Minería de Datos en Educación y Análisis del Aprendizaje

Cristóbal Romero Morales
(cromero@uco.es)

Departamento de Informática y Análisis Numérico.
Grupo de Investigación KDIS.
Universidad de Córdoba

UNIVERSIDAD D CÓRDOBA

KDIS Lab
KNOWLEDGE DISCOVERY AND INTELLIGENT SYSTEMS

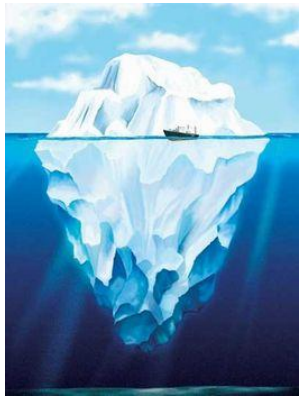
Contenido

* Introducción
* Tipos de Datos
* Pre-procesado de Datos
* Principales Tareas y Aplicaciones
* Técnicas de DM empleadas
* Software específico
* Líneas Futuras
* Publicaciones
* Enlaces Relacionados

Introducción

Introducción

- El desarrollo de **sistemas de enseñanza basada en web** se ha incrementado exponencialmente en los últimos años: Moodle, Ilias, WebCT, Atutor, etc.
- Estos sistemas almacenan una información de la interacción con los estudiantes que no se suele utilizar.Pero debido al **gran cantidad de datos de utilización** se necesitan herramientas para facilitar el descubrimiento de información, y **no sólo** utilizarlos **para el seguimiento y la evaluación** de los alumnos.



- Las **técnicas de minería de datos** se han aplicado con éxito en los sistemas de **e-commerce**, donde el objetivo es maximizar la compra de los clientes. Se están comenzando a utilizar en sistemas de **e-learning**, para maximizar el aprendizaje de los estudiantes.
- Aunque existen **sistemas genéricos de minería de datos**: Weka, Clementine, DBMiner, etc. se necesita **herramientas específicas** en educación debido a la especifidad de los usuarios y los objetivos.

Introducción

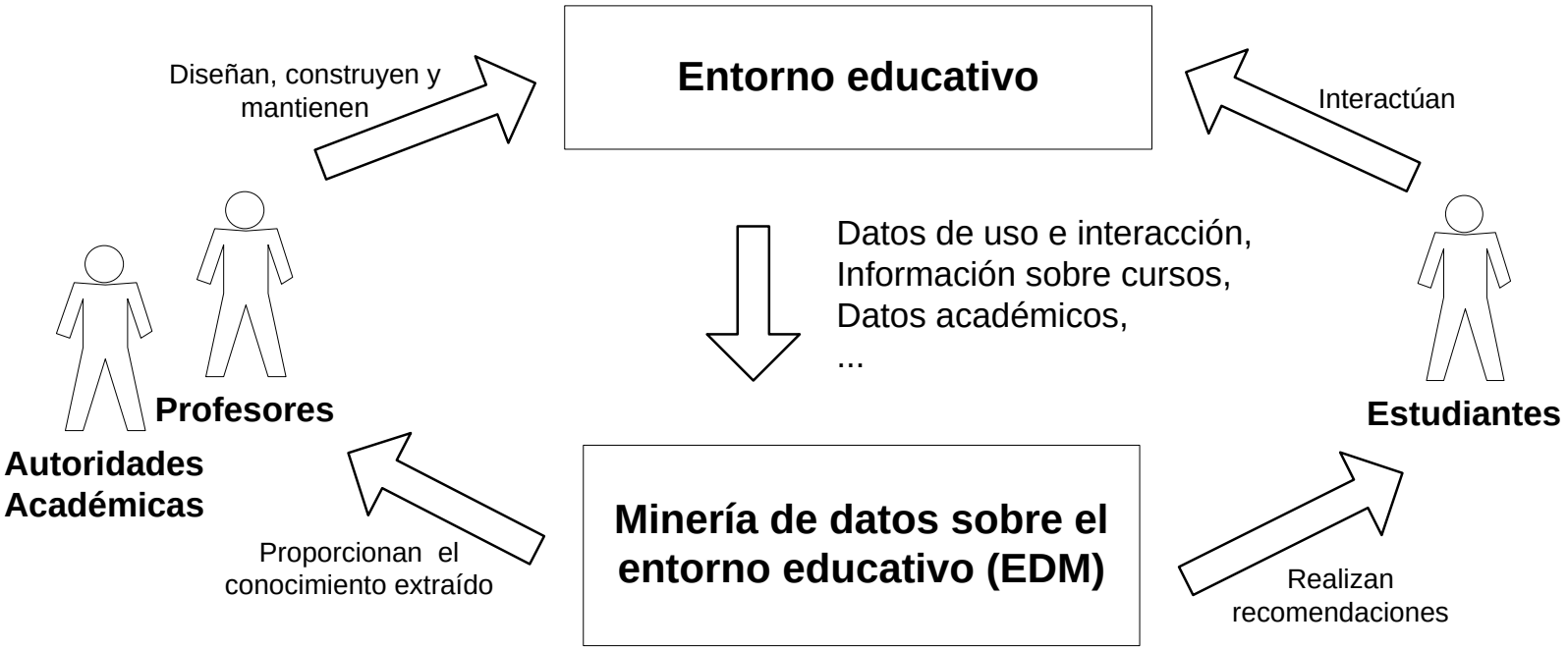
- La **Minería de Datos** es el proceso de descubrimiento de conocimiento para encontrar información no trivial, desconocida y potencialmente útil de grandes repositorios de datos.
- La Minería de Datos es uno de los pasos que componen el Proceso de **Descubrimiento de Conocimiento en Bases de Datos** o **KDD** (Knowledge Discovery in DataBases):

- La **Minería de datos Web** (Web Mining) consiste en la aplicación de técnicas de minería de datos para extraer conocimiento a partir de datos de la Web.
  - Minería de Contenidos Web (Web Content Mining)
  - Minería de Estructura Web (Web Structure Mining)
  - Minería de Utilización Web (Web Usage Mining)

Introducción

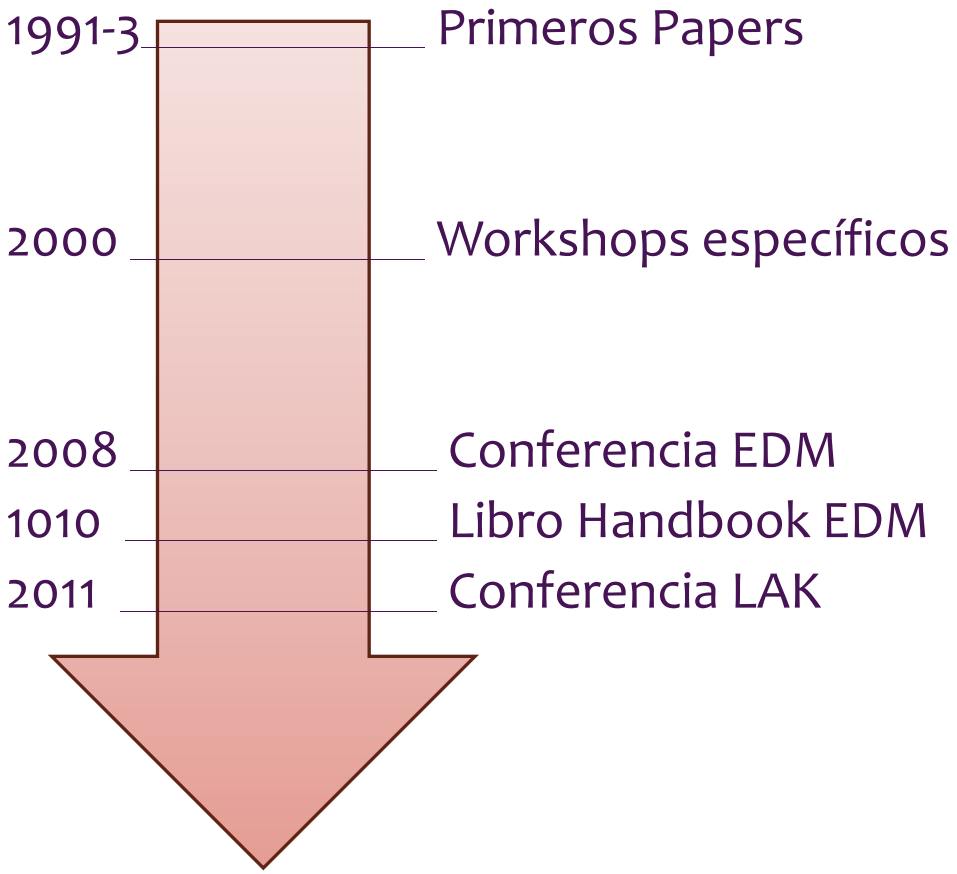
Minería de Datos

* La **Minería de Datos Educativa** (***educational data mining, EDM***) es la aplicación de técnicas de minería de datos a información generada en los entornos educativos.



Introducción

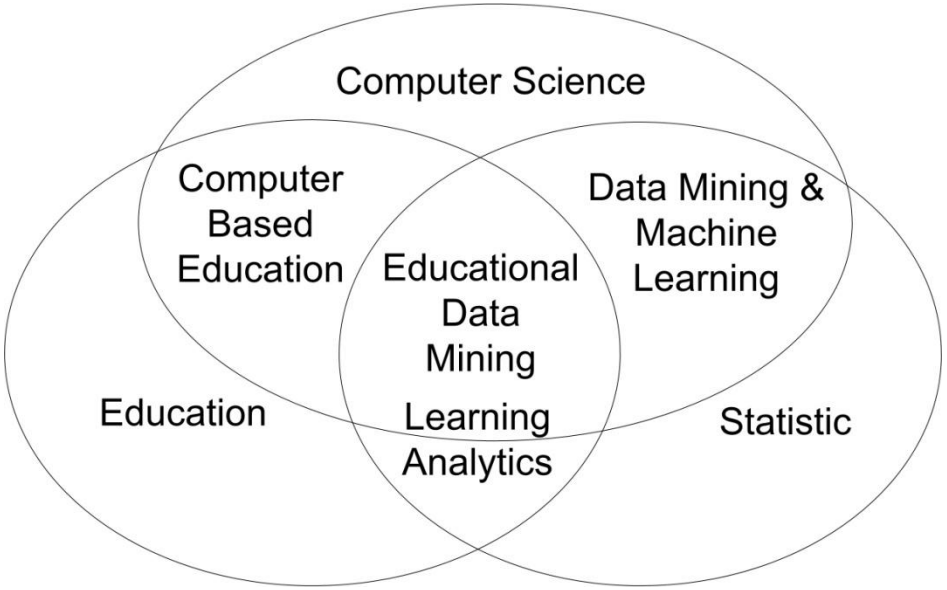
Orígenes



Introducción

Áreas relacionadas

* EDM surge de la combinación/integración de varias áreas: computación, educación y estadística.



* Un área más reciente y relacionada es **LAK** (**Learning and Analytics Knowledge**).

Introducción

Diferencias entre EDM y DM

* **Objetivos.** El objetivo en cada área de aplicación de DM es diferente. En EDM el objetivo principal es mejorar el proceso de aprendizaje del estudiante que es una tarea difícil de realizar y de cuantificar.
* **Datos.** En EDM hay muchos tipos de datos diferentes y específicos con información semántica intrínseca, relaciones con otros datos y múltiples nivel de significado jerárquico.
* **Técnicas.** Los problemas específicos que intenta resolver EDM hace necesario que se tengan que adaptar las técnicas de DM existentes tanto a los datos como al problema.

Introducción

Diferencias entre EDM y LAK

* **Técnicas.** LAK utiliza: social network analysis, sentiment analysis, influence analytics, discourse analysis, learner success prediction, concept analysis, sense making models. EDM, utiliza: visualization, classification, clustering, bayesian modeling, relationship mining and discovery with models.
* **Orígenes**. LAK proviene de semantic web, intelligent curriculum y systemic interventions. EDM del educational software, student modeling y predicting course outcomes.
* **Énfasis.** LAK da más énfasis a la descripción de los datos y aplicación de los resultados. EDM a las técnicas de DM.

Introducción

Analytics, Learning Analytics y Academic Analytics

* **Analytics.** Es el descubrimiento y comunicación de patrones significantes en datos, para la toma de decisiones dirigida por datos.
* **Learning analytics.** Es la medida, recoleción, análisis e informe de datos sobre estudiantes/aprendizes y su contexto, con el propósito de comprender y optimizar el aprendizaje y el entorno donde ocurre.
* **Academic analytics.** Es la aplicación de técnicas estadísticas y de minería a datos institucionales para producir inteligencia de empresa y soluciones a universidades y administradores.

Introducción
Objetivos del EDM

- El conocimiento que puede extraerse de los sistemas educativos es muy diverso.
- El objetivo que nos marcamos al intentar aplicar técnicas de EDM depende de:
  - Entorno en el que nos situamos
    - Enseñanza presencial
    - Enseñanza a distancia
  - A quién va dirigido el conocimiento que extraigamos
    - Alumnos
    - Profesores
    - Investigador
    - Autoridades académicas

Introducción

Objetivos del EDM – Punto de vista del estudiante

- Recomendar qué actividades, recursos y tareas podrían mejorar su rendimiento académico.
- Recomendar qué actividades se ajustan mejor al perfil de un determinado alumno.
- Recomendar qué camino recorrer para obtener un resultado concreto:
  - Basándonos en conocimiento del camino ya recorrido por el alumno y su éxito.
  - Por comparación con lo realizado por otros alumnos de características análogas.

Introducción

Objetivos del EDM – Punto de vista del profesor

- Cuantificar la efectividad del proceso de enseñanza-aprendizaje
- Organizar los contenidos de un curso
- Mejorar o corregir la estructura del curso
- Clasificar o agrupar alumnos en base a sus características
  - Tutorización
  - Asesoramiento
  - De cara a monitorizar conocimiento interesante
- Buscar patrones de comportamiento en alumnos
  - Patrones generales
  - Patrones anómalos
- Evaluar las actividades realizadas en un curso
  - Efectividad
  - Motivación
- Monitorizar actividades:
  - Errores más frecuentes en la realización de actividades
  - Grado de dificultad de una actividad
- Personalizar y adaptar el contenido de cursos
  - Diseñar planes de instrucción

Introducción

Objetivos del EDM – Punto de vista del investigador

* Desarrollar Herramientas, Entornos y Métodos específicos para EDM.
* Desarrollar/adaptar algoritmos para minar datos educacionales procedentes de: evaluaciones, navegación, interacción, etc.
* Realizar estudios de replicación. Aplicar técnicas anteriormente utilizadas en otros dominios.
* Aplicar minería de procesos educacional. Extraer conocimiento relacionada con el proceso a partir de eventos logs registrados.
* Integrar DM con teoría pedagógica. Utilizar conocimiento tanto educacional como psicológico para mejorar la búsqueda en DM.
* Evaluar las intervenciones del profesor y/o sistema. Determinar que acciones del profesor o del sistema son más beneficiosas.
* Personalización y Adaptación dirigida por datos. Utilizar DM para mejorar la personalización y adaptación del alumno.

Introducción

Objetivos del EDM – Punto de vista de las instituciones educativas

- Analizar y evaluar los comportamientos de los profesores
  - Detectar buenos y malos profesores.
  - Recomendar acciones en base a otros profesores.
- Mejorar la organización de los recursos institucionales
  - Diseño de horarios
  - Adquisición de material
- Mejora de la oferta educativa
  - Programas orientados a demanda
  - Orientación de alumnos en base a
    - Objetivos
    - Capacidades

Tipos de Datos

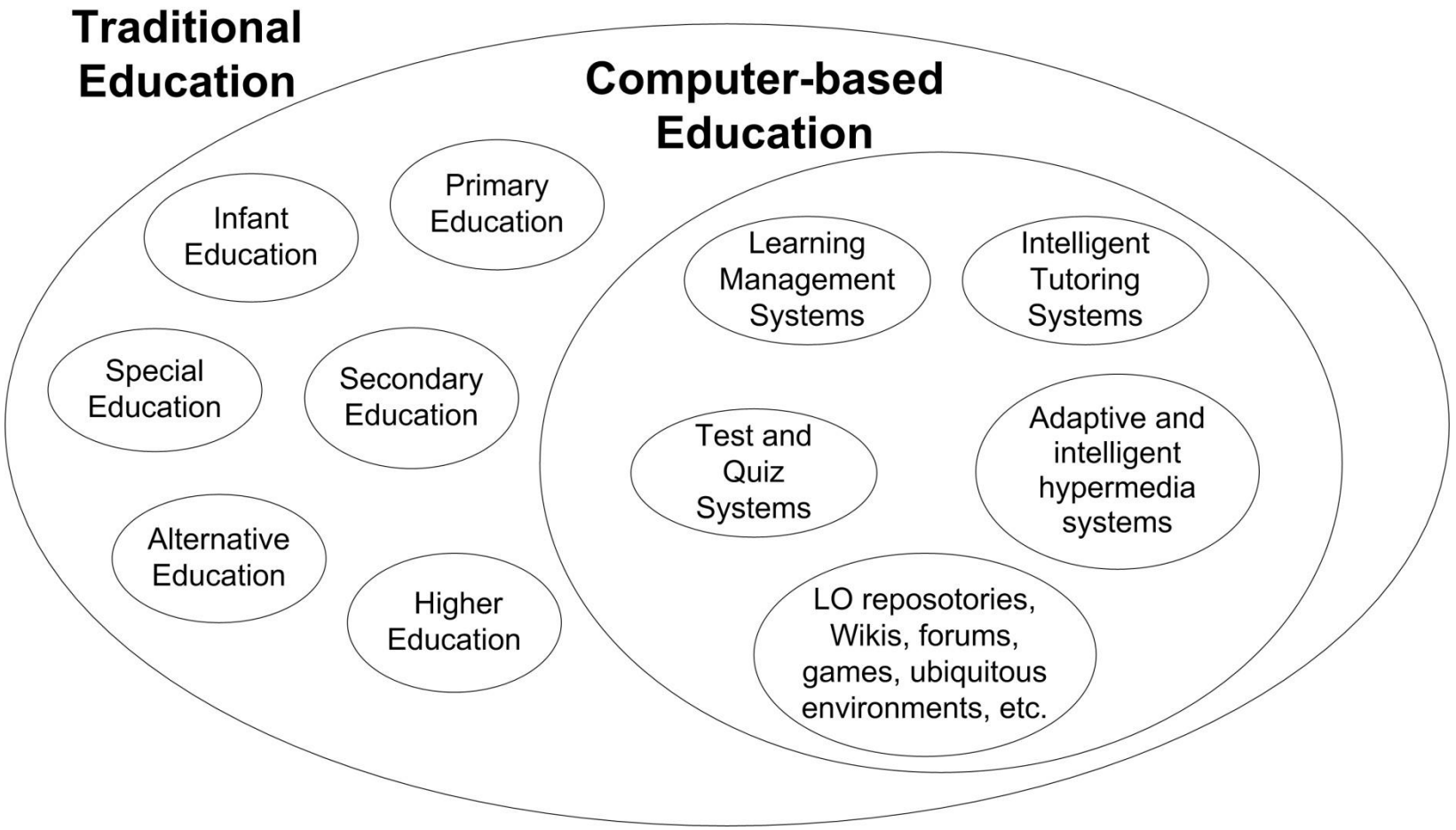
Tipos de Datos

Tipos de Sistemas Educativos

- **Sistemas Educativos Tradicionales.** Enseñanza presencial, cara a cara. (primaria, secundaria, superior, especial, etc.) Recomendación de matriculación en cursos, detectar posibles fracasos, etc.
- **Sistemas de enseñanza a distancia basados en web.**
  - **Cursos propios y foros educativos en web.** Recomendar material, noticias, etc..
  - **Sistemas de Manejo de Cursos (CMS), Aprendizaje (LMS) y MOOC.** Realizar sugerencias y reestructuración del curso.
  - **Sistemas Adaptativos y/o Inteligentes Adaptativos para Educación basada en Web (AHS/ITS).** Mejorar adaptación y personalización.
  - **Sistemas de Test (Quiz systems).** Mejorar y personalizar los test.
  - **Otros sistemas.** Repositorios LO, Wikis, juegos, 3D, móviles, etc.

