

Aplicación de técnicas de Inteligencia de Negocio al Seguimiento del Aprendizaje en MERLÍN

Mario A. Muñoz
Departamento de Ingeniería
de Sistemas Telemáticos
Universidad Politécnica de Madrid
mario@gsi.dit.upm.es

Jorge Gonzalo Alonso
Departamento de Ingeniería
de Sistemas Telemáticos
Universidad Politécnica de Madrid
jorge@gsi.dit.upm.es

Carlos A. Iglesias
Germinus XXI
cif@germinus.com

Resumen—Este artículo propone la aplicación de técnicas de inteligencia de negocio para el seguimiento de alumnos en la plataforma educativa MERLÍN, basada en portlets. Se hace una revisión de las funcionalidades de seguimiento existentes en las plataformas de *e-learning* más populares y se propone un módulo de seguimiento integrado basado en código abierto. Este módulo desarrolla un panel de control del estudiante que permite analizar su progreso desde diferentes perspectivas, que se componen con tecnología de portlets.

I. INTRODUCCIÓN

Durante los últimos años se ha popularizado y extendido el uso de las plataformas de aprendizaje a distancia o *e-learning*. El principal elemento diferenciador de estos sistemas de enseñanza frente a métodos tradicionales se encuentra en que profesores y alumnos pueden estar separados físicamente, comunicándose de forma bidireccional a través de canales asíncronos. El uso de Internet como medio de comunicación y distribución del conocimiento permite que el alumno sea responsable y gestor de su propio aprendizaje, siempre supervisado por tutores externos.

La principal ventaja de estas plataformas radica en su *flexibilidad*, puesto que permite el uso de una gran variedad de herramientas de comunicación tanto síncronas (chat, videoconferencia) como asíncronas (foros de debate, e-mail, grupos de noticias, etc.) que facilitan la comunicación entre alumnos y tutores.

Los elementos que conforman una solución típica de *e-learning* son la plataforma, los contenidos educativos y las herramientas de comunicación. Las plataformas actuales más populares, como Moodle [1] o Dokeos [2], facilitan la tarea del personal docente en cuanto a distribución y organización de la información, además permiten a los alumnos el acceso a esta información y fomentan la participación a través de foros de debate y encuestas. Estas plataformas, sin embargo, presentan ciertas limitaciones en la actualidad, como la dificultad de *personalización* y la rigidez a la hora de crear y distribuir contenidos. Otra importante carencia de las plataformas de *e-learning* actuales se encuentra en el *seguimiento* de los alumnos. Moodle, por ejemplo, ofrece información detallada del comportamiento del alumno en la plataforma, pero muy orientada a su historial de navegación, lo que dificulta realizar un seguimiento efectivo de la actividad realizada por los

alumnos.

Este trabajo de investigación se enmarca en el proyecto MERLÍN, cuyo objetivo es construir una plataforma educativa personalizable y participativa, empleando tecnologías web2.0, y empleando un contenedor de portlets JSR-168 [3] como plataforma. El artículo se centra en el módulo de seguimiento de los alumnos de MERLÍN, para el que se han aplicado técnicas de inteligencia de negocio (OLAP [4], presentación de datos, cuadros de control, etc.) que permiten obtener y presentar de forma organizada y comprensible datos representativos del aspecto que se pretende analizar.

El resto del artículo se organiza como sigue. La sección II presenta brevemente el proyecto MERLÍN en el que se enmarca el trabajo de investigación. A continuación, se realiza una revisión de las funcionalidades de seguimiento de alumnos disponibles en las principales plataformas de *e-learning* en la sección III. La sección IV-B presenta el módulo de seguimiento de MERLÍN. Por último, la sección V recoge las principales conclusiones y los trabajos futuros derivados de ese trabajo.

II. EL PROYECTO MERLÍN

El proyecto MERLÍN [5] es un proyecto PROFIT coordinado por Idea Informática, y en el que participan además la Universidad Politécnica de Madrid, la Universidad Complutense de Madrid e Infinity Group. El proyecto se encuentra en su segundo año de desarrollo. El proyecto nace gracias a la amplia experiencia de Idea Informática en el sector educativo, en que ha desarrollado los portales educativos EducaMadrid, EduCantabria, Scola Lliurex, Internet en el aula o EducarEx, orientados principalmente a los ciclos de primaria y secundaria. En estos entornos, la plataforma educativa se complementa con soluciones de *e-learning* como Moodle, mientras que la gestión de sitios web, comunidades, y herramientas de colaboración se realiza con una plataforma de portlets como Liferay [6]. Tras analizar las funcionalidades ofrecidas por las principales plataformas de E-Learning, MERLÍN propone integrar estas funcionalidades en una plataforma de portlets, e integrar funcionalidades web2.0 para favorecer que los alumnos puedan participar más activamente en el aprendizaje. Actualmente, las plataformas como Moodle están principalmente pensadas para el profesor, ofreciendo

una capacidad muy limitada a los alumnos. La plataforma MERLÍN permitirá que los alumnos puedan anotar con etiquetas los recursos educativos, comentarlos, puntuarlos, publicar sus propios blogs,... En definitiva, MERLÍN es una plataforma de Educación 2.0. Entre las diferentes líneas de investigación de MERLÍN, destacan:

- la gestión de contenidos educativos conforme a estándares como IMS LOM y SCORM
- integrar un motor de aventuras educativas
- integrar herramientas de evaluación con IMS QTI
- investigar en la mejora de usabilidad y experiencia de usuario para profesor y alumno en la creación y acceso a los recursos educativos
- investigar en sistemas avanzados de seguimiento de los progresos de los alumnos, y en su interrelación con el módulo de evaluación de los alumnos, cuyos resultados se presentan en este artículo.

