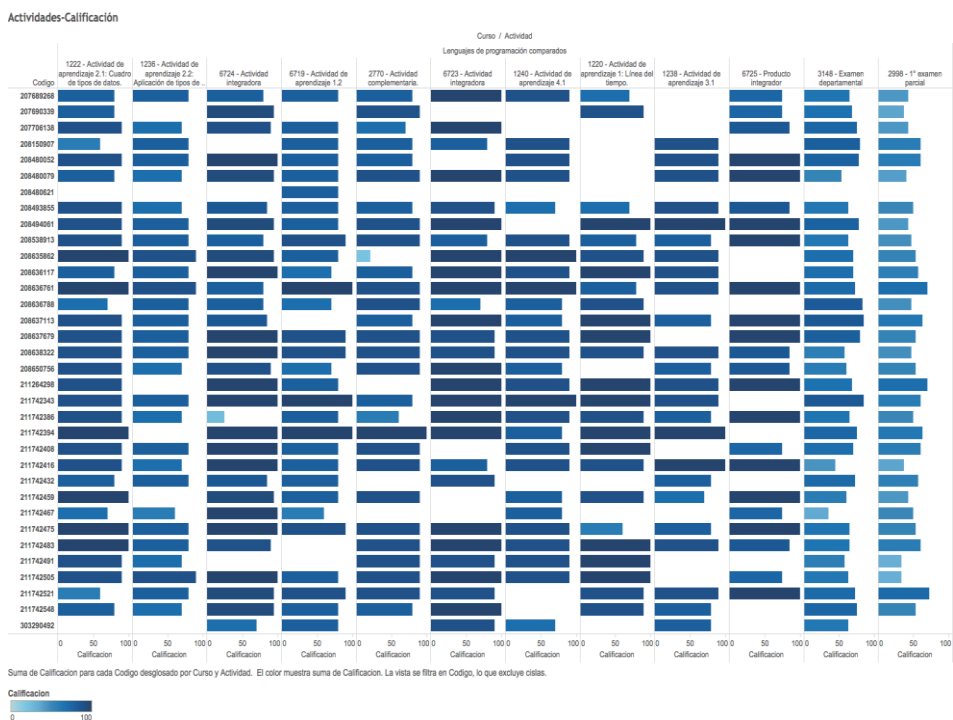


Fig. 5. Calificaciones por actividad.

3.4 Tiempo para elaborar un examen: el tiempo que designa un profesor para elaborar un examen son 40 minutos, en la Fig. 6 se muestra que los estudiantes utilizan un rango entre 8 y 29 minutos del tiempo, y no por ello tienen un mejor DA. El color rojo indica un bajo DA y el azul un mejor DA.

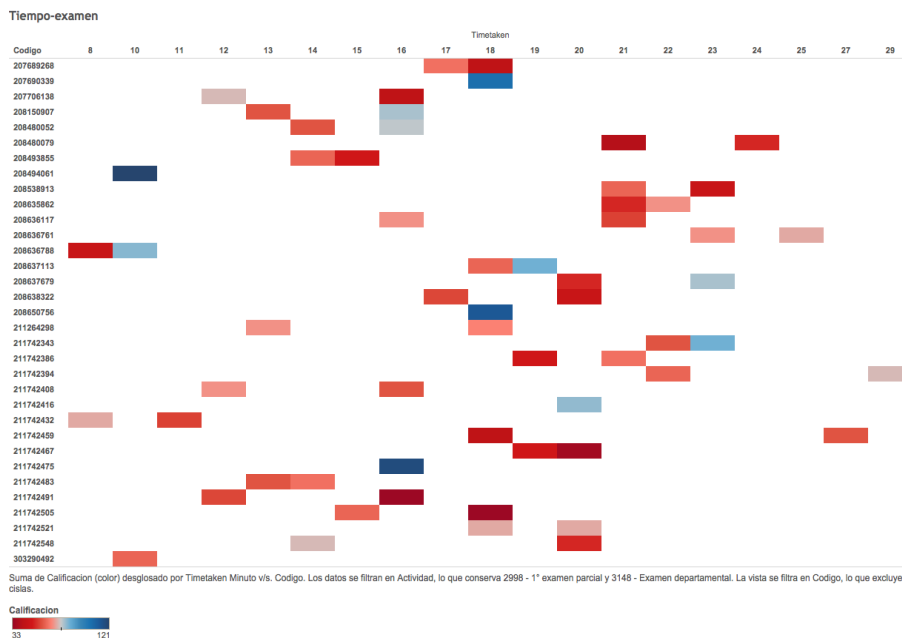


Fig. 6. Tiempo en minutos para elaborar un examen.



Fig. 7. Tiempo de anticipación para enviar una tarea.

3.5 Tiempo para el envío de tareas: la Fig. 7 muestra el tiempo anticipado con el que un estudiante envía una tarea o AA. En ella se comprueba que los estudiantes envían sus AA en las últimas horas (4) y es por ello que en algunos casos no existe participación, porque se les termina el tiempo para el envío de la misma (Fig. 5, rectángulos en blanco).



#### 4. Discusión

Con respecto a la clasificación de los cursos *b-learning*, de los 17 mostrados en la Fig. 3 y de acuerdo a las reglas designadas en el apartado 2.1, Moovisys solo catalogó a uno de ellos en esta categoría, porque Moodle, de acuerdo a la administración del CUP, registra a los estudiantes por un periodo de tiempo y al término, éstos son dados de baja automáticamente del curso, por lo tanto ya no cumplen con el límite de participantes.