Tipos de Datos

Tipos de Sistemas Educativos



Tipos de Datos

* Existen multitud de sistemas o entornos educativos diferentes.
* Cada tipo de sistema educacional suele proporcionar datos diferentes.
* Cada tipo de dato permiten resolver problemas y tareas diferentes.
* Algunos tipos de datos típicos son: transaccional, relacional, secuencial, textual, multimedia, etc.

Tipos de Datos

Datos Relacionales

* Las bases de datos relacionales son muy utilizadas en entornos educativos.
* Consiste en una coleción de tablas formada por un conjunto de atributos (columnas o campos) que almacenan un conjunto de tuplas (filas o registros).
* Cada tupla en una base de datos relacional representa un objeto identificado por una clave única y descrita por un conjunto de valores de sus atributos.
* Los datos relacionales se pueden acceder mediante consultas a la base de datos escritas en un lenguje de consulta relacional como SQL (Structured Query Language), o mediante la ayuda de un interfaz gráfico.

Tipos de Datos

Datos Relacionales

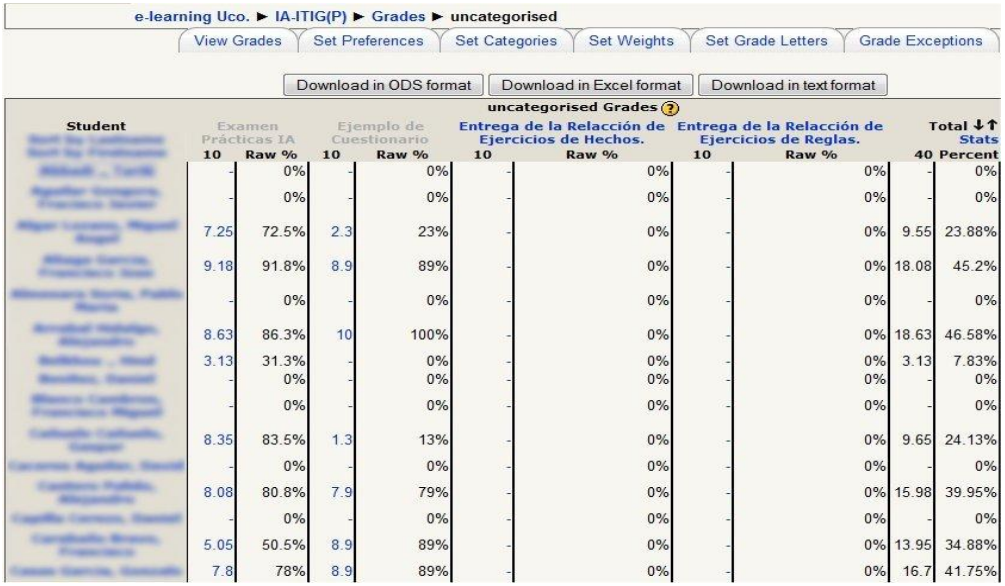
* Por ejemplo, Moodle utiliza una base de datos relacional con más de 200 tablas (todas comienza con *mdl_* seguidas de una palabra descriptiva). Algunas de las más importantes son:

Name	Description
mdl_user	Information about all the users.
mdl_user_students	Information about all students.
mdl_log	Logs every user's action.
mdl_assignement	Information about each assignment.
mdl_assignment_submissions	Information about assignments submitted.
mdl_forum	Information about all forums.
mdl_forum_posts	Stores all posts to the forums.
mdl_forum_discussions	Stores all forum discussions.
mdl_message	Stores all the current messages.
mdl_ message_reads	Stores all the read messages.
mdl_quiz	Information about all quizzes.
mdl_quiz_attempts	Stores various attempts at a quiz.
mdl_quiz_grades	Stores the final quiz grade.

Tipos de Datos

Datos Transaccionales

- Un dataset transaccional consiste en un fichero o tabla donde cada registro representa una transacción.
- Una transacción normalmente incluye un identificador y una lista de elementos que forman la transacción.
- Por ejemplo, Moodle proporciona alguna info de este tipo.



e-learning Uco. ► IA-ITIG(P) ► Grades ► uncategorised

View Grades | Set Preferences | Set Categories | Set Weights | Set Grade Letters | Grade Exceptions

Download in ODS format | Download in Excel format | Download in text format

uncategorised Grades

Student	Examen Prácticas IA		Ejemplo de Cuestionario		Entrega de la Relacción de Ejercicios de Hechos.		Entrega de la Relacción de Ejercicios de Reglas.		Total ↓↑ Stats	
	10	Raw %	10	Raw %	10	Raw %	10	Raw %	40	Percent
[redacted]	-	0%	-	0%	-	0%	-	0%	-	0%
[redacted]	-	0%	-	0%	-	0%	-	0%	-	0%
[redacted]	7.25	72.5%	2.3	23%	-	0%	-	0%	9.55	23.88%
[redacted]	9.18	91.8%	8.9	89%	-	0%	-	0%	18.08	45.2%
[redacted]	-	0%	-	0%	-	0%	-	0%	-	0%
[redacted]	8.63	86.3%	10	100%	-	0%	-	0%	18.63	46.58%
[redacted]	3.13	31.3%	-	0%	-	0%	-	0%	3.13	7.83%
[redacted]	-	0%	-	0%	-	0%	-	0%	-	0%
[redacted]	-	0%	-	0%	-	0%	-	0%	-	0%
[redacted]	8.35	83.5%	1.3	13%	-	0%	-	0%	9.65	24.13%
[redacted]	-	0%	-	0%	-	0%	-	0%	-	0%
[redacted]	8.08	80.8%	7.9	79%	-	0%	-	0%	15.98	39.95%
[redacted]	-	0%	-	0%	-	0%	-	0%	-	0%
[redacted]	5.05	50.5%	8.9	89%	-	0%	-	0%	13.95	34.88%
[redacted]	7.8	78%	8.9	89%	-	0%	-	0%	16.7	41.75%

Tipos de Datos

Datos Temporales y Secuenciales

* Los datos de series temporales y secuenciales constan de secuencias de valores o eventos que cambian con el tiempo.
* Una base de datos temporal, normalmente almacena datos relacionales que incluyen atributos relacionados con el tiempo.
* Una base de datos secuencial almacena secuencias de eventos ordenador con o sin una concreta noción del tiempo.

Tipos de Datos

Datos Temporales y Secuenciales

* Por ejemplo, Moodle proporciona la tabla Log

Time	IP Address	Full name	Action	Information
Fri 15 January 2010, 12:49 PM	150.214.110.166	[redacted]	forum view forum	Foro de discusión sobre los Ejercicios de Reglas
Fri 15 January 2010, 12:05 PM	150.214.110.166	[redacted]	resource view	Ejercicios resueltos de Reglas
Fri 15 January 2010, 12:05 PM	150.214.110.166	[redacted]	resource view	Relacción de Ejercicios Reglas
Fri 15 January 2010, 12:05 PM	150.214.110.166	[redacted]	course view	Prácticas IA
Thu 14 January 2010, 07:43 PM	150.214.110.166	[redacted]	resource view	Ejercicios resueltos de Reglas
Thu 14 January 2010, 07:38 PM	150.214.110.166	[redacted]	resource view	Relacción de Ejercicios Reglas
Thu 14 January 2010, 07:38 PM	150.214.110.166	[redacted]	resource view	Relacción de Ejercicios Hechos
Thu 14 January 2010, 07:38 PM	150.214.110.166	[redacted]	assignment view	Entrega de la Relacción de Ejercicios de Hechos.
Thu 14 January 2010, 07:38 PM	150.214.110.166	[redacted]	upload upload	Relacion_Hechos.rar
Thu 14 January 2010, 07:38 PM	150.214.110.166	[redacted]	assignment upload	Entrega de la Relacción de Ejercicios de Hechos.
Thu 14 January 2010, 07:02 PM	150.214.110.166	[redacted]	assignment view	Entrega de la Relacción de Ejercicios de Reglas.
Thu 14 January 2010, 07:01 PM	150.214.110.166	[redacted]	assignment view	Entrega de la Relacción de Ejercicios de Hechos.
Thu 14 January 2010, 07:01 PM	150.214.110.166	[redacted]	assignment view all	
Thu 14 January 2010, 07:01 PM	150.214.110.166	[redacted]	assignment view	Entrega de la Relacción de Ejercicios de Hechos.
Thu 14 January 2010, 06:10 PM	150.214.110.166	[redacted]	resource view	Introducción a CLIPS
Thu 14 January 2010, 06:09 PM	150.214.110.166	[redacted]	resource view	Hechos
Thu 14 January 2010, 06:09 PM	150.214.110.166	[redacted]	course view	Prácticas IA
Thu 14 January 2010, 05:56 PM	150.214.110.166	[redacted]	forum view discussion	duda ejercicio 8
Thu 14 January 2010, 05:56 PM	150.214.110.166	[redacted]	forum view discussion	duda en ejercicio 8

Tipos de Datos

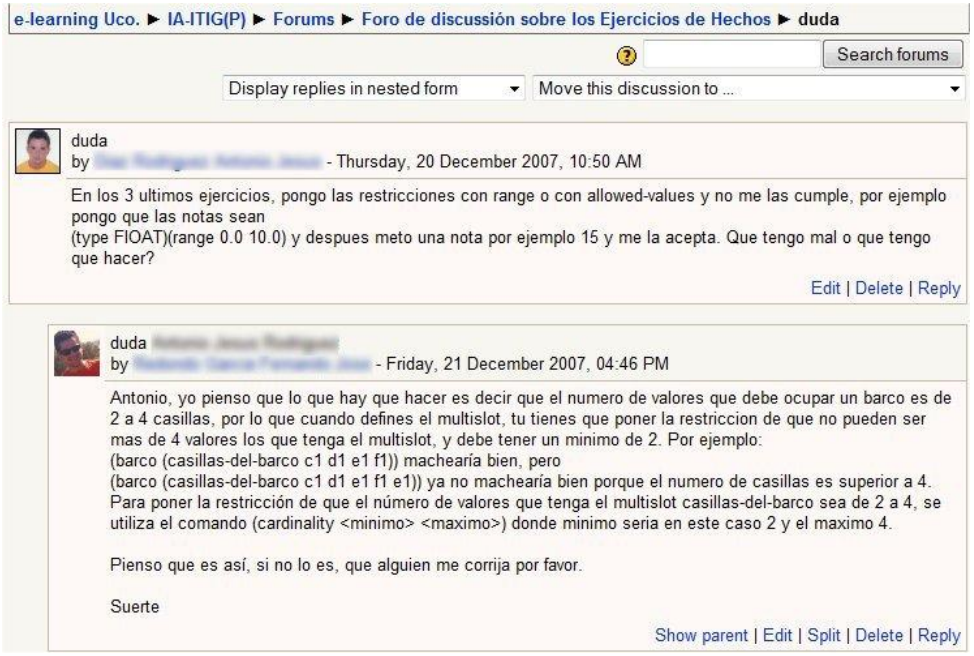
Datos de Texto

* Las bases de datos de texto (o base de datos de documentos de texto) consisten en coleciones de documentos de varias fuentes, como artículos de noticias, papers de investigación, libros, librerías digitales, mensajes de correo electrónico, mensajes de chats y foros, etc.
* Los textos de las bases de datos de texto puede estar poco estructuradas (contenidos de páginas Web) o bien estructuradas (páginas XML.)

Tipos de Datos

Datos de Texto

* Por ejemplo, Moodle proporciona mucha información textual como documentos de texto y web, mensajes a foros y chats, etc.



Tipos de Datos

Datos Multimedia

* Las bases de datos Multimedia almacenan datos de texto, video, audio e imágenes.
* Por ejemplo, Moodle almacena todos los datos subidos tanto por los profesores como por los propios alumnos.

Tipos de Datos

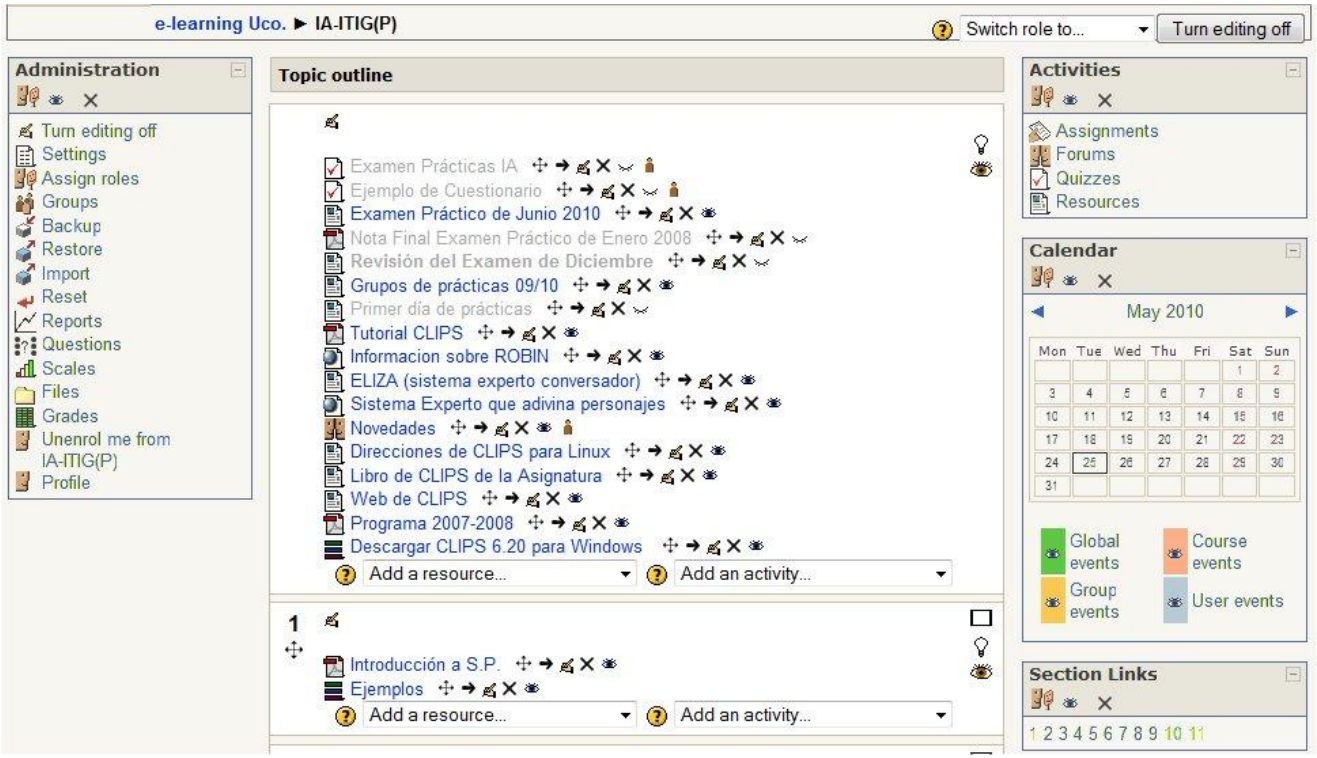
Datos Web

- La World Wide Web (WWW) proporciona tres tipos principales de fuentes de datos:
  - **Contenido de las páginas web**: Es el contenido de las propias páginas, tanto en formato HTML, XML, ect. como los ficheros incluidos de distintos tipos: sonido, video, etc.
  - **Estructura entre páginas o Hiper-links:** Datos que describen la organización del contenido o enlaces entre diferentes página o dentro de una misma página.
  - **Utilización de las páginas. User usage data:** Datos que describen el uso o patrones de navegación que realiza los usuarios al utilizar la Web.

Tipos de Datos

Datos Web

* Moodle al ser un sistema Web, tiene los mismo tipos de fuentes de datos que otros sistemas Web.



Tipos de Datos

Otros Tipos de Datos

Existen otros tipos diferentes de tipos de datos como:

* Datos objecto-relationales data.
* Datos espaciales.
* Datos espaciontemporales.
* Datos heterogeneos.
* etc

Pre-procesado de Datos

Pre-procesado de Datos

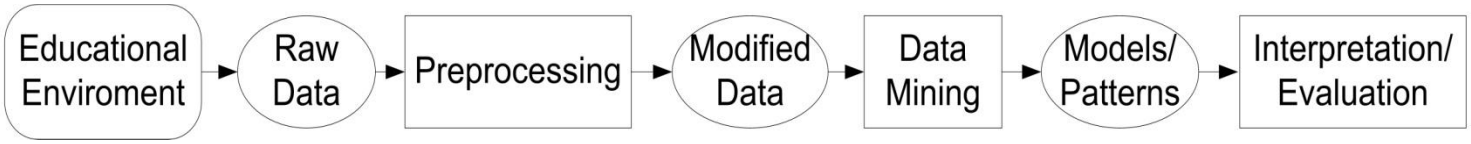
Características específicas

* La gran cantidad de información generada suele provenir de diferentes fuentes de información.
* Existe muchos datos incompletos y perdidos, ya que no todos los estudiantes realizan todas las actividades.
* No es necesaria realizar una identificación de usuarios y sesiones.
* Hay un gran número de instancias y atributos disponibles de los alumnos que suele requerir de tareas de filtrado y selección de atributos, para seleccionar los más importantes.
* Los datos educacionales suelen tener diferentes niveles de granularidad.
* Transformaciones en los datos como la discretización son muy utilizadas para aumentar la comprensibilidad de los datos.

Pre-procesado de Datos

Introducción

* El pre-procesado de datos es la primera etapa del **proceso de minería de datos o KDD :**



* Las principales **tareas/etapas del pre-procesado de datos** son:

Pre-procesado de Datos

Introducción

No todas las tareas de pre-procesado hay que aplicarlas siempre, sino que depende de los datos concretos:

* Agregación/Integración sólo si hay diferentes fuentes.
* Limpieza sólo si hay datos erróneos, perdidos o incompletos.
* Identificación de usuario y sesión sólo si no se dispone de esta información.
* Filtrado de datos y selección de atributos sólo si hay una gran cantidad de datos y atributos, respectivamente.