El proyecto MERLÍN es un proyecto basado en código abierto y que será explotado siguiendo el modelo de código abierto.

III. SEGUIMIENTO DE ALUMNOS EN LOS ENTORNOS DE *e-learning*

En esta sección se revisan la funcionalidad de seguimiento de alumnos ofrecida por principales entornos de *e-learning*, como Moodle (sección III-A), Dokeos (sección III-B), Sakai (sección III-C), ATutor (sección III-D) y WebCT (sección III-D).

III-A. Moodle

Moodle [1] es un sistema de gestión de cursos de libre distribución (Course Management System, CMS) que ayuda a los educadores a crear comunidades de aprendizaje en línea.

Una de las características más atractivas de Moodle, que también aparece en otros gestores de contenido educativo, es la posibilidad de que los alumnos participen en la creación de glosarios, y en todas las lecciones se generan automáticamente enlaces a las palabras incluidas en estos.

Algunas actividades pueden ser un poco mecánicas, dependiendo mucho del diseño instruccional para Moodle, lo que lo hace más dinámico y atractivo. Por estar basado en tecnología PHP, la configuración de un servidor con muchos usuarios debe ser cuidadosa para obtener el mejor rendimiento.

El seguimiento de alumnos en Moodle es uno de sus puntos débiles. La información que ofrece no está demasiado estructurada, por lo que puede resultar complejo y costoso en tiempo extraer conclusiones útiles a partir de los datos suministrados por la aplicación. En sus versiones más habituales, Moodle ofrece dos áreas diferenciadas para realizar seguimiento de alumnos:

Informes (Reports): ofrece dos tipos de informes, que se detallan a continuación.

Registros de uso: proporciona información sobre la actividad de los usuarios registrados en el curso dentro del portal. Los datos se presentan en lista ordenada por tiempo

que contiene las acciones realizadas por cada usuario. Los datos que incluye son:

- Fecha y hora de la acción.
- Dirección IP desde donde se realizó.
- Acción realizada.
- Información sobre la acción.

La información proporcionada es amplia y muy variada. Sin embargo, la forma de presentarla, sin apenas tratamiento y casi tal como se obtiene de la base de datos de la aplicación, reduce en gran medida la utilidad que puede obtener el profesor en actividades de seguimiento.

Informe de actividades: en este caso, se ofrece un listado con las actividades programadas por el profesor en el portal, divididas por temas y mostrando los distintos tipos de ejercicios propuestos y el material de consulta disponible para los alumnos.

Niveles (Grades): este módulo muestra las calificaciones obtenidas por los estudiantes dentro de las actividades propuestas por el profesor en un curso determinado. La información, como en los informes anteriores, se muestra en forma de lista y sin apenas tratamiento. La tabla, muestra los estudiantes y las calificaciones obtenidas en cada actividad. Ofrece funcionalidades de ordenamiento (por orden alfabético del estudiante) y un pequeño análisis de las notas obtenidas (se muestra estadísticos como máximo, mínimo, media, mediana y desviación típica en las notas). De nuevo, la forma de presentar la información reduce su utilidad para el profesor.

III-B. Dokeos

Dokeos es un entorno de e-learning y una aplicación de administración de contenidos de cursos y también una herramienta de colaboración. Es software libre y está bajo la licencia GNU GPL2, el desarrollo es internacional y colaborativo. También está certificado por la OSI3 y puede ser usado como un sistema de gestión de contenido (CMS) para educación y educadores. Esta característica para administrar contenidos incluye distribución de contenidos, calendario, proceso de entrenamiento, chat en texto, audio y vídeo, administración de pruebas y guardado de registros.

Las principales metas de Dokeos son ser un sistema flexible y de muy fácil uso mediante una interfaz de usuario sumamente amigable. Ser una herramienta de aprendizaje, especialmente recomendada a usuarios que tengan nociones mínimas de computación cuyo objetivo es la preocupación por el contenido. [2]

En cuanto a seguimiento de alumnos, Dokeos es superior a Moodle en cuanto a presentación y calidad de la información presentada. El seguimiento se realiza accediendo a la página del portal "Informes", donde en una lista se muestran distintos campos de la evolución del alumno en el portal:

- Nombre y apellidos.
- Tiempo de permanencia en el portal.
- Progreso en %.
- Puntuación obtenida en %.
- Número de trabajos enviados.
- Mensajes publicados.

- Fecha de la última conexión.