El curso catalogado como *b-learning* corresponde al calendario 2013A (de febrero a junio de 2013). En Moodle cada vez que se utiliza el mismo curso para otro semestre, la participación de los estudiantes queda registrada en otras tablas de la base de datos, y no en las que se utilizaron para esta investigación.

Las AA en las que los estudiantes tienen un mejor DA, pero existe menos participación son las tareas (Fig. 5), y se observa que en los exámenes, los estudiantes tienen un bajo DA y participan la mayoría.

El tiempo que un estudiante consume (un rango de 8 a 29 minutos) para elaborar un examen no siempre significa un mejor DA (Fig. 7). El tiempo de anticipación que un estudiante utiliza para el envío de su tarea, es al término del día de entrega, se observa una carga mayor (rectángulos de color azul) en las últimas cuatro horas, siendo la hora cero (0) en donde se concentran mayores envíos (rectángulos de azul intenso); con ello se infiere que el estudiante deja las tareas al final, y los tiempos designados, podrían no alcanzar para enviarla, por lo cual existe menos participación en algunas de ellas.

El tiempo de pre-procesamiento de los datos dentro de *Moovisys* es menor, comparado con el que se utiliza manualmente. Moodle permite la descarga manual en el apartado de calificaciones; sin embargo esos datos aún deben procesarse y transformarse por cada curso que se desee. *Moovisys* realiza el pre-procesado de los datos de diversos cursos; en este artículo solo se utilizó uno, ya que las reglas y la base de datos de Moodle restringieron la obtención de una cantidad mayor de datos, sin embargo, no se limita a utilizar solo uno.

Las AA dentro de Moodle son variadas, sin embargo el profesor de este curso solo utiliza exámenes y tareas, por lo tanto se sugiere utilizar otras AA para motivar a los estudiantes. En las prácticas docentes, se propone concientizar a los estudiantes para que cumplan con sus AA en el tiempo debido. El tiempo que se utiliza en los exámenes es adecuado, sin embargo pueden existir diversos problemas por lo que el estudiante tiene un bajo DA en esta actividad.

#### 5. Trabajo futuro

*Moovisys* integrará módulos para la recolección de datos históricos de Moodle y a su vez de visualización, para que las autoridades académicas tomen decisiones para mejorar el DA de los cursos *b-learning* en un corto periodo de tiempo. Las decisiones estarán a cargo de la autoridad académica cuando la aplicación *Moovisys* muestre los gráficos de DA, sin embargo, para los resultados que se presentan en este artículo, una decisión podrá ser que el profesor debe poner atención en sus prácticas de evaluación (exámenes) porque es donde los estudiantes muestran un bajo DA.

## Referencias

1. Aguilera, A., Riascos, S.: Direccionamiento estratégico apoyado en las TIC. *Estudios Gerenciales*, Vol. 25, No. 111, pp. 127–143 (2009)
2. Romero, C., Ventura, S.: Data mining in education. *Data Mining and Knowledge Discovery*, Vol. 3, No. 1, pp. 12–27 (2013)
3. Callagher, M.: How can Student Interactivity be Enhanced through the use of a Blended Learning Approach (2008)
4. Ortiz, L.: Campus Virtual: la educación más allá del LMS. *Revista de la Universidad y Sociedad del Conocimiento*, Vol. 4, No. 1, pp. 1–7 (2007)
5. Han, J., Kamber, M., Pei, J.: *Data Mining Concepts and Techniques*. 3rd ed. Morgan Kaufmann, Waltham, MA, USA (2011)
6. Cios, K., Pedrycz, W., Swiniarski, R., Kurgan, L.: *Data mining: a knowledge discovery approach*. Springer (2007)
7. Maimon, O., Rokach, L. (eds): *Data Mining and Knowledge Discovery Handbook*. 2nd ed. Springer US (2010)
8. Witten, I., Frank, E., Hall, M.: *Data Mining Practical Machine Learning Tools and Techniques*. 3 ed. Morgan Kaufmann, Burlington, MA, USA (2011)
9. Zaïane, O.: Introduction to Data Mining. *Proceedings, CMPUT690 Principles of Knowledge Discovery in Databases*, University of Alberta, pp. 1–15 (1999)
10. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P.: From Data Mining to Knowledge Discovery in Database. *AI Magazine*, Vol. 17, No. 3, pp. 37–54 (1996)
11. Pal, N., Jain, L. (Eds.): *Advanced techniques in knowledge discovery and data mining*. Springer-Verlag, London (2005)
12. Bousbia, N., Belamri, I.: Which Contribution Does EDM Provide to Computer-Based Learning Environments? In: Peña-Ayala, A. (Ed.) *Educational Data Mining: Applications and Trends*, Vol. 524, Springer International Publishing (2014)
13. Tejedor, F., García-Varcárcel, A.: Causas del bajo rendimiento del estudiante universitario (en opinión de los profesores y alumnos): propuestas de mejora en el marco del EEES. *Revista de educación*, No. 343, pp. 443–473 (2007)
14. Martínez, R., Heredia Y.: Tecnología educativa en el salón de clase. *Revista Mexicana de Investigación Educativa*, pp. 371–390 (2010)
15. Edel, R.: El rendimiento académico: concepto, investigación y desarrollo. *Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, Vol. 1, No. 2 (2003)
16. Mazza, R., Milani, C.: GISMO: a Graphical Interactive Student Monitoring Tool for Course Management Systems. *International Conference on Technology Enhanced Learning*, pp. 1–8, Milan, Italy (2004)