Pre-procesado de Datos

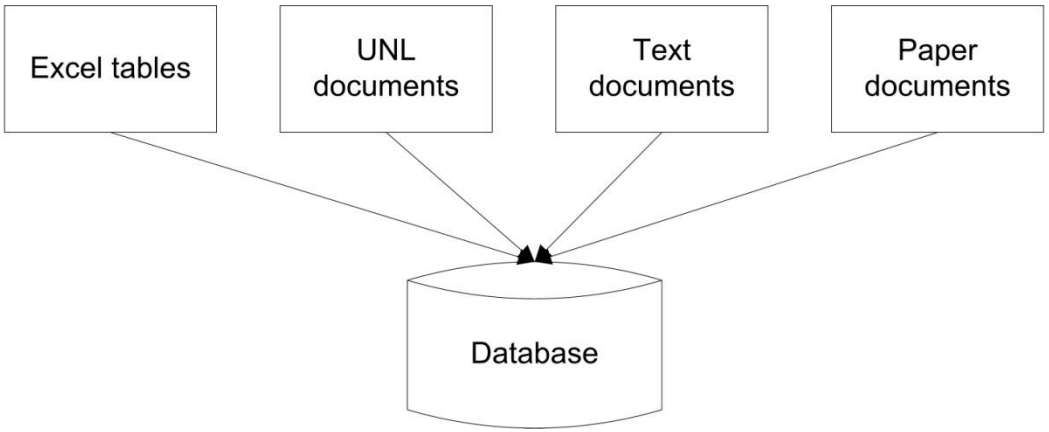
Recogida de Datos

* Recogida de datos consiste en localizar todos los datos importantes para nuestro problema. Algunos términos relacionados con la recogida y almacenamiento de datos son data warehousing, data smart, repositorio central de datos, etc.
* Lo sistemas educacionales suelen recoger datos de varias fuentes, debido a que son generados en diferentes lugares y en diferentes momentos:
  * **Datos de perfil** (profile). Datos administrativos que contienen información personal sobre los alumnos y los profesores.
  * **Datos de clase**. Asistencia a clase, a prácticas, tutorías, participación en clase, notas de prácticas, exámenes, etc.
  * **Datos de e-learning**. Datos de uso e interacción con los recursos, comunicación entre alumnos, datos de actividades realizadas, etc.
  * **Otros datos**: Notas en otros cursos o cursos anteriores, relaciones en redes sociales, etc.

Pre-procesado de Datos

Agregación e Integración de Datos

* El objetivo es agregar/integrar todo la información proveniente de diferentes fuentes en una única recopilación coherente (normalmente una base de datos).
* Los sistemas educacionales proporcionan diferentes fuentes, normalmente con diferentes formatos, por ejemplo:

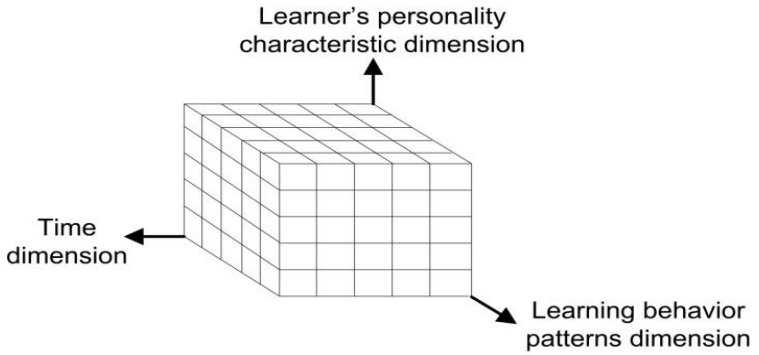


Pre-procesado de Datos

Agregación e Integración de Datos

Los sistemas más utilizados son:

- **Base de datos relacionales**, donde la información esta categorizada en tablas que se pueden acceder mediente el lenguaje SQL. Por ejemplo Moodle la utiliza.
- **Almacén de datos** (datawarehouse) específicamente estructurado para la consulta y el análisis de datos, y herramientas para extraer, transformar y cargar datos. Un tipo son los cubos de Información o cubo OLAP (*OnLine Analytical Processing*). Es un cubo con un número de dimensiones (atributos relativos a las variables).



Pre-procesado de Datos

Agregación e Integración de Datos

* **Tabla o fichero sumario.** Recoge un resumen de toda la información de los estudiantes. Un ejemplo de resumen de actividad en Moodle :

Name	Description
id_student	Identification number of the student.
id_course	Identification number of the course.
num_sessions	Number of sessions.
num_assigment	Number of assignments done.
num_quiz	Number of quizzes taken.
a_scr_quiz	Average score on quizzes
num_posts	Number of messages sent to the forum.
num_read	Number of messages read on the forum.
t_time	Total time used on Moodle.
t_assignment	Total time used on assignments.
t_quiz	Total time used on quizzes.
t_forum	Total time used on forum.
f_scr_course	Final score the student obtained in the course.

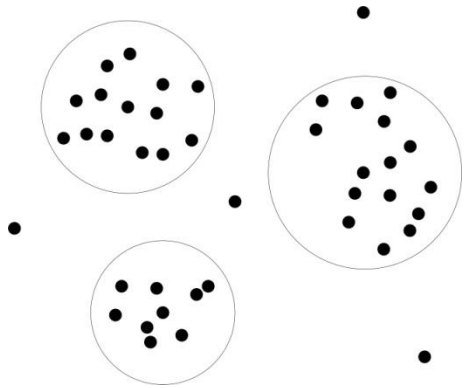
La tabla generada se guarda en un fichero que puede tener formato: .txt, .csv, .xls, .arff, .dat, .xml, etc.

Pre-procesado de Datos

Limpieza de Datos

La limpieza de datos consiste en detectar diferentes tipos de inexactitudes en los datos como: datos perdidos , ruidosos e inconsistentes.

* Los **valores perdidos** ocurren cuando no hay valores para una variable. Una solución puede ser descartar la instancia completa o por el contrario utilizar etiquetas como ? O NULL.
* Los **valores inconsistentes** o **ruidosos** son los que se diferencian mucho del resto de datos sin una razón aparente. Algunos motivos pueden ser errores en la introducción de datos. Una forma de detectarlos son utilizar técnicas de **outliers:**



Pre-procesado de Datos

Identificación de Usuario y Sesión

* La **identificación de usuario** consiste en identificar a cada usuario, a través de la IP, utilizando cookies, ID, etc.
* La **identificación de sesión** consiste en determinar el periodo de actividad desde que el usuario se conecta y desconecta.
* En los sistemas educacionales basados en Web no son necesarias, ya que se realizan automáticamente. Aunque si puede ser interesante identificar episodios concretos de interacción, realización de tareas o actividades.
* Es importante prevenir la **privacidad** y **anonimato** de los usuarios. Para ello se suele reemplazar el nombre por un ID numérico.

Pre-procesado de Datos

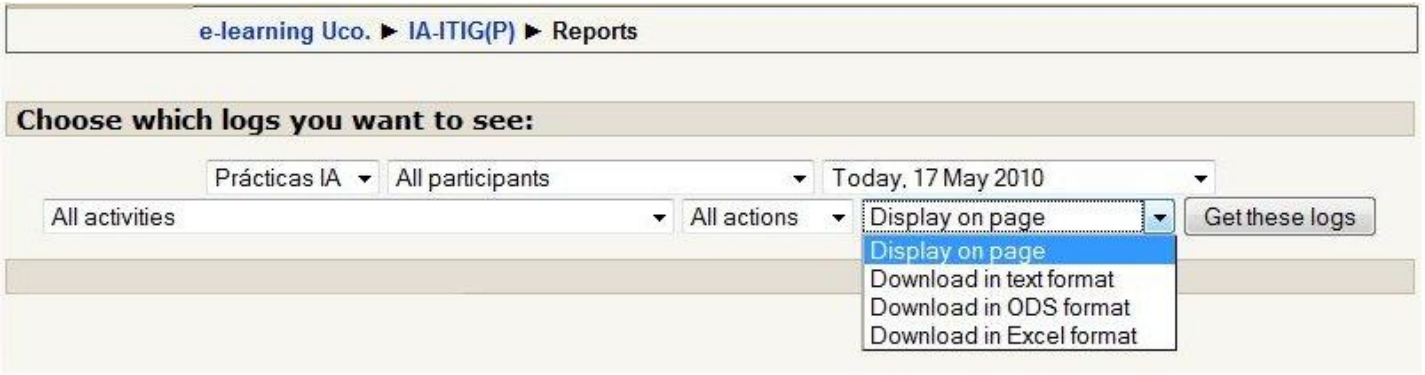
Selección de Atributos

* La **selección** o **reducción de atributos** o **características** o variables, consiste en seleccionar un subconjunto relevante de atributos (columnas) de entre todos los disponibles.
* En los sistemas educacionales suele haber muchos atributos, y algunos pueden ser irrelevantes, redundantes, o estar correlados.
  * Ejemplos de atributos irrelevantes en educacion son: password, student's e-mail, student's phone number, student's address, student's picture, etc.
  * Existen muchas técnicas para la selección de los mejores atributos y determinación de atributos correlados y los irrelevantes.

Pre-procesado de Datos

Filtrado de Datos

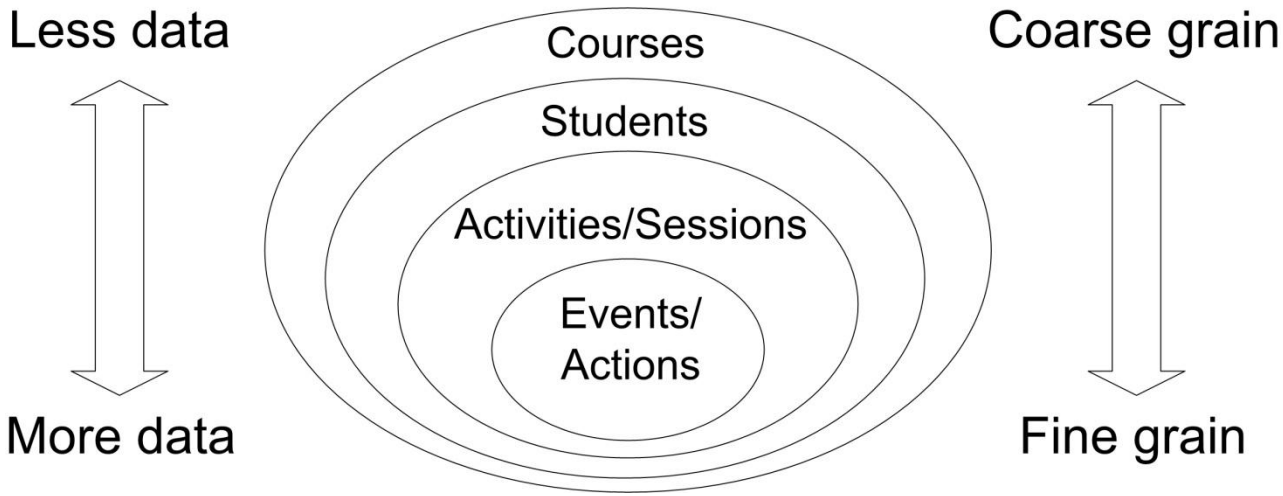
* El **filtrado de datos** o **selección de instancias (filas)** consiste en seleccionar un subconjunto representativo de los datos para convertir grandes datasets en datasets más manejables.
* En educación un tipo de filtrado muy utilizado es seleccionar subconjuntos de datos referentes a una tarea, evento o actividad concreta. Por ejemplo, Moodle permite filtrar la información de los logs:



Pre-procesado de Datos

Filtrado de Datos

* Otro tipo de filtrado típico en educación es utilizar **diferentes niveles de granularidad**: keystroke level, answer level, session level, student level, classroom level, and school level:



Pre-procesado de Datos

Transformación de Datos

La transformación de datos deriva nuevos atributos a partir de los atributos existentes, mediante técnicas de normalización, discretización y derivación.

La **normalización** transforma los valores de los atributos escalándolos dentro de un rango específico como [-1..1] o [0..1]. En educación el método más utilizado para normalizar es el min-max que mapea el valor, *v* a *v'* en el rango [*new minA*, *new maxA*]:

$$v' = \frac{v - \min_A}{\max_A - \min_A}(new\_\max_A - new\_\min_A) + new\_\min_A$$

Para normalizar en el rango [0.0, 1.0] sería: $v' = \frac{v - \min_A}{\max_A - \min_A}$

Pre-procesado de Datos

Transformación de Datos

La **discretización** de datos transforma datos numéricos en categóricos.

* En educación normalmente el uso de etiquetas proporciona una mayor comprensibilidad en los datos.
* Los métodos más utilizados son: particionado igual en anchura (equal-width) , igual en frecuencia (equal-frequency) y manual.
* Un ejemplo de particionado manual de notas es:
  * *FAIL*: si la nota es < 5
  * *PASS*: si la nota es ≥ 5 y < 7
  * *GOOD:* si la nota es ≥ 7 y < 9
  * *EXCELLENT*: si la nota es ≥ 9

Pre-procesado de Datos

Transformación de Datos

La **derivación** crea nuevos atributos a partir de los actuales.

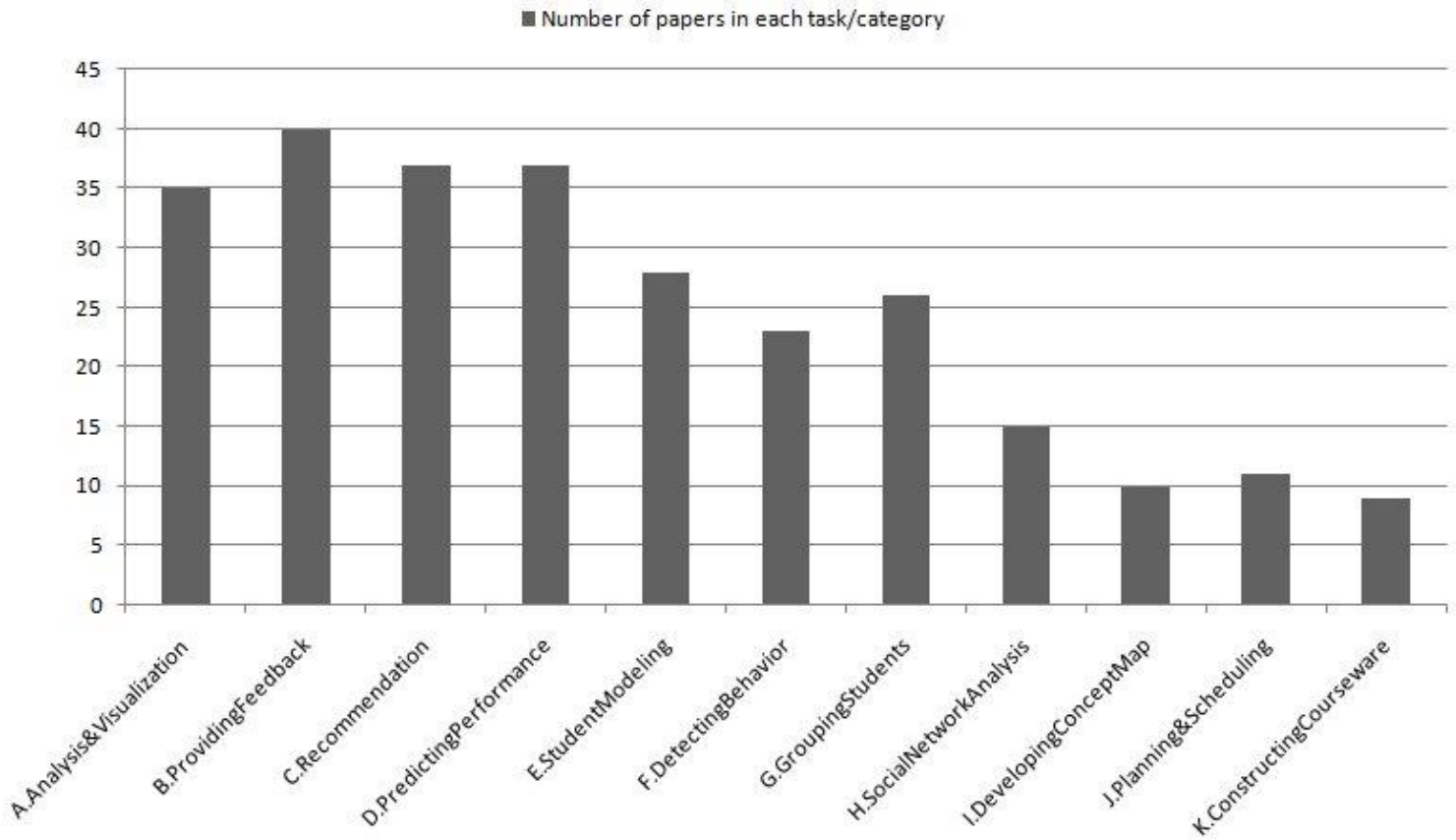
* Un nuevo atributo se puede obtener a partir de una transformación matemática. Por ejemplo, el atributo tiempo se puede pasar de segundos a minutos, horas o días.
* Un nuevo atributo se puede crear a partir de varios atributos. Por ejemplo la edad en años se puede obtener del día, mes y año.
* Otros ejemplos:

Attribute	Description
UserId	A unique identifier per user.
Performance	Percentage of correctly answered tests calculated as the number of correct tests divided by the total number of tests performed).
TimeReading	Time spent on pages (calculated as the total time spent on each page accessed) in a session.
NoPages	The number of accessed pages.
TimeTests	The time spent performing tests (calculated as the total time spent on each test).
Motivation	Engaged / Disengaged.

Principales Tareas y Aplicaciones de EDM

Tareas y Aplicaciones

Distribución del número de papers por tarea de aplicación



Tareas y Aplicaciones

Análisis y Visualización de los datos

* El objetivo es remarcar/indicar información útil y dar soporte para la toma de decisiones.
* Ayudan a los educadores y administradores de cursos a analizar las actividades de los estudiantes y la información de utilización para obtener una visión general del aprendizaje de los estudiantes.
* Las técnicas más utilizadas son:
  * **Estadísticas** que permiten obtener sumarios e informes.
  * **Visualización de la información** que utilizan gráficas para ayudar a comprender mejor los datos.

Tareas y Aplicaciones

Proporcionar información de retroalimentación (feedback) al profesor/autor/administrador

* El objetivo es proporciona información de ayuda al profesor/autor/administrador del curso en la toma de decisiones (sobre como mejorar el aprendizaje de los alumnos, organizar los recursos utilizado para la enseñanza más eficientemente, etc.) y permitir realizar apropiadas acciones proactivas y/o de remedio.
* Esta tarea es diferente a la anterior de análisis de datos y visualización, que sólo mostraba información básica directamente obtenida de los datos (informes y estadísticas). En cambio, ahora se descubre una información nueva, oculta e interesante.
* Varias técnicas de DM se han empleado en esta tarea (clustering, clasificación, análisis de patrones secuenciales, etc.), aunque la más común es la **minería de Reglas de Asociación**.

Tareas y Aplicaciones

Realizar recomendaciones a los estudiantes

* El objetivo es hacer recomendaciones o sugerencias directamente a los estudiantes con respecto a sus actividades, enlaces a visitar, siguiente tarea o problema por hacer. etc.
* Esta tarea permite personalizar o adaptar tanto los contenidos de aprendizaje, como los interfaces de usuario o la secuencia de aprendizaje a cada alumno en particular.
* Varias técnicas de DM se han aplicado en esta tarea como: minería de reglas de asociación, clustering, clasificación y patrones de secuencias.