Accediendo a la página “Detalles”, se muestra un panel con información acerca del usuario en cuestión. En este panel la información se muestra ya clasificada por campos:

- Datos del usuario.
 - Nombre.
 - Correo electrónico.
 - Teléfono.
 - Estado de conexión actual.
 - Seguimiento.
 - Última conexión.
 - Tiempo de permanencia en la plataforma.
 - Progreso.
 - Puntuación.
 - Acción.
 - Enviar correo.
- Datos del curso.
 - Itinerarios de aprendizaje, con tiempo empleado, puntuación, progreso, fecha de la última conexión. Ejercicios realizados, con puntuación e intentos.
 - Otras herramientas, con número de trabajos enviados, mensajes publicados, enlaces visitados y documentos descargados.

Esta forma de presentar la información ofrece una mayor utilidad para el profesor, puesto que se ofrece clasificada por distintos aspectos y centralizada en la aplicación.

III-C. Sakai

Sakai [7] es una comunidad formada por instituciones académicas y organizaciones comerciales para desarrollar un entorno colaborativo de aprendizaje. Sakai está basado en Java, lenguaje que permite desarrollar aplicaciones escalables, fiables, interoperables y extensibles.

El Proyecto Sakai fue concebido para desarrollar software educativo de código abierto. El objetivo del Proyecto Sakai es crear un entorno de colaboración y aprendizaje para la educación superior, que pueda competir con sus equivalentes comerciales Blackboard/WebCT [8] y que mejore otras iniciativas de código abierto como Moodle.

Sakai pretende ser una plataforma que aune las ventajas de un sistema gestor de contenidos (CMS) y un sistema de aprendizaje a distancia (LMS), pretende que se use para enseñar, investigar y colaborar. El Software de Sakai posee las características comunes de un sistema de aprendizaje LMS además de contar con múltiples funcionalidades de comunicación entre profesores y alumnos, entre las más importantes la plataforma cuenta con el lector de noticias RSS, la herramientas de distribución de material docente, de realización de exámenes, de gestión de trabajos, chat y wiki entre otros.

Las herramientas destinadas al seguimiento de alumnos que ofrece Sakai siguen el enfoque tradicional de las aplicaciones descritas en los apartados anteriores. Comparativamente, ofrece mayor funcionalidad que Moodle en este aspecto, y unas características similares a las de Dokeos.

El mecanismo de seguimiento de Sakai es el llamado “Libro de notas”. El libro de notas permite al profesor listar los cursos que tiene asignados y las notas de cada alumno, pudiendo calcular, almacenar y distribuir la información de las notas a los estudiantes haciendo uso de la web. Los cursos pueden ser evaluados con diferentes escalas.

III-D. Otras plataformas de tele-enseñanza

En este apartado se analizan de forma más general otras aplicaciones de tele-enseñanza a tener en cuenta:

WebCT [9] WebCT (*Web Course Tools*, o Herramientas para Cursos Web) es un sistema comercial de aprendizaje virtual online, el cual es usado principalmente por instituciones educativas para el aprendizaje a través de Internet. La flexibilidad de las herramientas para el diseño de clases hace este entorno muy atractivo tanto para principiantes como usuarios experimentados en la creación de cursos en línea. Los instructores pueden añadir a sus cursos WebCT varias herramientas interactivas tales como: tableros de discusión o foros, sistemas de correos electrónicos, conversaciones en vivo (chats), contenido en formato de páginas web y archivos PDF entre otros. Una importante crítica recibida por WebCT, especialmente la versión Vista, es que rompe muchas de las pautas de uso de Internet, no siendo accesible por alumnos con discapacidades.

ATutor [10] ATutor es un Sistema de Gestión de Contenidos de Aprendizaje (Learning Content Management System, LCMS) de código abierto basado en web y diseñado con el objetivo de lograr accesibilidad y adaptabilidad. Los administradores pueden instalar o actualizar ATutor en minutos. Los educadores pueden rápidamente configurar, empaquetar y redistribuir contenido educativo, y llevar a cabo sus clases online. Los estudiantes pueden formarse en un entorno de aprendizaje adaptativo. Contiene herramientas de gestión y administración de alumnos, tutores, cursos y evaluaciones en línea, autoría y colaboración. Incorpora las especificaciones de empaquetado de contenido IMS/SCORM4, permitiendo que los diseñadores creen contenido reutilizable que se puede intercambiar entre diversos sistemas de aprendizaje. El contenido creado en otros sistemas conforme a IMS o SCORM se puede importar en ATutor, y viceversa.

IV. LA PROPUESTA DE MERLÍN: INTELIGENCIA DE NEGOCIO

IV-A. Inteligencia de Negocio en software educativo

En todas las herramientas estudiadas y explicadas en el apartado anterior vemos una deficiencia en la minería de datos. Así pues, creemos necesario un procesamiento de los datos para extraer el conocimiento y mostrarlo de una manera que permita tener una mejor visión del estado de un cierto alumno, clase o profesor.