Tareas y Aplicaciones

Predicción del rendimiento de los estudiantes

* El objetivo de la predicción es estimar un valor desconocido de una variable que describe el rendimiento del estudiante. Este valor suele referirse al conocimiento, puntuación o la nota que tiene el alumno en un curso o concepto determinado.
* Este valor a predecir puede ser numérico/continuo (sería una tarea de regresión) o categórico/discreto (sería una tarea de clasificación).
* Esta tarea es la más antigua y más popular de EDM, y muchas técnicas se han empleado como: neural networks, Bayesian networks, rule-based systems, regression and correlation analysis.

Tareas y Aplicaciones

Modelado del Estudiante

* El objetivo del modelado del estudiante es desarrollar un modelo cognitivo del estudiante que incluya un modelo de las habilidades (skills) y del conocimiento declarativo.
* Para automatizar la creación de modelos de usuario la minería de datos se a utilizado para obtener características del usuario como: motivación, satisfacción, estilos de aprendizaje, estados afectivos, etc.
* Diferentes técnicas de DM se han utilizado, pero principalmente las redes Bayesianas.

Tareas y Aplicaciones

Detectar comportamientos no deseados de los estudiantes

* El objetivo es descubrir o detectar aquellos estudiantes que tienen algún tipo de comportamiento raro, inusual, no deseado, problemático, etc.
* Ejemplos de estos comportamientos son las acciones erróneas, baja motivación, engaños/juego(cheating/gamming), abusos/despilfarros (misuse), abandono (drop out), fracaso académico (academic failure), etc.
* Las principales técnicas de DM que se han utilizado para detectar estos comportamiento a tiempo y poder proporcionar algún tipo de ayuda han sido Clasificación y clustering.

Tareas y Aplicaciones

Agrupación de estudiantes

* El objetivo es crear grupos de estudiantes según una serie de características que presenten.
* A partir de los grupos de estudiantes se puede realizar una enseñaza personalizada, promover aprendizaje en grupo, proporcionar contenidos adaptados, etc.
* Las técnicas de DM más utilizadas en esta tarea son la clasificación (supervised learning) y el clustering (unsupervised learning).

Tareas y Aplicaciones

Análisis de Redes Sociales

- Análisis de redes sociales o Social Networks Analysis (SNA) estudia las relaciones entre individuos en lugar de entre atributos individuales o propiedades.
- Una red social se le puede considerar a un grupo de personas, organización o individuos que están conectados mediante una relación de amistad, cooperación, trabajo, intercambio de información, etc.
- Diferentes técnicas de minería de datos se han utilizado en entornos educativos, pero la más popular son las técnicas de asociación, clustering y filtrado/recomendación colaborativa.

Tareas y Aplicaciones

Construcción de mapas conceptuales

* El objetivo de la construcción de los mapas conceptuales es ayudar a los profesores/educadores en automatizar el proceso de construción de estos mapas.
* Un mapa conceptual es un grafo que muestra relaciones entre conceptos, que se utilizan como herramienta gráfica para organizar y representar el conocimiento, por ejemplo, ontologías.
* Las principales técnicas de minería de datos que se han utilizado son reglas de asociación y minería de textos.

Tareas y Aplicaciones

Construcción de cursos

* El objetivo es ayudar al instructor y/o desarrollador de cursos en el proceso de desarrollo/construcción de cursos y contenidos de aprendizaje (y poder llegar a hacerlo de forma automática).
* También promueve el intercambio y/o reutilización de los recursos de aprendizaje existentes entre diferentes usuarios y sistemas.
* Clasificación y clustering son las técnicas más utilizadas en esta tarea.

Tareas y Aplicaciones

Planificación/Programación

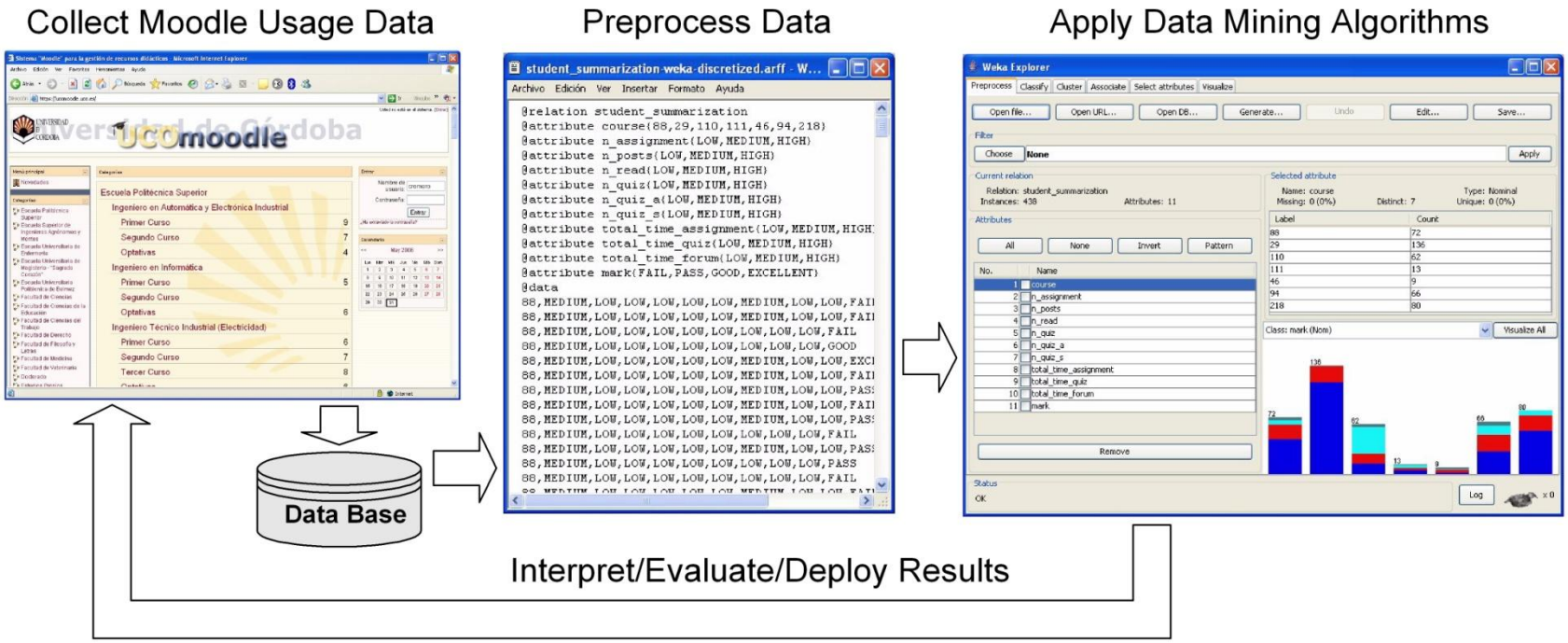
* El objetivo de la planificación (planning) / programación (scheduling) es mejorar el proceso tradicional de planificación de cursos, ayudar a los estudiantes en la programación de cursos, planificación en la asignación de recursos, ayudar en los procesos de admisión y asesoramiento, desarrollo curricular, etc.
* Las principales técnicas de DM utilizadas son reglas de asociación y clasificación.

Técnicas empleadas en EDM

Técnicas empleadas en EDM

- Las técnicas empleadas son las mismas que se utilizan en cualquier campo de aplicación de la minería de datos.
- Los usuarios finales de las herramientas son los agentes implicados en el proceso educativo, por lo que:
  - El objetivo final es mejorar el aprendizaje
  - Los algoritmos deben de ser fáciles de configurar
  - Los resultados deben de ser fáciles de interpretar
- Tareas de DM que se han utilizado en educación:
  - Estadísticas
  - Visualización de información
  - Clasificación
  - Clustering
  - Asociación y patrones de secuencia
  - Minería de Textos

Proceso específico de ejemplo: Minería de Datos Moodle



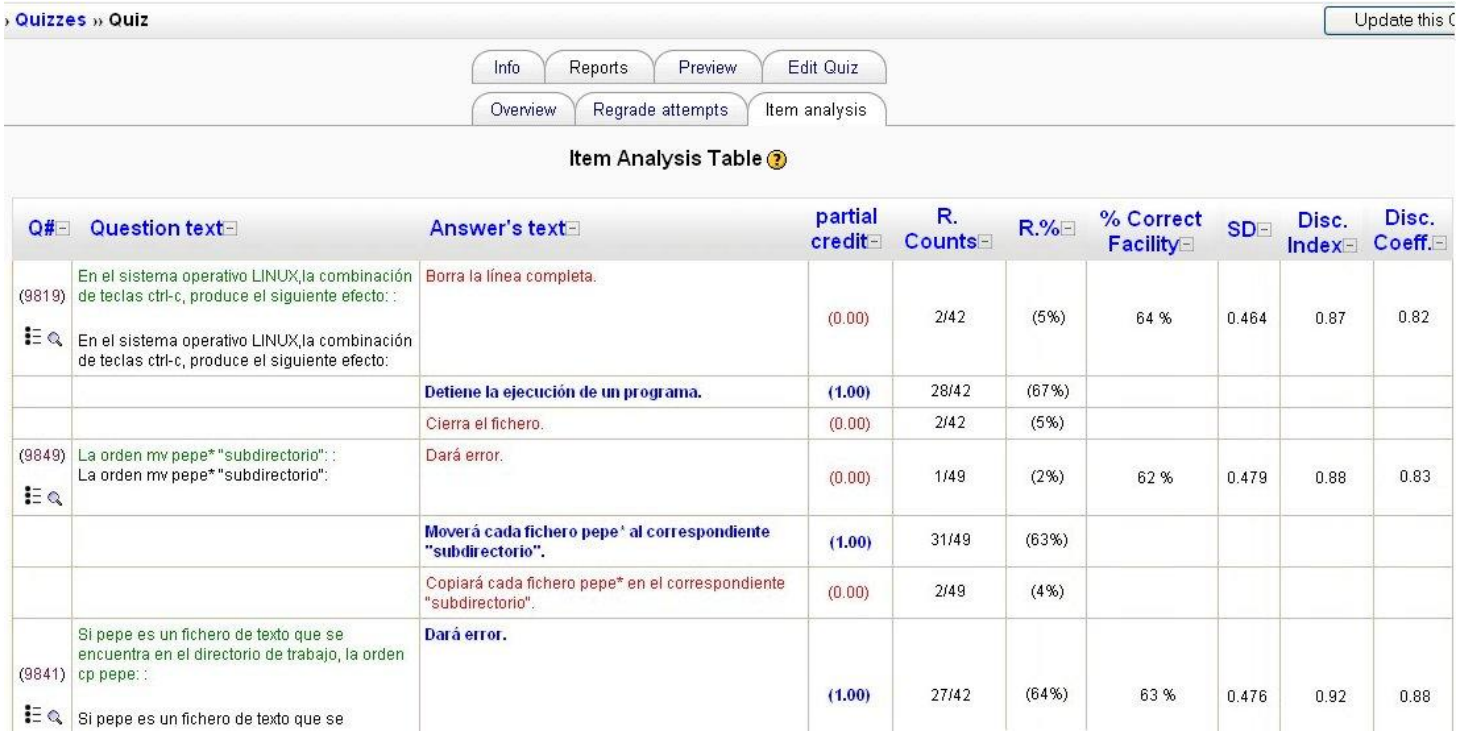
Técnicas empleadas en EDM
Estadísticas

- Las estadísticas sobre la utilización del curso es la primera técnica de evaluación empleada en los sistemas de e-learning, aunque no se suele considerar como minería de datos.
- Algunos ejemplos de estadísticos empleados son:
  - Número total de visitas al curso.
  - Número total de visitas por página y/o actividad.
  - Páginas/Actividades más y menos visitadas
  - Tiempos de acceso al curso y a páginas/actividades.
  - Medias de las puntuaciones obtenidas en las actividades/test.
  - …
- Conociendo estos datos, el profesor puede tener información general sobre como mejorar el curso, incluso podría detectar algunos problemas evidentes.

Técnicas empleadas en EDM

Estadísticas - Moodle

Ejemplo: Análisis de los Items o Preguntas de un Quiz/Test de Moodle.
Objetivo: Permite mejorar el rendimiento de un test o crear un Test Adaptativos (IRT).



› Quizzes » Quiz

Update this (

Info | Reports | Preview | Edit Quiz

Overview | Regrade attempts | Item analysis

Item Analysis Table

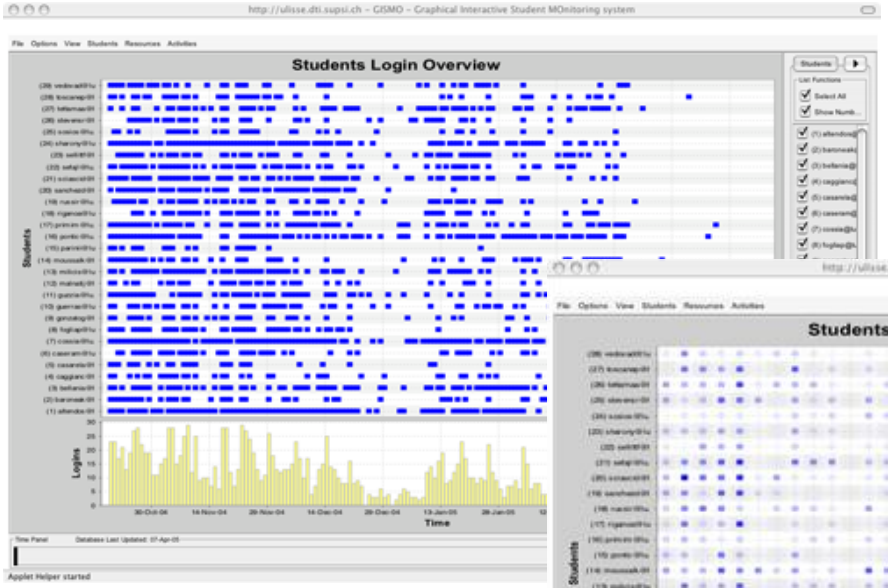
Q#	Question text	Answer's text	partial credit	R. Counts	R.%	% Correct Facility	SD	Disc. Index	Disc. Coeff.
(9819)	En el sistema operativo LINUX,la combinación de teclas ctrl-c, produce el siguiente efecto: : En el sistema operativo LINUX,la combinación de teclas ctrl-c, produce el siguiente efecto:	Borra la línea completa.	(0.00)	2/42	(5%)	64 %	0.464	0.87	0.82
		Detiene la ejecución de un programa.	**(1.00)**	28/42	(67%)				
		Cierra el fichero.	(0.00)	2/42	(5%)				
(9849)	La orden mv pepe* "subdirectorio": : La orden mv pepe* "subdirectorio":	Dará error.	(0.00)	1/49	(2%)	62 %	0.479	0.88	0.83
		Moverá cada fichero pepe* al correspondiente "subdirectorio".	**(1.00)**	31/49	(63%)				
		Copiará cada fichero pepe* en el correspondiente "subdirectorio".	(0.00)	2/49	(4%)				
(9841)	Si pepe es un fichero de texto que se encuentra en el directorio de trabajo, la orden cp pepe: : Si pepe es un fichero de texto que se	**Dará error.**	**(1.00)**	27/42	(64%)	63 %	0.476	0.92	0.88

Técnicas empleadas en EDM
Visualización de información

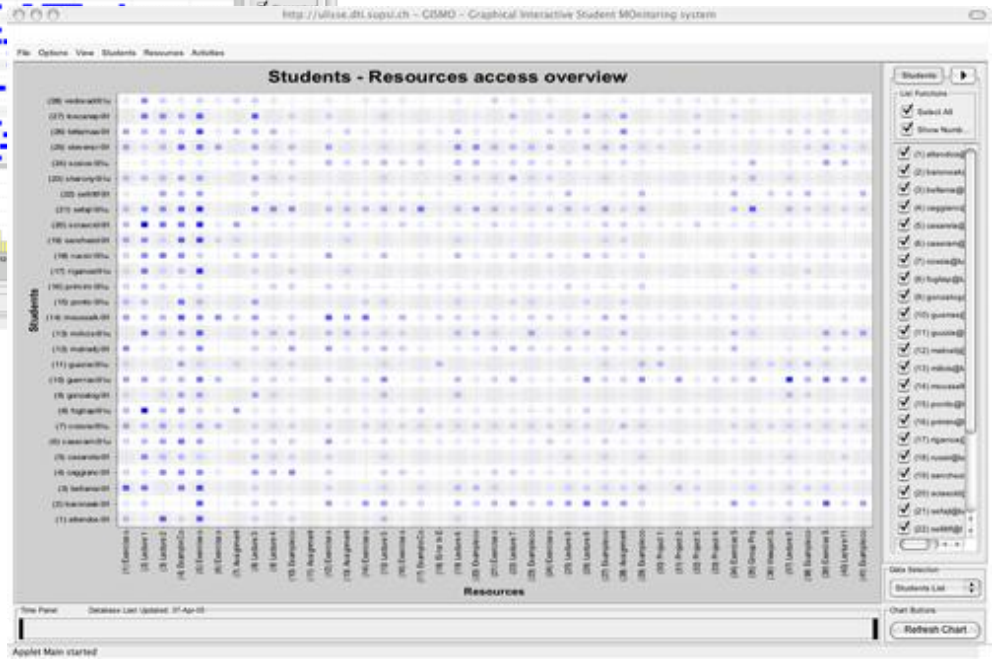
- Construyen imágenes digitales interactivas o animadas orientadas a que los usuarios puedan comprender grandes cantidades de información.
- Existen varias herramientas de visualización para los sistemas educativos:
  - **CourseVis** (http://www.comp.leeds.ac.uk/vania/umuas/coursevis.html) es una herramienta que permite visualizar información generada en los ficheros de log de WebCT.
  - **GISMO** (http://gismo.sf.net) es un proyecto análogo al anterior (de hecho, el autor es la misma persona – R. Mazza) pero que extrae la información de las tablas que almacenan la información en el sistema Moodle.

Técnicas empleadas en EDM

Visualización de información - GISMO



Gráfica que muestra la información de acceso a un curso



Gráfica que muestra la información de acceso a los recursos por parte del alumnado

Técnicas empleadas en EDM
Clasificación

* A partir de un conjunto de patrones de entrenamiento etiquetados hemos de ser capaces de etiquetar nuevos patrones.
* Método de aprendizaje **supervisado**.
* Métodos precisos vs. métodos interpretables. Suelen preferirse los segundos, para:
  * Poder contrastar las conclusiones alcanzadas con el conocimiento de los expertos humanos.
  * Que de los modelos pueda extraerse información útil por parte de los usuarios del proceso de EDM
* Algoritmos empleados:
  * Árboles de Decisión.
  * Inducción de Reglas.
  * Softcomputing: redes neuronales, algoritmos evolutivos, etc.

Técnicas empleadas en EDM

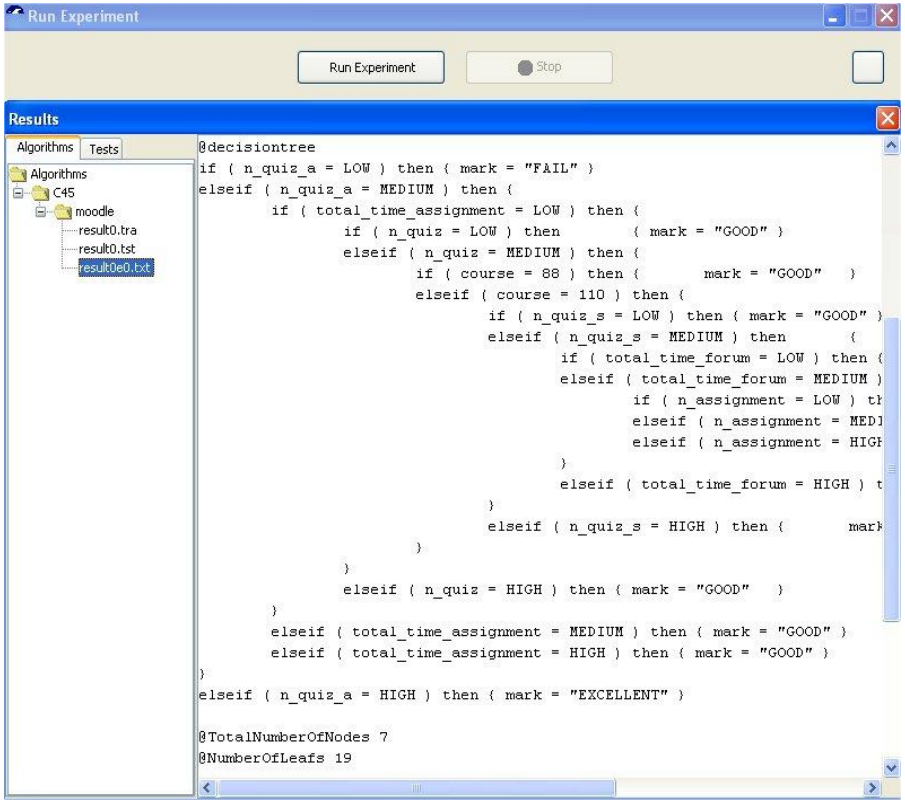
Clasificación – Algunas aplicaciones

* Descubrir grupos potenciales de estudiantes con características similares, para definir una determinada estrategia pedagógica (Chen et al, 2000).
* Predecir el rendimiento de estudiantes y su calificación final (Minaei-Bidgoli & Punch, 2003)
* Detectar estudiantes que hacen un mal uso de las instalaciones o que juegan (Baker et al., 2004).
* Agrupar los estudiantes en (a) guiados a través de consejos y (b) a través de fallos y encontrar los conceptos erróneos que presentan con más frecuencia (Yudelson et al., 2006).
* Identificar alumnos con una motivación baja y encontrar remedio al problema de abandono de los estudios (Cocea & Weibelzahl, 2006).

Técnicas empleadas en EDM

Clasificación – Un ejemplo con datos de Moodle

- Datos de alumnos de la Universidad de Córdoba.
- Intentamos establecer una relación entre la calificación final y la participación en las actividades planteadas a los alumnos a través de la plataforma Moodle.
- Información disponible:
  - Participación en las distintas actividades
  - Calificación final discretizada (excellent, good, pass & fail)
- Utilizamos el algoritmo C4.5 disponible en la herramienta de Data Mining KEEL (http://www.keel.es).
- Las reglas extraídas del árbol de decisión son indican, por ejemplo, que los alumnos que sacan una elevada calificación en los cuestionarios suele sacar una elevada calificación final…



Técnicas empleadas en EDM

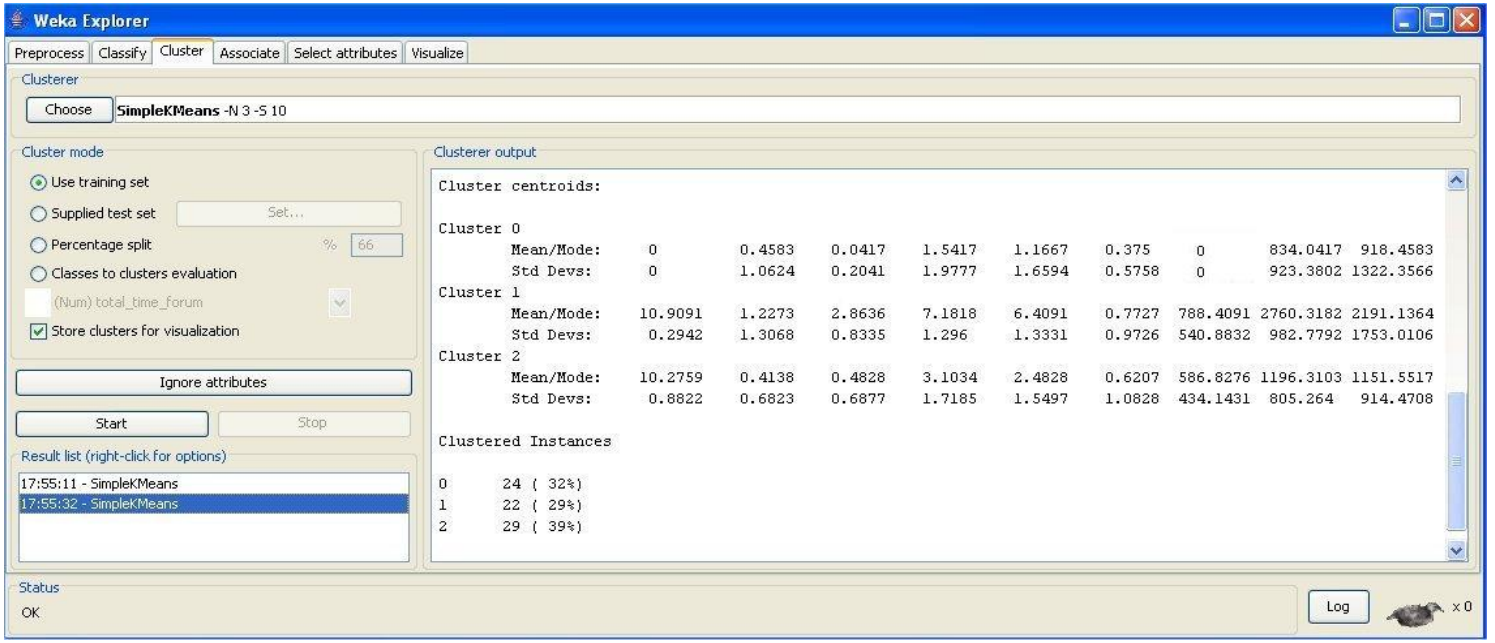
Clustering

- Establecer grupos de objetos que presentan características similares.
- Método no supervisado.
- Algoritmos principales:
  - Jerárquicos: single-link, complete-link
  - Basados en función objetivo: **K medias**, expectation maximization
- Algunas aplicaciones:
  - Descubrir patrones que reflejen comportamientos análogos en los usuarios, de cara a que, cuando se les incluya en espacios de colaboración comunes, se asegurar un incremento de la actividad (Talavera & Gaudioso, 2004).
  - Agrupar estudiantes para establecer itinerarios de educación personalizados (Mor y Minguillon, 2004).
  - Agrupar estudiantes según sus destrezas y otras características, para a realizar tutorías de forma personalizada (Hamalainen et al., 2004).
  - Agrupar alumnos de características similares para promover un aprendizaje colaborativo basado en grupos (Tang & McCalla, 2005).
  - Agrupan test y cuestiones en grupos relacionados basándose en datos de una matriz de puntuaciones (Spacco et al., 2006).

Técnicas empleadas en EDM

Clustering – Un ejemplo con datos de Moodle

* El objetivo es agrupar a los estudiantes de un determinado curso en diferentes grupos, relacionados con las actividades realizadas en Moodle.
* Para llevar a cabo los experimentos debemos tomar la información de la base de datos y llevarla a una aplicación de minería de datos (Moodle no dispone aún de un sistema de extracción de conocimiento integrado).
* Weka (http://www.cs.waikato.ac.nz/ml/weka/) es la herramienta elegida para realizar el clustering con los datos obtenidos. Utilizamos algoritmo K-medias.



Técnicas empleadas en EDM

Clustering – Un ejemplo con datos de Moodle

Se obtienen tres grupos de características distintas:

* **Grupo 0**. Formado por alumnos que no realizan las tareas asignadas, que leen una muy baja proporción de mensajes de los foros, realizan muy pocos cuestionarios, y pasan muy poco tiempo en las actividades tarea, cuestionario y foro (es decir, participan muy poco).
* **Grupo 1**. Alumnos que envían bastantes mensajes al foro (1.22 en media), leen alrededor de 3 mensajes, realizan un elevado número de cuestionarios, acertando un porcentaje elevado de estos y pasan un tiempo elevado en las actividades tarea, cuestionario y foro.
* **Grupo 2**. Valores un poco inferiores a los del grupo 1 pero superiores a los del grupo 0.

Técnicas empleadas en EDM

Asociación

- El objetivo de la minería de reglas de asociación es establecer reglas que asocian conceptos que se encuentran en columnas (atributos) diferentes de una misma base de datos.
- Principales algoritmos:
  - **Apriori**. Es el primero y más popular
  - Variantes del A priori: A priori-TID, DIC, Eclac, FP-Growth…
- Algunas aplicaciones de los algoritmos de asociación:
  - Búsqueda de relaciones entre cada patrón de comportamiento de los estudiantes (Yu et al, 2001).
  - Construcción de agentes que recomiendan y generan de forma inteligente materiales didácticos para los estudiantes (Zaïane, 2002).
  - Guiar la búsqueda de modelos de comportamiento del estudiante más fiables (Freyberger et al., 2004).

Técnicas empleadas en EDM
Asociación

* Algunas otras aplicaciones de los algoritmos de asociación:
  * Guiar la actividad del estudiante de forma automática y generar y recomendar automáticamente materiales didácticos (Lu, 2004).
  * Buscar errores de los estudiantes que suelen ocurrir conjuntamente (Merceron & Yacef, 2004).
  * Identificar atributos que caracterizan patrones de disparidad de rendimiento entre grupos de estudiantes (Minaei-Bidgoli et al., 2004).
  * Descubrir relaciones interesantes entre la información generada por los estudiantes en un sistema adaptativo (usage information), orientadas a retroalimentar el curso (Romero et al., 2004).
  * Para determinar qué materiales didácticos son los más apropiados para recomendar a los alumnos (Markellou et al, 2005).
  * Para optimizar el contenido de un portal de e-learning determinando qué es lo que más interesa a los usuarios (Ramli, 2005).

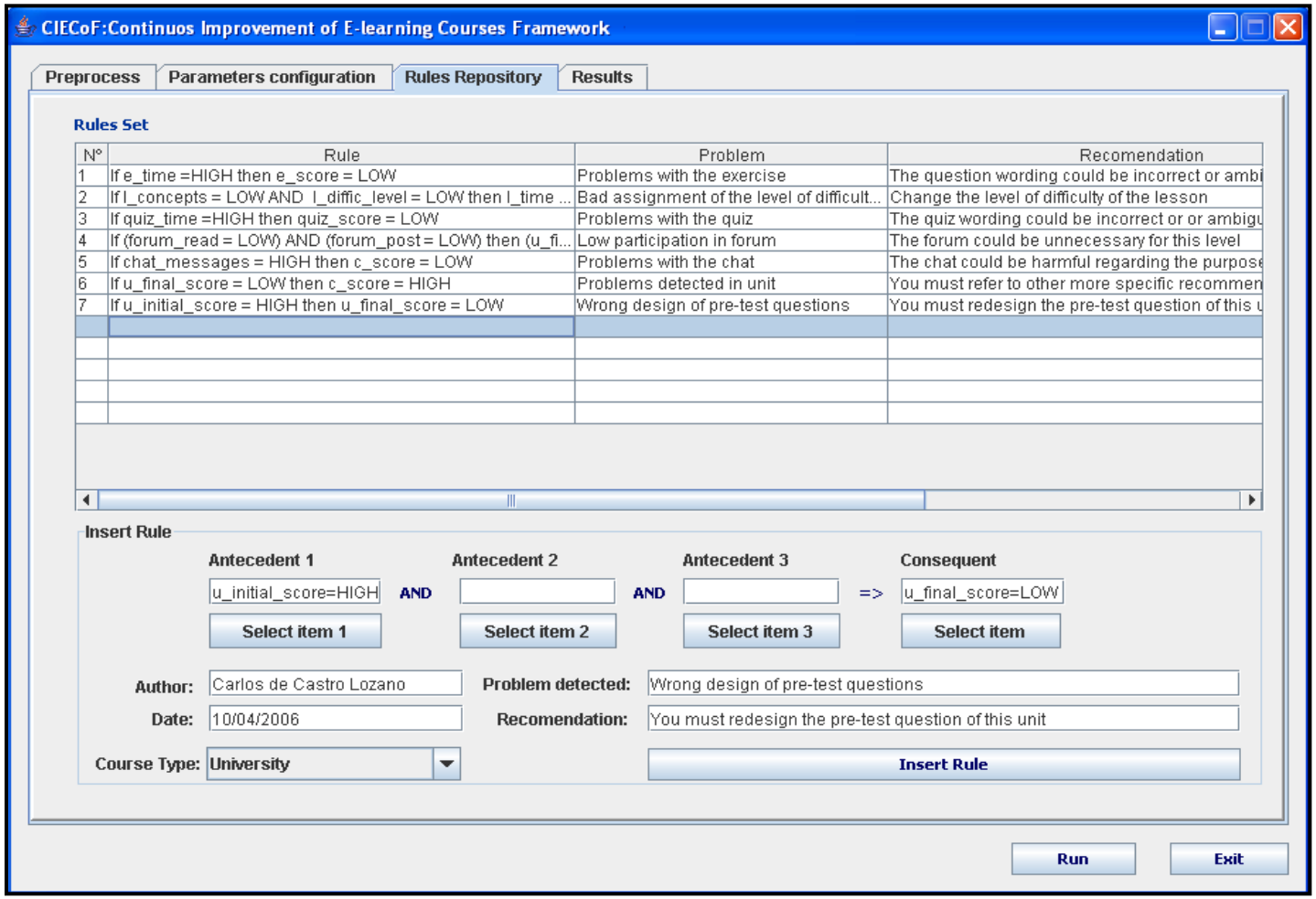
Técnicas empleadas en EDM

Asociación – Ejemplo con datos de INDESHAC

- Herramienta **INDESHAC**:
  - La distribución estándar de Moodle no dispone de cursos adaptativos.
  - El grupo EATCO de la Universidad de Córdoba ha desarrollado una herramienta autor INDESHAC que permite la construcción de cursos adaptativos dentro del sistema Moodle de una forma sencilla.
  - Estos cursos pueden utilizar todos los recursos disponibles en Moodle, y organizan su contenido por niveles de dificultad.
- **Algoritmos de Asociación**:
  - **A priori**. A veces no es sencillo elegir los valores de soporte y confianza para que el número de reglas no sea demasiado amplio o demasiado reducido.
  - **A priori Predictivo**. Se define un único parámetro, el número de reglas que se presentan al usuario.

Técnicas empleadas en EDM

Asociación – Ejemplo con datos de INDESHAC



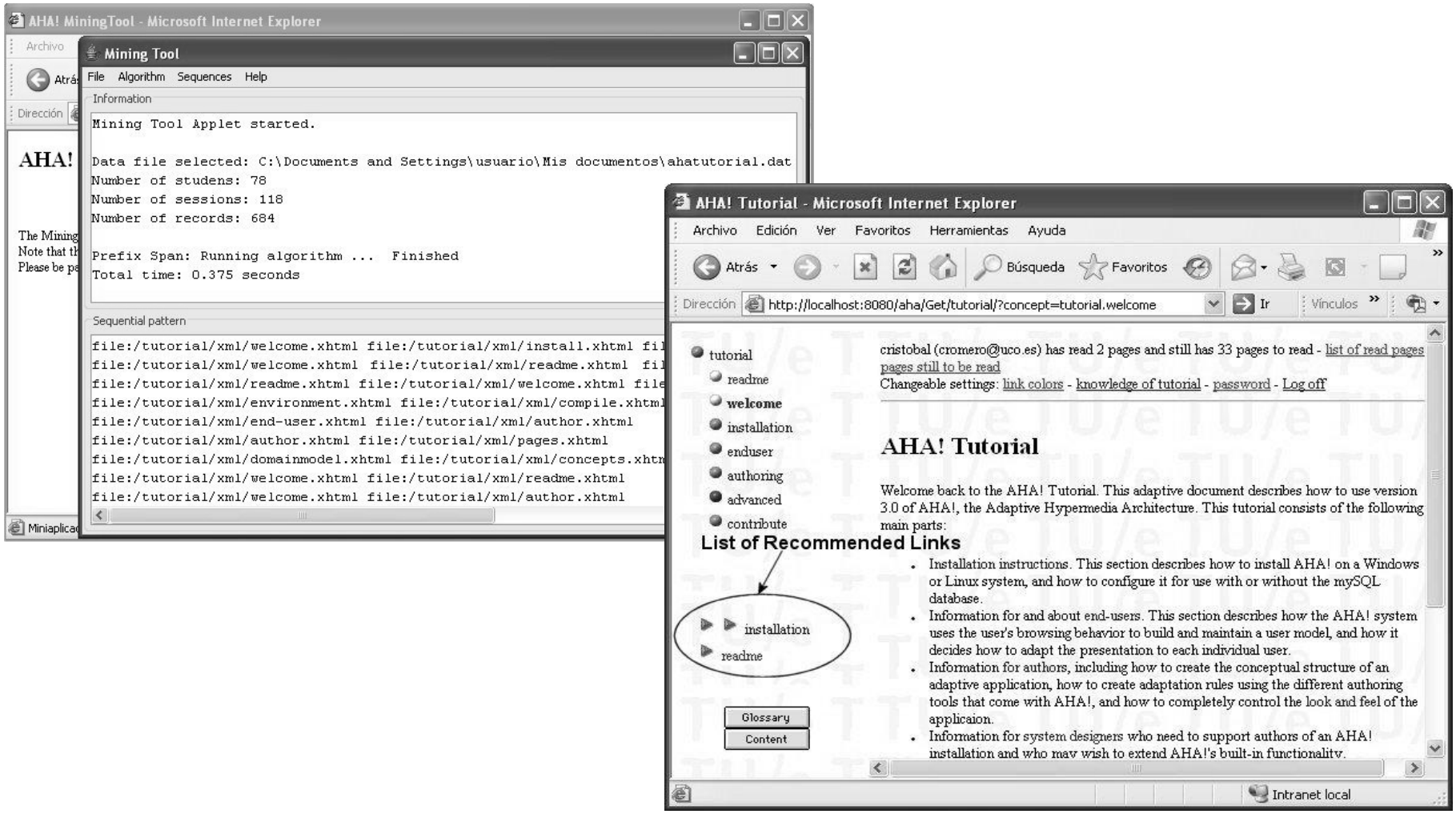
Técnicas empleadas en EDM

Patrones de secuencia

- Descubrir patrones entre sesiones.
- Algoritmos principales: AprioriAll, GSP, SPADE, PrefixSpan, CloSpan, FreSpan
- Algunas aplicaciones:
  - Dar una indicación de cómo organizar mejor el espacio educativo web y ser capaz de hacer sugerencias a los estudiantes que comparten características similares (Ha et al., 2000).
  - Evaluar las actividades del estudiante y personalizar el envío de recursos (Zaïane & Luo, 2001).
  - Llevar a cabo la evluación y validación de diseños de sitios web educativos (Machado & Becker, 2003).
  - Comparar los caminos extraídos con otros patrones de comportamiento ideales, especificados por el diseñador del curso o por el educator (Pahl & Donnellan, 2003).
  - Generar actividades personalizadas para distintos grupos de estudiantes (Wang et al., 2004).
  - Identificar secuencias de interacción indicativas de problemas y patrones que son indicativos de éxito (Kay et al., 2006).