La aplicación de centros de control inteligentes a plataformas educativas es un campo poco explorado en la actualidad, por lo que resulta de interés la investigación en este ámbito. Se entiende por un centro de control inteligente a una consola de gestión de un sistema con características que

añadan un valor añadido a información cruda o de bajo nivel para hacerla más comprensible y manipulable por parte de un agente humano, como pueden ser:

- Adaptabilidad de información de bajo nivel a métricas de mayor nivel de abstracción.
- Correlación de información para la interpretación de diferentes eventos, como incidencias.
- Predicción de comportamientos determinados a partir de las distintas métricas.

Este tipo de centros de control son de aplicación generalizada en el campo de la inteligencia de negocio, que, atendiendo a un punto de vista empresarial, podría definirse como el proceso de transformación de datos en información y de información en conocimiento. Análogamente, pueden ser aplicados a otros campos como el aprendizaje asistido por tecnologías de la información (*e-learning*), ámbito en el que la utilización de centros de control dentro de una plataforma educativa facilitaría el acceso a información masiva de alumnado de una forma enriquecida para posibilitar análisis detallados de aspectos del alumnado.

En los centros de control inteligentes tenemos alternativas comerciales como Business Objects o Hyperion y, como alternativa de código abierto, destaca Pentaho BI [11], que ofrece una arquitectura J2EE y que puede ser integrado en entornos de portlets. Pentaho BI ofrece componentes de visualización de los datos, así como componentes de análisis, incluyendo módulos de informes, procesado analítico en línea u OLAP (del inglés *On Line Analytical Processing*) [12], minería de datos, centros de control y tarjetas de puntuación.

IV-B. Seguimiento inteligente en MERLÍN

Como parte del proyecto MERLÍN se pretende analizar el uso e integración de sistemas de inteligencia de negocio para el seguimiento de alumnos, tanto personalizada como en grupo, de forma que la gestión de la información de negocio se realice de forma centralizada en la propia plataforma. Además, se investigan las ventajas resultantes de la aplicación de técnicas inteligentes en el seguimiento de la evolución y los progresos conseguidos por los alumnos a lo largo del curso.

Este trabajo propone y desarrolla una arquitectura para la integración de un sistema de inteligencia de negocio en una plataforma de portlets aplicada al ámbito de *e-learning*. Con este fin, se ha realizado una integración entre el contenido de portlets Liferay [6], que es compatible con las especificaciones JSR-168 [3] y JSR-286 [13].

Los principios de definición de la arquitectura han sido:

- *Orientación a servicios*. Aprovechando la flexibilidad de la plataforma J2EE, y su escalabilidad, el servicio de análisis inteligente de datos se integra para que pueda ser dimensionado y no penalice el rendimiento global del portal. Por tanto, los componentes portlets de minería de dato invocarán este servicio, que puede ser ofrecido en local o en remoto.
- *Personalización y facilidad de uso*. Para que los profesores acepten la herramienta de seguimiento, ésta

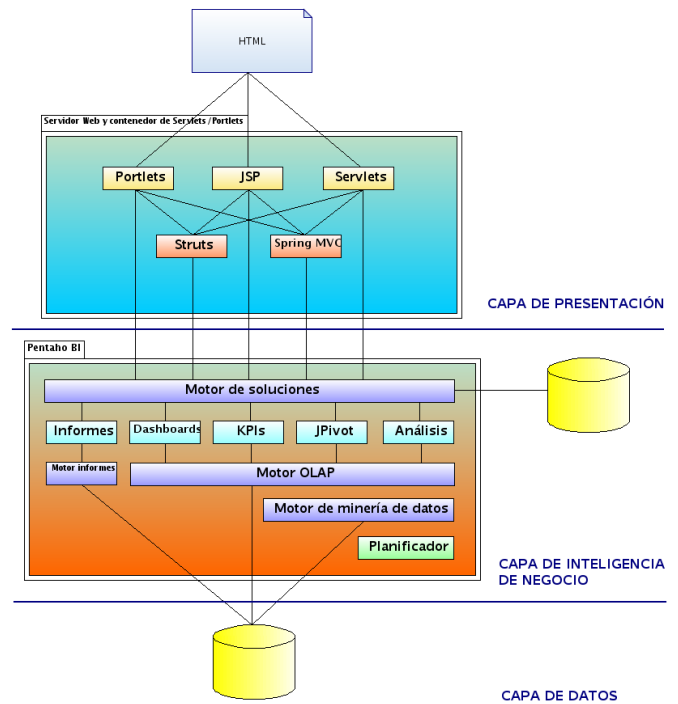


Figura 1. Arquitectura de MERLÍN

debe ser fácil de uso y, a la vez, permitir su personalización. Con este objeto, se ha desarrollado una interfaz de panel de control que permite que el profesor seleccione las vistas que en que está interesado, y empleando el mecanismo estándar de comunicación entre portlets, son actualizadas al seleccionar un alumno. Se han desarrollado informes básicos para facilitar el uso de la herramienta, y mediante la colaboración social, el sistema puede incluir fácilmente nuevos informes. Además, las herramientas como OLAP ofrecen una interfaz interactiva que permite la composición de consultas visualmente.