Técnicas empleadas en EDM

Patrones de secuencia – Un ejemplo con datos de AHA!



Técnicas empleadas en EDM

Text Mining

- Extensión de las tareas de minería de datos a datos textuales.
- Conjunto de disciplinas que incluyen la minería de datos, recupera-ción de información y procesamiento del lenguaje natural.
- Algunas aplicaciones:
  - Dar soporte a los autores en el desarrollo de materiales (Grobelnik et al, 2002.)
  - Buscar y organizar material utilizando información semántica (Tane et al., 2004)
  - Para evaluar el progreso de un foro de discusión y ver qué contribuciones se están haciendo al debate (Dringus & Ellis, 2005).
  - Agrupar documentos en base a temas y similaridades. Producir resúmenes de documentos (Hammouda & Kamel, 2006).
  - Detectar el foco de la conversación en hilos de discusión, clasificando temas y estimando la profundidad técnica de una contribución (Kim et al., 2006).

Técnicas empleadas en EDM

Text Mining – Un ejemplo con datos de Moodle

- Pretendemos extraer términos concretos del contenido de los foros de un determinado curso Moodle.
- Software KEA http://www.paynter.info/academia/Kea.php para minería de textos.
- Existen dos tablas relacionadas con los datos de los foros:
  - forum_read: Relacionada con los mensajes que se han leído
  - forum_post: Relacionada con los envíos de mensajes

```
SELECT message FROM moodle.mdl_forum_post where discussion=93
```

- La información extraída se pone en ficheros de texto. El algoritmo extraerá palabras clave analizando la información contenida en estos ficheros (contenido de los foros).
- Podemos analizar si las palabras clave descubiertas por el algoritmo coinciden con las proporcionadas por el profesor (a través de un fichero aparte).
  - En caso afirmativo, el uso de los foros es correcto.
  - En otro caso, puede que se estén usando inadecuadamente

Software específico

Software específico
Herramientas de DM Genéricas

- Existe actualmente una gran cantidad de herramientas de minería de datos de tipo general, tanto:
  - **Comeriales**: DBMiner, SPSS Clementine, IBM DB2 Intelligent Miner, etc.
  - **Gratuitas**: Weka , Keel, Rapid Miner, etc.
- Sin embargo, estas herramientas no estan diseñadas para propósitos educacionales/pedagógicos y sobrepasan a un usuario tipo educador (no experto en DM) ya que estan diseñadas más pensando en su potencia que en su simplicidad de uso.
- Se están desarrollando un número creciente de herramientas específicas para resolver diferentes problemas educacionales.

Software específico

Herramientas de DM Específicas

Tool	Objective	Reference
WUM tool	To extract patterns useful for evaluating on-line courses.	(Zaïane and Luo, 2001)
EPRules	To discover prediction rules to provide feedback for courseware authors.	(Romero et al., 2004)
GISMO/CourseVis	To visualize what is happening in distance learning classes.	(Mazza and Milani, 2004)
TADA-ED	To help teachers to discover relevant patterns in students' online exercises.	(Merceron and Yacef, 2005)
O3R	To retrieve and interpret sequential navigation patterns.	(Becker et al., 2005)
Synergo/ColAT	To analyze and produce interpretative views of learning activities.	(Avouris et al., 2005)
LISTEN Mining tool	To explore huge student-tutor interaction logs.	(Mostow et al., 2005)
MINEL	To analyze the navigational behavior and the performance of the learner.	(Bellaachia and Vommina, 2006)
LOCO-Analyst	To provide teachers with feedback on the learning process.	(Jovanovic et al., 2007)
Measuring tool	To measure the motivation of online learners.	(Hershkovitz and Nachmias, 2008)
DataShop	To store and analyze click-stream data, fine-grained longitudinal data generated by educational systems.	(Koedinger et al., 2008)
Decisional tool	To discover factors contributing to students' success and failure rates.	(Selmoune and Alimazighi, 2008)
CIECoF	To make recommendations to courseware authors about how to improve courses.	(Garcia et al., 2009)
SAMOS	Student activity monitoring using overview spreadsheets.	(Juan et al., 2009)
PDinamet	To support teachers in collaborative student modeling.	(Gaudioso et al., 2009)
AHA! Mining Tool	To recommend the best links for a student to visit next.	(Romero et al., 2009)
EDM Visualization Tool	To visualize the process in which student solve procedural problem in logic.	(Johnson and Barnes, 2010)
Meerkat-ED	To analyze participation of students in discussion forums using social network analysis techniques.	(Rabbany et al. 2011)
E-learning Web Miner	To discover student's behavior profiles and models about how they work in virtual courses.	(García-Saiz and Zorrilla, 2011)
MMT tool	To facilitate the execution of all the steps in the data mining process of Moodle data for newcomers.	(Pedraza-Perez et al., 2011)

Líneas Futuras

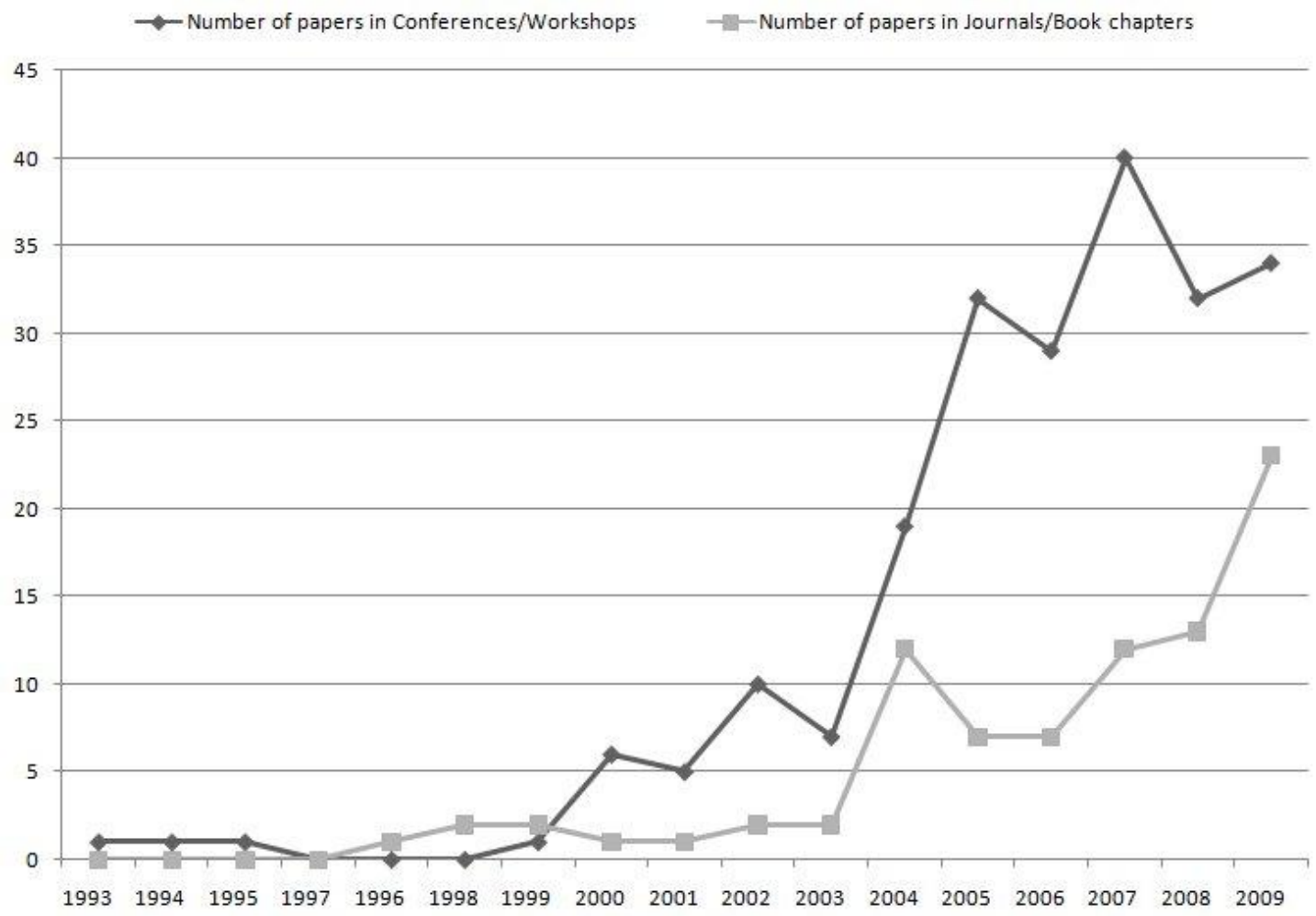
Líneas Futuras

* Aplicar otras tareas y técnicas de DM a educación:
  * Análisis de outliers
  * Análisis de redes sociales
  * Minería de datos en MOOC
  * Minería de semántica web
  * Minería en repositorios de objetos de aprendizaje
* Estandarización de métodos y datos en Educación
* Desarrollo de herramientas y algoritmos de minería de datos más fáciles e intuitivas de utilizar, orientadas para personas no expertas en minería de datos, sino en educación.
* Integración de las herramientas de minería de datos dentro de las propios sistemas autor de construcción y mantenimiento de los LMS, cursos, etc. Utilización de información semántica del dominio en los algoritmos.

Publicaciones

Publicaciones

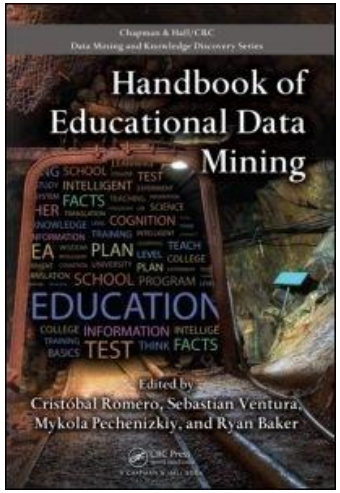
Publicaciones EDM hasta 2009



Publicaciones
Libros

* *Data Mining in E-Learning*.
  C. Romero & S. Ventura (Eds).
  Editorial WIT Press, 2006.

* Handbook of Educational Data Mining.
  C.Romero, S. Ventura,
  M. Pechenizky, R. Baker. (Eds).
  Editorial CRC Press, Taylor & Francis Group. 2010.



Publicaciones
Revisiones

* C. Romero & S. Ventura. Educational Data Mining: A survey from 1995 to 2005. *Expert Systems with Applications* 33:1, pp. 135-146, 2007.
* Castro, F., Vellido, A., Nebot, A. Mugica, F. Applying Data Mining Techniques to e-Learning Problems. In: Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence, 62, Springer-Verlag, 183-221. 2007.
* Baker, R., Yacef, K. The State of Educational Data Mining in 2009: A Review and Future Visions. Journal of Educational Data Mining, 1, 1, 3-17. 2009.
* C. Romero, S. Ventura. Educational Data Mining: A Review of the State-of-the-Art. IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews. 40:6, pp. 601 – 618. 2010.

Publicaciones
Revisiones

* Baker, R.S.J.d. Data Mining for Education. In McGaw, B., Peterson, P., Baker, E. (Eds.) International Encyclopedia of Education (3rd edition), vol. 7, pp. 112-118. Oxford, UK: Elsevier. 2010.
* Scheuer, O. & McLaren, B.M. Educational Data Mining. In the Encyclopedia of the Sciences of Learning, Springer. 2011.
* Karen Cator. Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics. Report of the U.S. Office of Educational Technology. 2012.
* C. Romero, S. Ventura. Data Mining in Education. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. Volume 3, Issue 1, pages 12–27, January/February 2013.

Publicaciones

Revistas

Title of the Journal	**Acronym**	**Editorial**	**Impact Factor 2010**
Journal of Educational Data Mining	JEDM	EDM Society	-
Journal of Learning Analyitics	JLA	SOLAR Society	-
Journal of Artificial Intelligence in Education	JAIED	AIED Society	-
Journal of Educational Psychology	JEP	American Psychological Association	3.583
Journal of the Learning Sciences	JLS	Taylor&Francis	3.644
Computer and Education	CAE	Elsevier	2.617
IEEE Transactions on Learning Technologies	TLT	IEEE	-
IEEE Transactions on Knowledge and Data Engineering	KDE	IEEE	2.290
User Modeling and User-Adapted Interaction	UMUAI	Springer	3.074
Internet and Higher Education	INTHIG	Elsevier	1.896
Decision Support Systems	DCS	Elsevier	2.135
Expert Systems with Applications	ESWA	Elsevier	1.924
Knowledge-Based Systems	KBS	Elsevier	1.574

Publicaciones
Números Especiales

* *Usage Analysis in Learning Systems: Existing Approaches and Scientific Issues. Special Issue in the Journal of Interactive Learning Research (JILR) Choquet, C., V. Luengo and K. Yacef, Eds . 18(1), 2007.*
* *Web mining and higher education. Special Issue in the Journal The Internet and Higher Education. Rafi Nachmias Ed. 14(2), 2011.*
* *Data Mining for Personalized Educational Systems. Special Issue in the Journal User Modeling and User-Adapted Interaction. C. Romero, S. Ventura Eds. 21(1-2), 2011.*
* *Educational Data Mining. Special Issue in the ACM SIGKDD Explorations.* Vol 13, N 2, December 2011.

Publicaciones
Conferencias

- *International Conference on Educational Data Mining*
  - *EDM'08 Montreal*
    - *http://www.educationaldatamining.org/EDM2008*
  - *EDM'09 Córdoba*
    - *http://www.educationaldatamining.org/EDM2009*
  - *EDM'10 Pittsburgh*
    - *http://www.educationaldatamining.org/EDM2010*
  - *EDM'11 Eindhoven*
    - *http://www.educationaldatamining.org/EDM2011*
  - *EDM'12 Creta*
    - *http://educationaldatamining.org/EDM2012/*
  - *EDM'13 Memphis*
    - *http://edm2013.memphis.edu*

Publicaciones
Conferencias relacionadas

* *Learning Analytics & Knowledge*
  * *LAK'11, https://tekri.athabascau.ca/analytics/*
  * *LAK'12. http://lak12.sites.olt.ubc.ca/*
  * *LAK'13. http://lakconference2013.wordpress.com/*
* *Artificial Intelligence in Education*
  * *AIED13* http://aied2013.iismemphis.org/
* *Intelligent Tutoring Systems.*
  * *ITS12* http://its2012.teicrete.gr/
* *Knowledge Discovery and Data Mining*
  * *KDD13* http://www.kdd.org/kdd2013/
* *User Modeling, Adaptation and Personalization*
  * *UMAP13* http://www.dia.uniroma3.it/~umap2013/

Publicaciones
WorkShops

* Workshop on Usage Analysis in Learning Systems at the 12th International Conference on Artificial Intelligence in Education (AIED 2005).
* Educational Data Mining Workshop, at the 13th International Conference on Artificial Intelligence in Education. 2007.
* International Workshop on Applying Data Mining in e-Learning, at the 2nd European Conference on Technology Enhanced Learning, 2007.
* KDD 2011 Workshop: Knowledge Discovery in Educational Data, at the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2011.
* The Hybrid Special Session on Educational Data Mining (EDM-12) at the IJCNN2012, FUZZY-IEEE2012, and CEC2012.