MERLÍN define una arquitectura en tres capas para la aplicación de seguimiento (Figura 1). Una capa superior de presentación, constituida en su mayor parte por el software de Liferay, ofrece la interfaz de usuario para la gestión y consulta de los informes. Estos informes son generados mediante Pentaho, integrado como una capa intermedia de inteligencia de negocio, que se encarga de ejecutar bajo demanda los procesos de negocio definidos en un repositorio. Los informes se alimentan a partir de los datos almacenados en la capa inferior, constituida por el sistema gestor de bases de datos donde se guarda toda la información de la aplicación siguiendo los criterios establecidos por los modelos de datos definidos para tal fin.

Panel de control del alumno Para alcanzar funcionalidades de seguimiento análogas a las de los sistemas usuales, MERLÍN propone la definición de un "Panel de control del alumno". Este panel de control está compuesto por varios

portlets gráficos que se configuran de forma centralizada a través de otro componente en forma de portlet, denominado "Filtro". Con este método, se alcanza una gran flexibilidad en cuanto a la configuración del panel, puesto que es posible realizarla en dos sentidos. En primer lugar, y gracias a las características del portal Liferay, el usuario tendrá la capacidad de seleccionar los portlets que componen su panel de seguimiento, cambiar su tamaño o reubicarlos, siguiendo la metáfora de escritorio usual en este tipo de portales. Además, a través del portlet Filtro, el usuario puede seleccionar los contenidos que desea que se muestren en los portlets visibles, además de alterar la configuración de los gráficos mostrados a su antojo.

Implementación de la comunicación entre portlets: Dada la filosofía del panel del alumno, en el que los portlets de información muestran el contenido seleccionado en el filtro, ha sido necesario implementar mecanismos de comunicación entre portlets. Estos mecanismos son provistos por la reciente especificación Portlets 2.0 en la JSR-286 [13].

Aplicación de técnicas inteligentes: reporting, OLAP y minería de datos En cuanto a la aplicación de tecnologías complejas de inteligencia de negocio, normalmente ligadas de manera íntima al mundo económico y empresarial, se propone la creación de un catálogo de informes accesible a través de un nuevo elemento en forma de portlets. Las tecnologías implicadas en este tipo de informes serán el reporting, las estructuras OLAP [4] y los algoritmos de minería de datos [14].

La inmensa mayoría de organizaciones utilizan técnicas de reporting en alguna de sus variantes. Aunque se trata de un componente con un grado de interactividad reducido, permite la creación rápida de listados que pueden incluir gráficos sobre los datos, codificación de colores, agregación de resultados, etc. Como resultado, se considera el reporting como una necesidad central de cualquier sistema de inteligencia de negocio, siendo normalmente la primera tecnología en desplegarse para su uso, por ser la más sencilla en cuanto a creación y utilización de informes. El uso de Pentaho como software de inteligencia, cubre ampliamente las necesidades de reporting de MERLÍN, tanto en la creación de informes como en su publicación dentro de la plataforma. La tecnología empleada permite extender en cierto modo las limitaciones comentadas en cuanto a interactividad, puesto que facilita la creación de informes parametrizables por el usuario, además de la exportación en varios formatos usuales, como documentos de MS Office o PDF).

El siguiente paso en la inclusión de técnicas inteligentes ha sido la incorporación de estructuras de análisis y visualización de datos basadas en OLAP. La tecnología OLAP se basa en la utilización de estructuras multidimensionales denominadas de forma genérica "Cubos" para el almacenamiento y posterior análisis de datos. La principal ventaja que supone el uso de esta tecnología radica en la rapidez de procesamiento de operaciones analíticas en este tipo de estructuras, puesto que para realizar una búsqueda sobre un dato concreto basta con indexar el miembro al que hace referencia en

las dimensiones deseadas. OLAP permite realizar este tipo de operaciones prácticamente en tiempo real, a diferencia de los sistemas gestores de bases de datos tradicionales, en los que este tipo de operaciones supondría la ejecución en cadena de consultas complejas que requerirían un elevado coste computacional. Además de esta evidente ventaja en lo referente al tiempo de cálculo en el análisis de datos, la combinación de estructuras OLAP con el adecuado software de presentación de la información posibilita la creación de informes altamente flexibles en los que es posible interactuar con el cubo de forma que se pueda mostrar casi cualquier información que contiene. Pentaho ofrece esta posibilidad a través de la incorporación de los componentes de código abierto Mondrian [15] y JPivot. Mondrian actúa como servidor OLAP, manteniendo en memoria las estructuras creadas a partir de un esquema definido sobre el modelo de datos de la aplicación. JPivot, por su parte, permite el acceso a la estructura del cubo desde una JSP, de forma que es posible mostrar el cubo en una tabla interactiva acompañada de menús que permiten alterar la estructura (por ejemplo, rotándolo o filtrando miembros y dimensiones), reordenar los datos, representar gráficos sobre ellos, o exportar el resultado del informe a una tabla de Excel o un fichero PDF. Utilizando esta técnica es posible la creación de estructuras que contengan información simultánea sobre campos dispares aplicados al seguimiento de alumnos. Por ejemplo, pueden crearse tablas sobre resultados históricos en ejercicios evaluados, o realizar comparativas por alumnos teniendo en cuenta datos de edad, asignaturas cursadas o nacionalidad, con lo que se lograría crear sectores de población diferenciados a la hora de realizar el análisis.

La última tecnología clave de inteligencia de negocio empleada ha sido la minería de datos. Se entiende por minería de datos la extracción de conocimiento procesable a partir de información implícita en una base de datos. Existe una gran variedad de algoritmos utilizados para realizar este tipo de tareas, pero la especificación inicial de MERLÍN preve que, dada su utilidad real en el caso del seguimiento de alumnos, se haga uso del grupo de algoritmos de clustering. Esta elección se debe a la naturaleza de las métricas de seguimiento definidas en la especificación de la aplicación, en su mayoría, datos de carácter numérico relacionados con las calificaciones obtenidas por los alumnos y estadísticos de estas (máximos, mínimos y medias).

En cuanto a los parámetros de clasificación, inicialmente, en el proyecto MERLÍN se ha experimentado con la ejecución de procesos de clustering a partir de las calificaciones obtenidas por los alumnos en las tareas definidas en la plataforma. Esto permite diferenciar grupos de alumnos por su rendimiento en un determinado curso, y de esta forma determinar posibles carencias educativas o grupos de riesgo en cuanto a comportamiento. La clasificación obtenida en el proceso de clustering se utiliza para alimentar una estructura OLAP como las ya explicadas, de modo que se pueda contar con toda la potencia de esta tecnología para el análisis de datos junto con la flexibilidad que ofrece en cuanto a presentación

de la información. La tecnología empleada para la ejecución de algoritmos de minería de datos ha sido la plataforma Weka [16]. Weka ofrece una aplicación independiente, junto con una serie de librerías Java, que incluye una colección completa de algoritmos de minería de datos, tanto supervisados como no supervisados. Dentro de las librerías se pueden encontrar herramientas complementarias que facilitan la creación de conjuntos de instancias sobre los que realizar las operaciones de minería y la gestión de éstos. Dada la gran variedad de algoritmos ofrecidos por Weka, cabe pensar que en futuras ampliaciones del proyecto puedan incluirse funcionalidades innovadoras y de interés en el campo de la tele-formación como pueden ser la creación de modelos de predicción a partir de datos históricos.

IV-C. Alternativas

Dentro de las aplicaciones libres de tele-formación analizadas para establecer el estado del arte de las tecnologías relacionadas con los objetivos de MERLÍN, no se ha encontrado ninguna que haga uso extensivo de tecnologías de inteligencia de negocio. Como principal alternativa libre a MERLÍN en cuanto a seguimiento de alumnos, puede señalarse Dokeos como la herramienta que realiza un cierto proceso de la información “cruda”, tal y como se almacena en las tablas del sistema gestor de bases de datos de la aplicación, para obtener información de mayor utilidad, como el registro de los progresos de los alumnos en los cursos.

En lo que concierne a equivalentes comerciales, empresas relacionadas con tecnologías de inteligencia de negocio, como SAS o Cognos, ofrecen soluciones destinadas al ámbito de la educación. En cualquier caso, y pese a ofrecer algunas funcionalidades de registro y presentación de los progresos de los alumnos, están más enfocadas a la gestión económica de los recursos educativos, por lo que constituyen aplicaciones más similares al software de gestión empresarial. Dado el carácter comercial de estas herramientas, no se ha profundizado en su análisis, por lo que los datos resumidos en este artículo son los que aportan las propias organizaciones sobre esta gama de productos.

IV-D. Resultados

Siguiendo las pautas de diseño establecidas en el desarrollo de la aplicación de seguimiento, se ha implementado una primera versión que presenta los siguientes componentes y portlets:

Panel de seguimiento de alumnos: se trata de un conjunto de portlets sencillos con los que puede configurarse un panel o cuadro de control, de manera que se ofrece una interfaz visual que proporciona información rápida sobre el estado actual y la evolución de los alumnos en la plataforma. Los componentes con los que puede construirse este panel son los siguientes (Figura 2):

- **Filtro** proporciona funcionalidades de configuración sobre el resto de elementos del panel, de forma que el usuario puede elegir el formato de los gráficos o la información a visualizar.