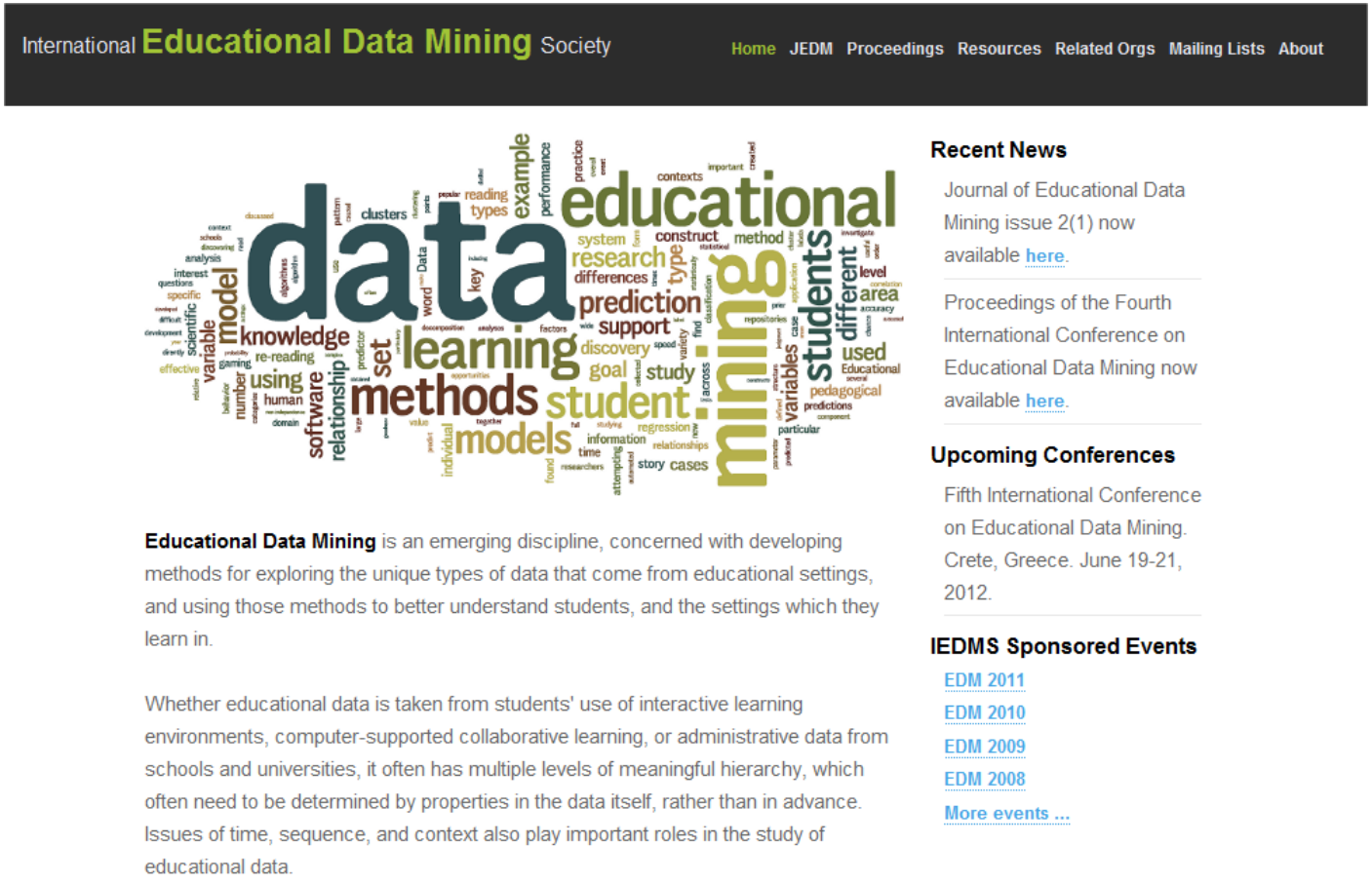
Publicaciones
Propias

* http://www.uco.es/users/in1romoc/html/publications.html

Enlaces Relaccionados

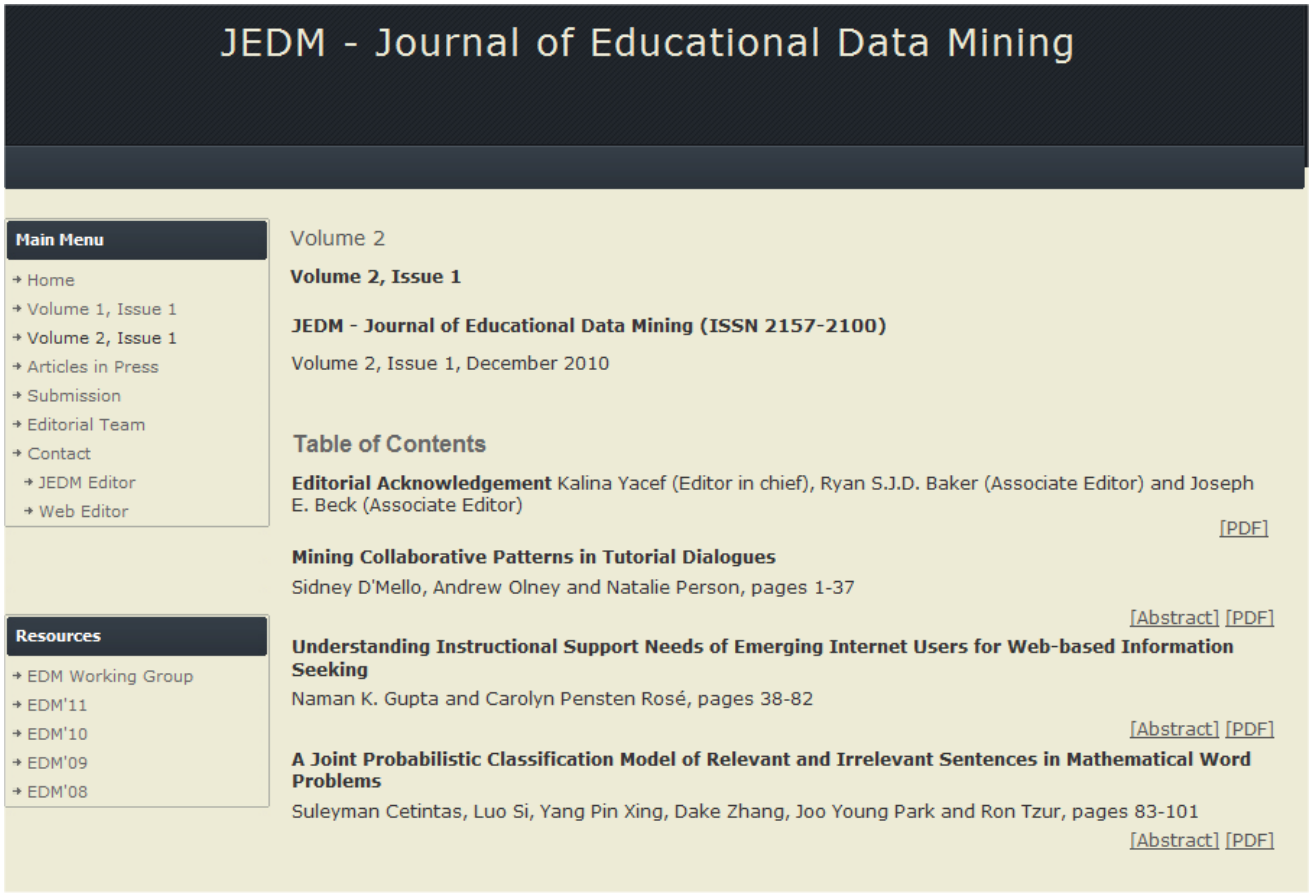
Educational Data Mining Society

* http://www.educationaldatamining.org/



Journal of EDM

* http://www.educationaldatamining.org/JEDM/



JEDM - Journal of Educational Data Mining

Main Menu

- Home
- Volume 1, Issue 1
- Volume 2, Issue 1
- Articles in Press
- Submission
- Editorial Team
- Contact
  - JEDM Editor
  - Web Editor

Resources

- EDM Working Group
- EDM'11
- EDM'10
- EDM'09
- EDM'08

Volume 2

Volume 2, Issue 1

JEDM - Journal of Educational Data Mining (ISSN 2157-2100)

Volume 2, Issue 1, December 2010

Table of Contents

Editorial Acknowledgement Kalina Yacef (Editor in chief), Ryan S.J.D. Baker (Associate Editor) and Joseph E. Beck (Associate Editor)

[PDF]

Mining Collaborative Patterns in Tutorial Dialogues

Sidney D'Mello, Andrew Olney and Natalie Person, pages 1-37

[Abstract] [PDF]

Understanding Instructional Support Needs of Emerging Internet Users for Web-based Information Seeking

Naman K. Gupta and Carolyn Pensten Rosé, pages 38-82

[Abstract] [PDF]

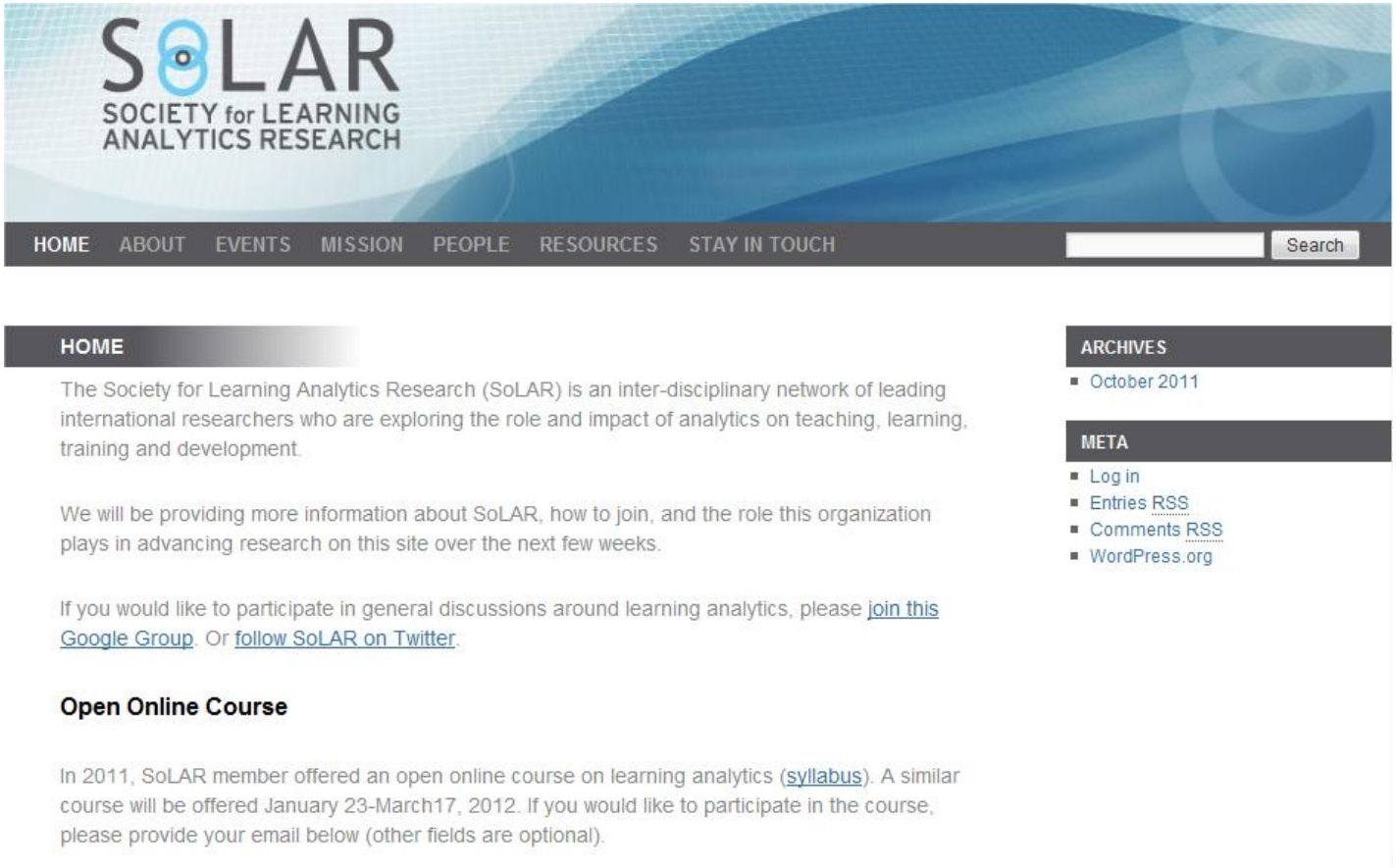
A Joint Probabilistic Classification Model of Relevant and Irrelevant Sentences in Mathematical Word Problems

Suleyman Cetintas, Luo Si, Yang Pin Xing, Dake Zhang, Joo Young Park and Ron Tzur, pages 83-101

[Abstract] [PDF]

Society for Learning Analytics

* http://www.solaresearch.org/

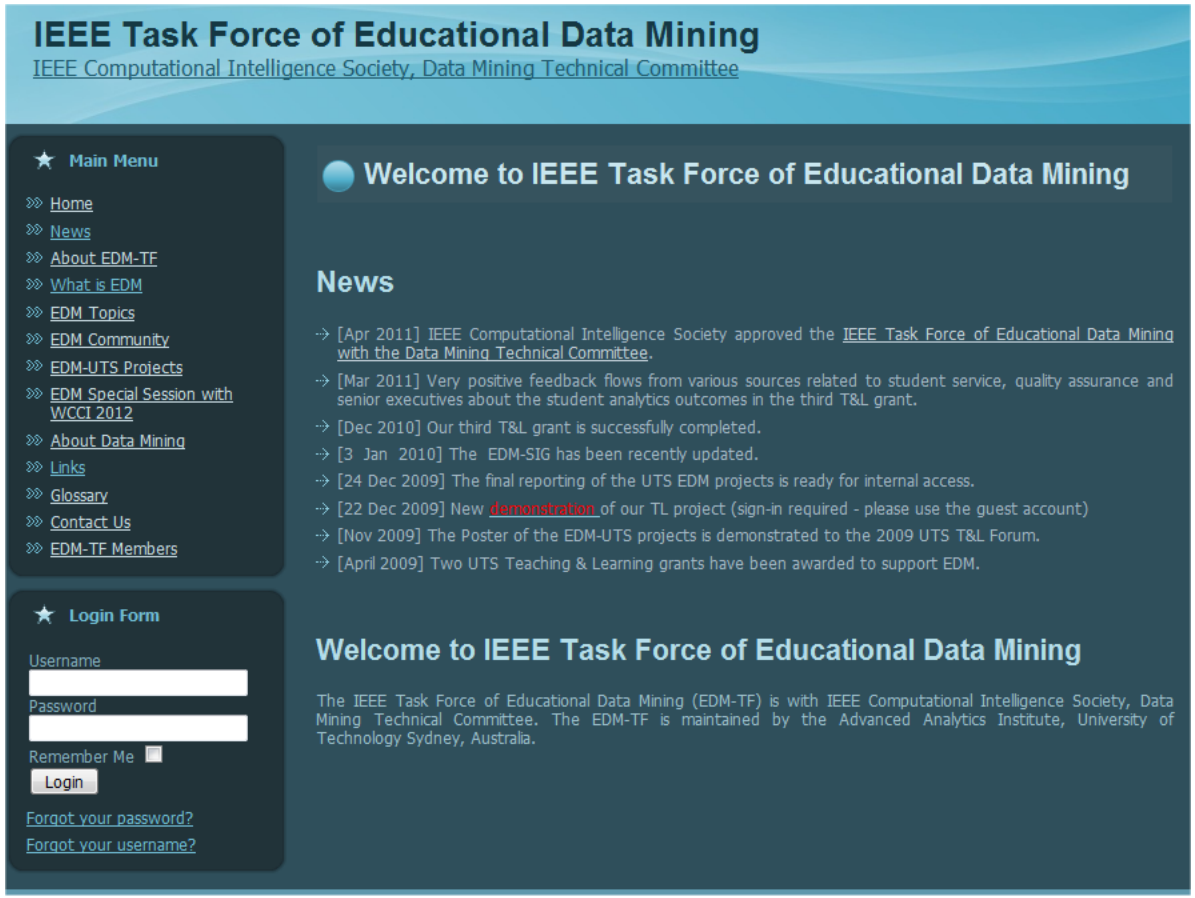


Learning and Knowledge Analytics

* http://www.learninganalytics.net/

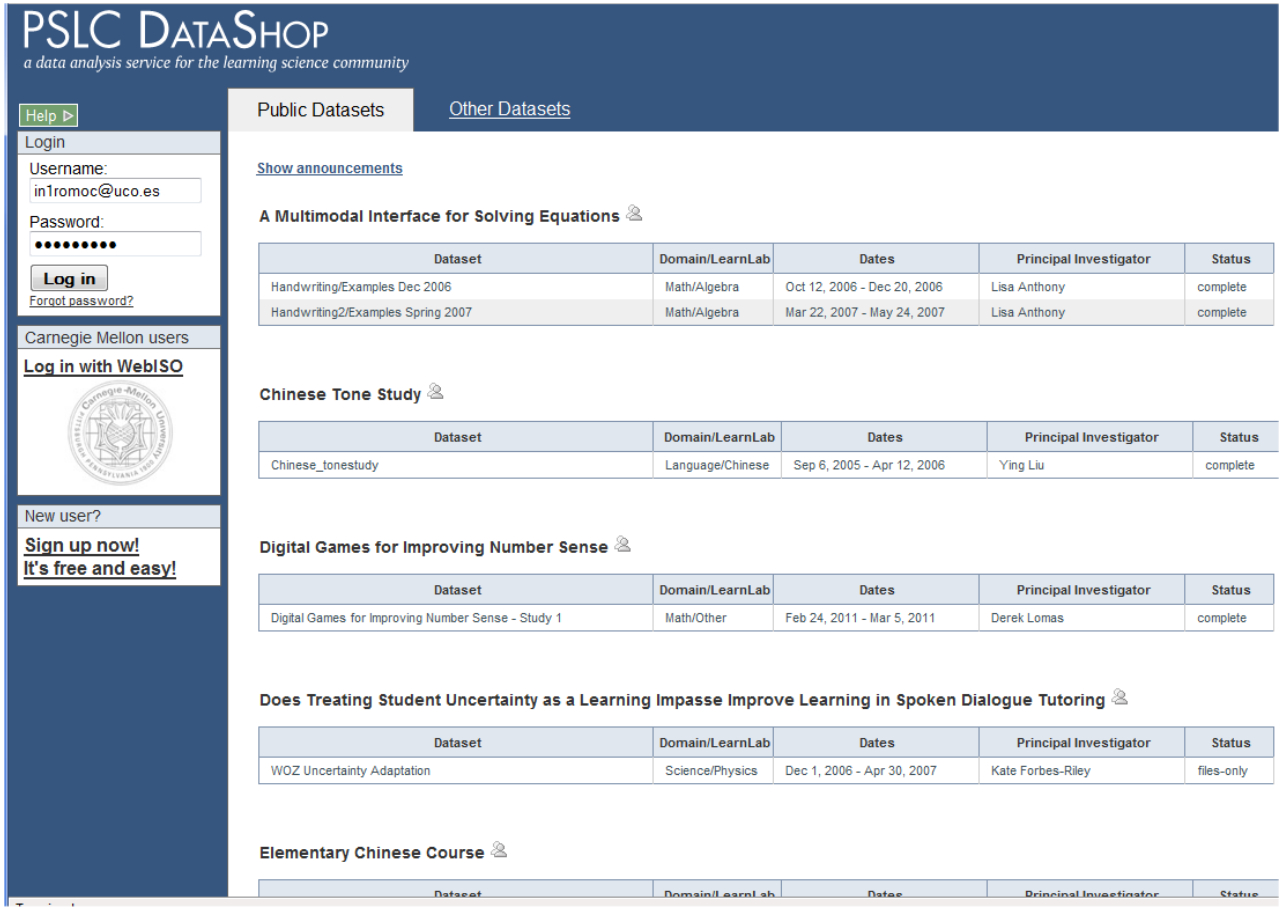
IEEE Task Force of Educational Data Mining

* http://datamining.it.uts.edu.au/edd/



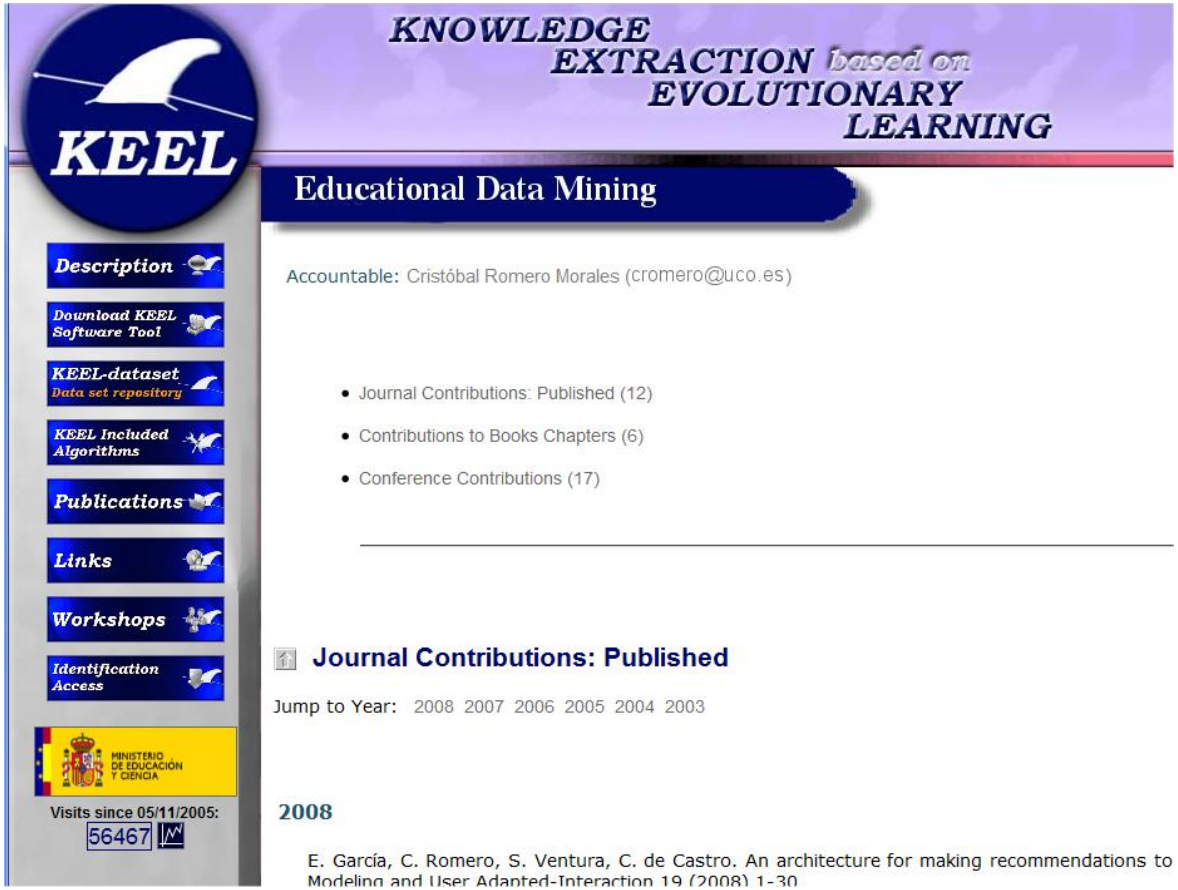
PSLC Data Shop

* https://pslcdatashop.web.cmu.edu/



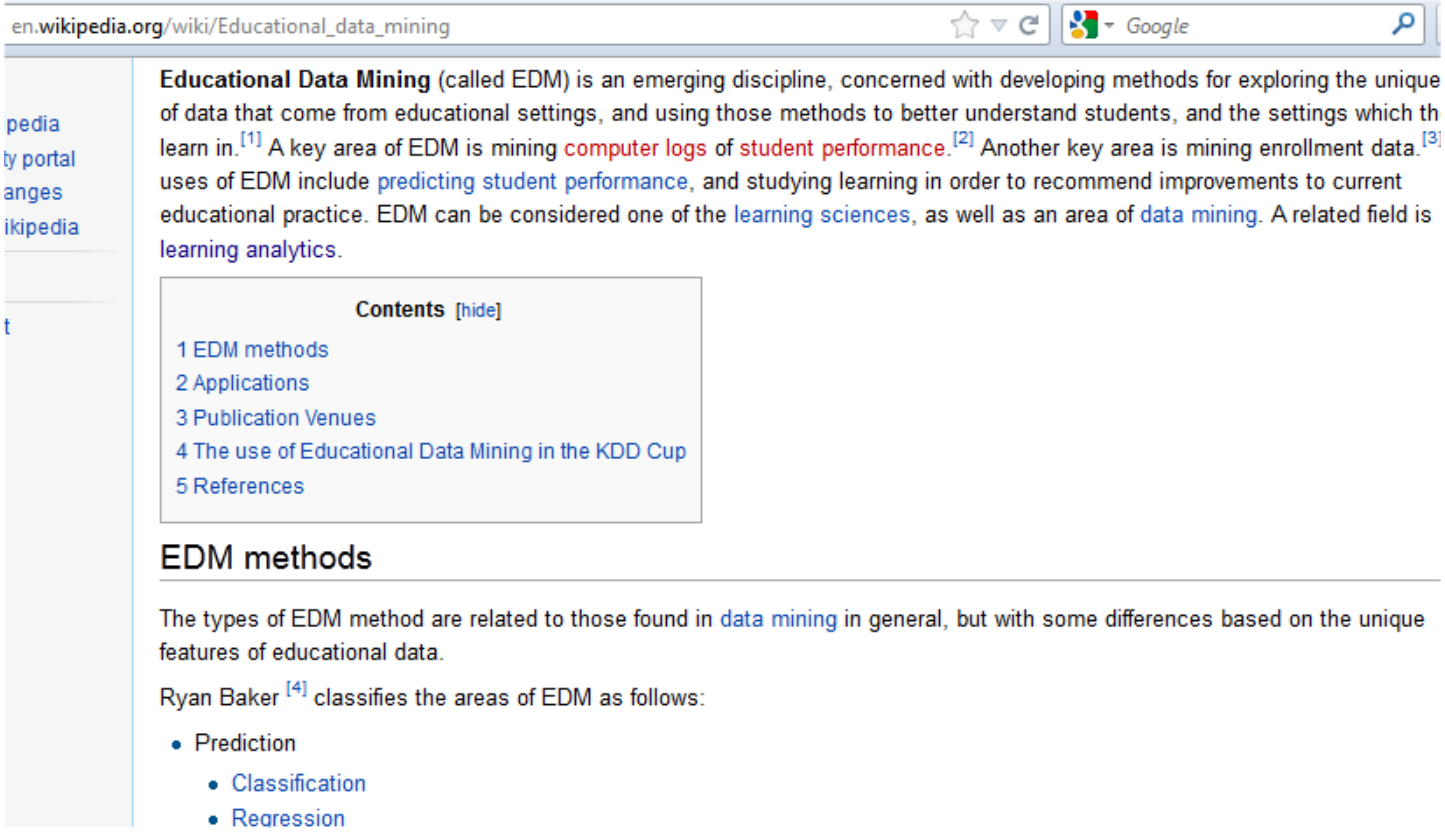
Artículos EDM proyecto Keel

* http://sci2s.ugr.es/keel/specific.php?area=66



Wikipedia EDM

* http://en.wikipedia.org/wiki/Educational_data_mining



en.wikipedia.org/wiki/Educational_data_mining

Google

pedia
y portal
anges
ikipedia

t

Educational Data Mining (called EDM) is an emerging discipline, concerned with developing methods for exploring the unique of data that come from educational settings, and using those methods to better understand students, and the settings which th learn in.[1] A key area of EDM is mining computer logs of student performance.[2] Another key area is mining enrollment data.[3] uses of EDM include predicting student performance, and studying learning in order to recommend improvements to current educational practice. EDM can be considered one of the learning sciences, as well as an area of data mining. A related field is learning analytics.

Contents [hide]

1 EDM methods
2 Applications
3 Publication Venues
4 The use of Educational Data Mining in the KDD Cup
5 References

EDM methods

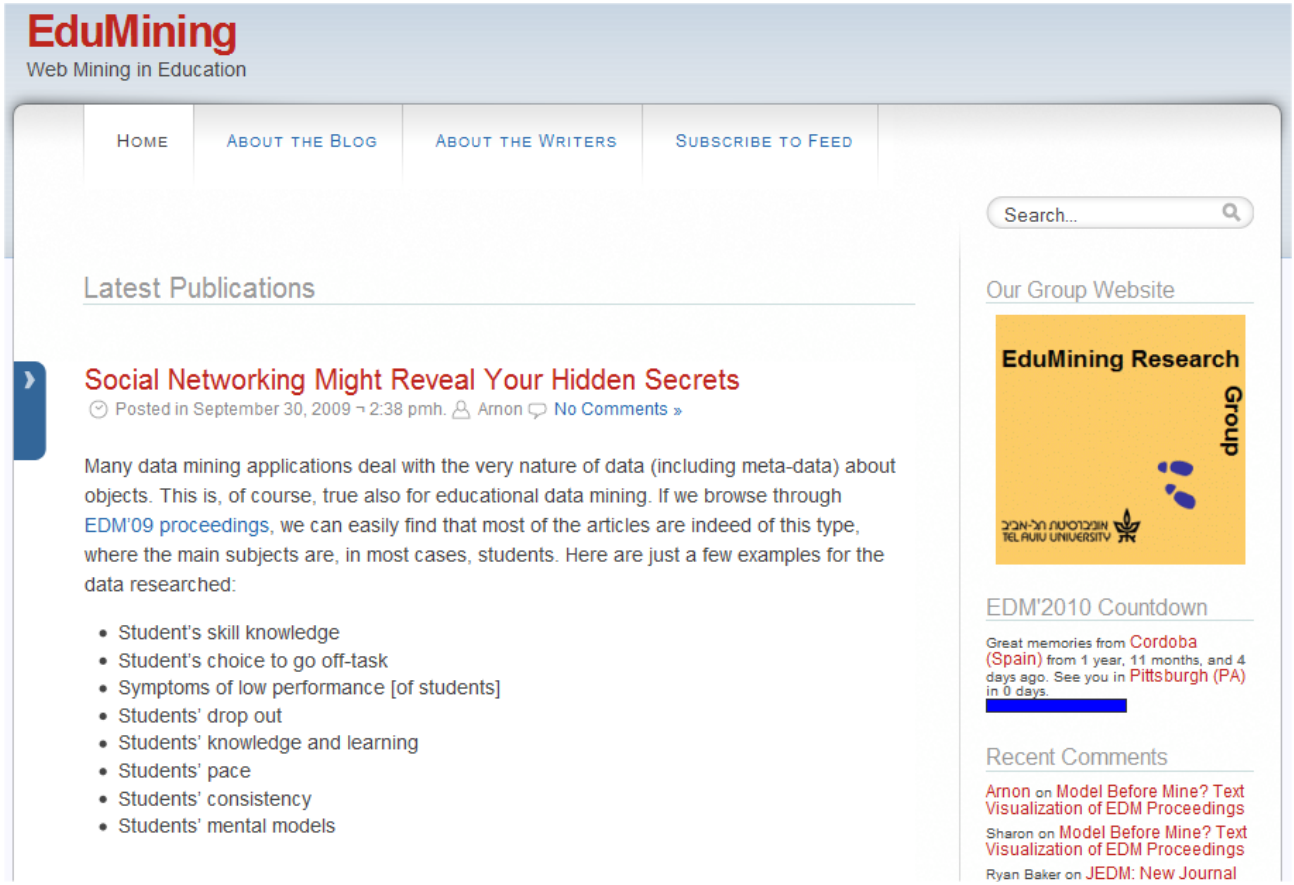
The types of EDM method are related to those found in data mining in general, but with some differences based on the unique features of educational data.

Ryan Baker [4] classifies the areas of EDM as follows:

- Prediction
  - Classification
  - Regression

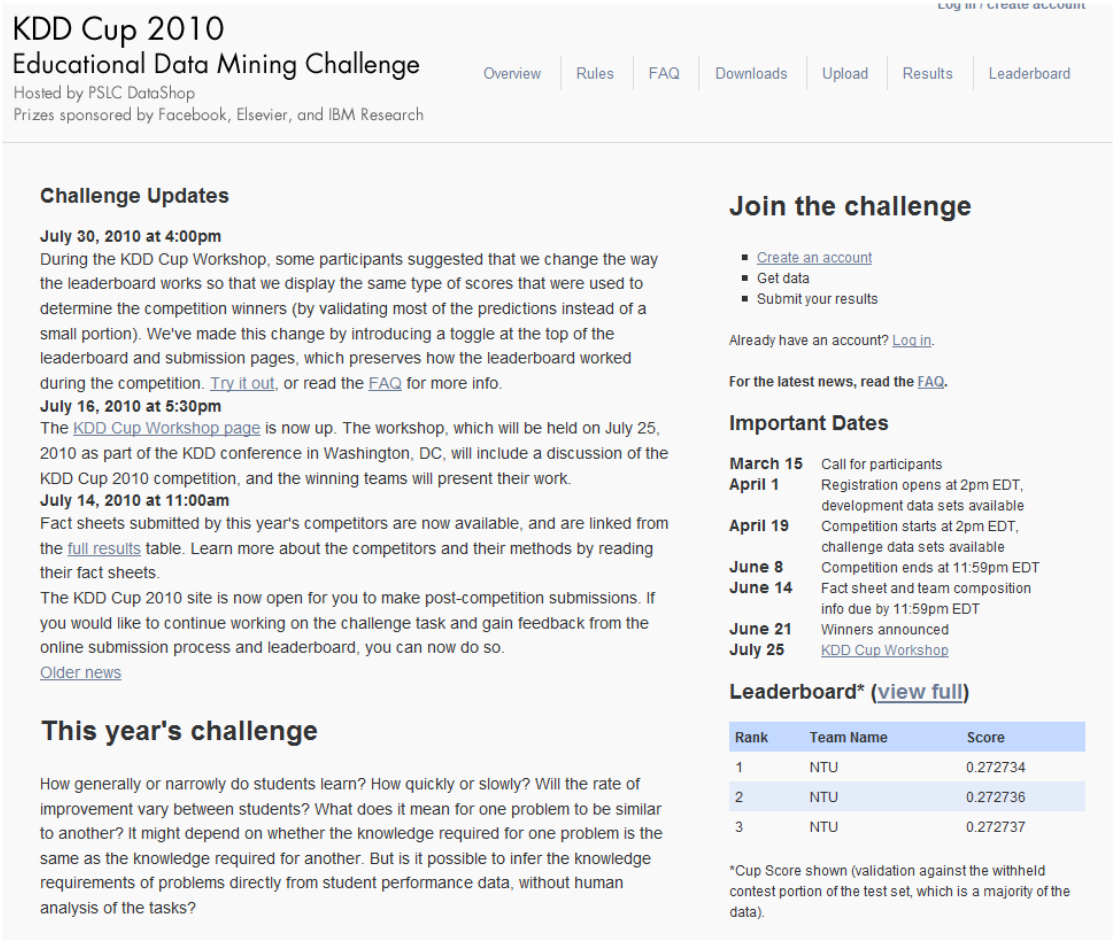
EduMining Blog

* http://blog.edumining.info/



KDD Cup 2010

* https://pslcdatashop.web.cmu.edu/KDDCup/



KDD Cup 2010
Educational Data Mining Challenge
Hosted by PSLC DataShop
Prizes sponsored by Facebook, Elsevier, and IBM Research

Overview | Rules | FAQ | Downloads | Upload | Results | Leaderboard

Challenge Updates

July 30, 2010 at 4:00pm
During the KDD Cup Workshop, some participants suggested that we change the way the leaderboard works so that we display the same type of scores that were used to determine the competition winners (by validating most of the predictions instead of a small portion). We've made this change by introducing a toggle at the top of the leaderboard and submission pages, which preserves how the leaderboard worked during the competition. Try it out, or read the FAQ for more info.
July 16, 2010 at 5:30pm
The KDD Cup Workshop page is now up. The workshop, which will be held on July 25, 2010 as part of the KDD conference in Washington, DC, will include a discussion of the KDD Cup 2010 competition, and the winning teams will present their work.
July 14, 2010 at 11:00am
Fact sheets submitted by this year's competitors are now available, and are linked from the full results table. Learn more about the competitors and their methods by reading their fact sheets.
The KDD Cup 2010 site is now open for you to make post-competition submissions. If you would like to continue working on the challenge task and gain feedback from the online submission process and leaderboard, you can now do so.
Older news

This year's challenge

How generally or narrowly do students learn? How quickly or slowly? Will the rate of improvement vary between students? What does it mean for one problem to be similar to another? It might depend on whether the knowledge required for one problem is the same as the knowledge required for another. But is it possible to infer the knowledge requirements of problems directly from student performance data, without human analysis of the tasks?

Join the challenge

- Create an account
- Get data
- Submit your results

Already have an account? Log in.

For the latest news, read the FAQ.

Important Dates

March 15	Call for participants
April 1	Registration opens at 2pm EDT, development data sets available
April 19	Competition starts at 2pm EDT, challenge data sets available
June 8	Competition ends at 11:59pm EDT
June 14	Fact sheet and team composition info due by 11:59pm EDT
June 21	Winners announced
July 25	KDD Cup Workshop

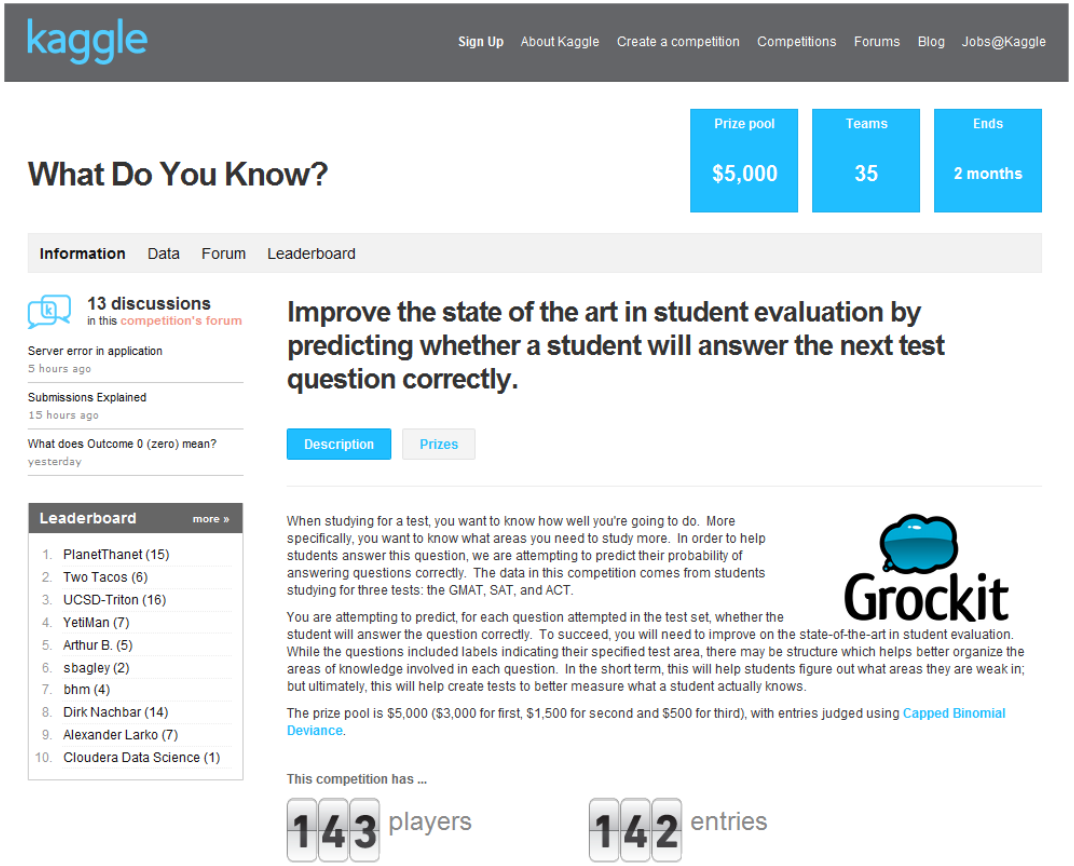
Leaderboard* (view full)

Rank	Team Name	Score
1	NTU	0.272734
2	NTU	0.272736
3	NTU	0.272737

*Cup Score shown (validation against the withheld contest portion of the test set, which is a majority of the data).

Kaggle competition

* http://www.kaggle.com/c/WhatDoYouKnow



The 10 Ways Data Mining Is About To Change Education

* http://edudemic.com/2012/08/data-education-evolutiong/

The 10 Ways Data Mining Is About To Change Education

Added by Katie Lepi on 2012-08-17

Me gusta 20 +1 3 Share 3

The following is a cross-post from our content partners at Online Universities:

Data mining, for better or worse, is having a major impact on numerous facets of American life. While many of the changes have been related to business, especially online business, education is also tapping into the power of data mining in a big way.

Much like Netflix and Amazon use consumer data to recommend products and tailor customer experiences, colleges are using student data to help recruit students, offer them career advice, or even to help them excel in their courses.

While the practice has its critics, many of whom believe it's an invasion of privacy and creates a watered-down, prescriptive education system, there is no doubt that the applications and the impact of the data mining will grow in the coming decade. These are just a few of the ways that dating mining will transform higher ed in the coming years, whether students and teachers like it or not.

It will change how students work together.

Data mining is already changing how students learn and collaborate in their college courses.

Coursera: Big Data in Education

* https://www.coursera.org/course/bigdata-edu