Tarea: Tarea 7		
ID de usuario	Nombre de usuario	Resultado
10804		0
10826		0
11601		4
11623		8
11644		2
11666		0
Media en la tarea Tarea 7		2,33

Tarea: Tarea 8		
ID de usuario	Nombre de usuario	Resultado
10804		7
10826		9
11601		7
11623		1
11644		8
11666		2
Media en la tarea Tarea 8		5,67

Tarea: Tarea 9		
ID de usuario	Nombre de usuario	Resultado
10804		2
10826		6
11601		2
11623		0
11644		1
11666		7
Media en la tarea Tarea 9		4,33
Media de todos los resultados		4,02

Figura 3. Generación de diferentes documentos

- **Evolución del alumno en las tareas** muestra un gráfico configurable con la evolución de los resultados en las tareas obtenidos por el alumno elegido.
- **Evolución del grupo en las tareas** gráfico configurable con la evolución de grupo de alumnos que componen el curso en las tareas. Este gráfico tiene en cuenta tres métricas por tarea: calificación máxima, mínima y media.
- **Evolución del alumno en las preguntas de test** muestra un gráfico configurable con los resultados obtenidos por el alumno elegido en las preguntas del test seleccionado.
- **Evolución del grupo en las preguntas de test** gráfico configurable mostrando los resultados del conjunto de alumnos que conforman el curso en las preguntas de test seleccionado. De nuevo, se utilizan como métricas la calificación máxima, mínima y media.
- **Estado del alumno** gráfico en forma de dial con codificación de color que representa la situación en cuanto a resultados del alumno elegido respecto a sus compañeros en el curso.
- **Estado del curso** dial con codificación de color que representa la situación en cuanto a resultados del curso respecto al resto de cursos registrados en la plataforma.

Catálogo de informes predefinidos: proporciona acceso a un conjunto de informes predefinidos disponibles, de tipo estático o creados a partir de cubos OLAP con o sin funcionalidades de minería de datos.

Los informes de tipo estático: utilizan herramientas de reporting para generar el documento seleccionado. Su utilidad principal consiste en proporcionar listados de resultados en tareas y test exportables de forma inmediata a varios formatos (PDF, excel, word, CSV). Dentro de la propuesta inicial, se incluyen informes con listados sobre evolución en tareas y test, que pueden incluir o no información relativa a la nacionalidad de los alumnos registrados en la plataforma (Figura 3).

Los informes de tipo OLAP: proporcionan una tabla interactiva con capacidad para realizar distintas operaciones sobre los datos mostrados. Estos informes pueden alimentarse directamente de la capa de datos del portal, o de información obtenida a partir de procesos de clustering (minería de



Figura 2. Panel de seguimiento de los alumnos

datos) que permiten clasificar los alumnos en grupos dependiendo de distintos factores. En los ejemplos de la figura, pueden observarse las tablas que representan las estructuras multidimensionales, con los controles y el menú superior que permite interactuar con ellas. A través de los controles incluidos en la propia tabla, se pueden realizar operaciones como la extensión de alguna de las dimensiones (puede pasarse de mostrar el agregado de una dimensión a sus detalles), o el reordenamiento de los datos contenidos en las columnas de medidas. El menú superior ofrece mayores opciones de interactividad. En primer lugar, es posible realizar operaciones de filtrado e intercambio de filas y columnas individuales, con lo que se consiguen de forma inmediata facilidades para alterar la estructura del cubo en el grado que se desee. Dentro del mismo menú pueden encontrarse opciones adicionales para rotar la estructura del cubo, intercambiando el lugar de todas las filas y columnas. Una funcionalidad muy interesante de estos informes consiste en la representación de gráficos con la estructura seleccionada para el cubo en un momento determinado. Se ofrece una interfaz completa para la configuración de estos gráficos, pudiendo seleccionarse entre varios tipos (barras, líneas, área, sectores, etc.), así como aspectos como la orientación, el tamaño o la leyenda de los ejes. Se ofrecen además opciones de configuración para la impresión del informe así como para su exportación a una hoja de cálculo de Microsoft Excel. Dentro de los informes OLAP definidos inicialmente en MERLÍN, pueden encontrarse tablas de análisis sobre resultados obtenidos en tarea teniendo en cuenta distintos parámetros (nacionalidad, tipo de tarea, alumno, etc.), e informes de análisis teniendo en cuenta como

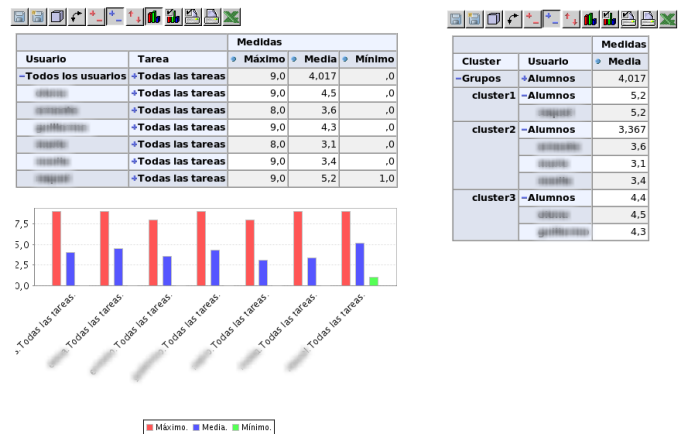


Figura 4. Minería de datos en informes OLAP

dimensión la clasificación realizada sobre el grupo de alumnos en el proceso de clustering (Figura 4).

V. CONCLUSIONES Y TRABAJOS FUTUROS

En el proyecto MERLÍN se propone una nueva visión de las plataformas educativas dando un salto cualitativo con respecto a las ya existentes gracias, en parte, a la utilización de técnicas inteligentes para el seguimiento de alumnos.

En este artículo se ofrece un estudio de las herramientas actuales para luego desarrollar la aportación que el proyecto MERLÍN hace sobre ellas. Entre estas aportaciones destacan, en el campo de las técnicas inteligentes, la minería de datos

(en concreto, agrupación de interacciones mediante algoritmos de clustering), OLAP y los cuadros de control.

A la vista del trabajo realizado, se hace notable la importancia de la calidad de la información, concepto que remarca la utilidad real de los datos para el usuario que la recibe. La utilización de técnicas de inteligencia de negocio ofrece soporte para conseguir esta información útil, puesto que su concepción es la de extraer conocimiento que facilite las tareas de toma de decisiones.

MERLÍN pone de manifiesto, aún más si cabe, la importancia del software libre. Todos los componentes del proyecto han sido desarrollados empleando estándares y aplicaciones libres, lo que demuestra la posibilidad de desarrollar aplicaciones potentes y tecnológicamente avanzadas sin necesidad de recurrir al software propietario para ello. Del mismo modo, MERLÍN también se ofrece como software libre al contrario que el resto de plataformas estudiadas.

Durante el desarrollo del proyecto la contribución a la comunidad ha sido muy importante en tanto que se ha logrado una integración entre el portal Liferay y la aplicación Pentaho. Y, así mismo la arquitectura propuesta permite un alto grado de acoplamiento y escalabilidad. Además, dada la arquitectura diseñada para la creación de portlets e informes, es relativamente sencilla su extensión así como la creación y publicación de nuevos informes, utilizando las herramientas previstas para tal fin. En la intercomunicación entre portlets para su actualización ante cualquier cambio de uno de ellos, MERLÍN es uno de los primeros proyectos que implementa la nueva especificación de Portlets 2.0 JSR-286. [13]

Entre los trabajos futuros y mejoras sobre la implementación inicial, cabe destacar la necesidad prevista de creación y mantenimiento de un datawarehouse específico para almacenar los datos de seguimiento. Con ello se conseguiría una reducción drástica de los tiempos de ejecución de las consultas necesarias para mantener los procesos de negocio, lo que supondría un mayor rendimiento del portal y una medida de prevención frente a posibles bloqueos. Para la implementación de esta solución, se recomienda la utilización de herramientas ETL (Extracción, Carga y Transformación), que alimenten al datawarehouse con la información generada en la capa de datos de la aplicación. La tecnología que soporta los procesos de negocio definidos en MERLÍN, Pentaho, ofrece una potente herramienta dirigida a la realización de operaciones de este tipo, el proyecto Pentaho Data Integration-Kettle.

AGRADECIMIENTOS

Este trabajo de investigación ha sido cofinanciado por el Ministerio de Industria, Comercio y Turismo a través del programa PROFIT MERLÍN. *Plataforma de Aprendizaje a Distancia basada en tecnología de portlets y web2.0 para una enseñanza participativa* FIT-360000-2007-23. Los autores quieren agradecer la colaboración del resto de socios de MERLÍN por sus contribuciones y sugerencias y, muy especialmente, al coordinador, David Jiménez de Idea

Informática, por su tenacidad y diligencia en la coordinación para la consecución de los objetivos del proyecto.

REFERENCIAS

- [1] (2008) Página web oficial de moodle. [Online]. Available: <http://moodle.org>
- [2] (2008) Página web oficial de dokeos. [Online]. Available: <http://www.dokeos.com/es/>
- [3] (2003) Especificación portlets 1.0 JSR-168. [Online]. Available: <http://jcp.org/en/jsr/detail?id=168>
- [4] E. Thomsen, *OLAP solutions: building multidimensional information systems*. New York, NY, USA: John Wiley & Sons, Inc., 1997.
- [5] (2008) Web oficial de merlín. [Online]. Available: <http://merlin.germinus.com>
- [6] (2008) Web oficial de liferay. [Online]. Available: <http://www.liferay.com>
- [7] (2008) Página web oficial de sakai. [Online]. Available: <http://sakaiproject.org>
- [8] (2008) Página web oficial de blackboard. [Online]. Available: <http://www.blackboard.com/>
- [9] (2008) Página web oficial de WebCT. [Online]. Available: <http://www.webct.com>
- [10] (2008) Página web oficial de atutor. [Online]. Available: <http://www.atutor.ca>
- [11] (2008) Web oficial de pentaho. [Online]. Available: <http://www.pentaho.com>
- [12] M. L. E.Vitt and S. Miner, *Business Intelligence*. Microsoft Press, 2002.
- [13] (2008) Especificación portlets 2.0 JSR-286. [Online]. Available: <http://jcp.org/en/jsr/detail?id=286>
- [14] I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed. Morgan Kaufmann, 2005.
- [15] (2008) Documentación del proyecto mondrian. [Online]. Available: <http://mondrian.pentaho.org/documentation/doc.php>
- [16] (2008) Documentación del proyecto weka. [Online]. Available: <http://www.cs.waikato.ac.nz/ml/weka/>