

Universidad Politécnica de Valencia  
Departamento de Ingeniería de Sistemas y Automática  
Master Universitario en Automática e Informática Industrial



---

TRABAJO DE INVESTIGACIÓN

SISTEMA TUTOR INTELIGENTE ADAPTATIVO  
PARA LABORATORIOS  
VIRTUALES Y REMOTOS

---

Verónica Amela Tarongí

Tutores: José Luís Díez Ruano y Marina Vallés Miquel

Valencia, Julio 2010



# Agradecimientos

A mi madre, a mi hermana Patricia y a mi abuela, les agradezco el poder contar siempre con ellas, en los momentos difíciles, son las que me ayudan y aconsejan para seguir adelante. A mi padre, que aunque no esté con nosotras, es el que nos ha inculcado el amor por los libros y el estudio y sé que estaría muy orgulloso de nuestras trayectorias.

A Juan, que siempre con buen humor y una sonrisa, ha aguantado mis neuras en esta etapa final del proyecto, haciendo que cada día sea único y especial a su lado.

A mis directores del Proyecto, José Luis y Marina, sin ellos aún estaría perdida esbozando esta tesina. Ha sido muy gratificante trabajar juntos, me han transmitido su saber y me han guiado durante todo el proceso, corrigiéndome y mostrándome alternativas y soluciones a cada traba que me ha surgido.

A María Luisa, profesora de la Universidad Pública de Navarra, por facilitarme el Cuestionario de Metas para Adolescentes y ofrecerme toda su ayuda en temas de pedagogía. A Fátima, por permitirme utilizar sus bases de datos y estar siempre dispuesta a responder las consultas que me han surgido.

A los profesores del Master en Automática e Informática Industrial y en concreto, a los compañeros que he tenido en las diferentes asignaturas, de los cuales he aprendido mucho después de horas codo con codo haciendo prácticas de laboratorio.

Y por último, a mis amigos, que indirectamente también han sufrido el tiempo que no les he podido dedicar mientras elaboraba el proyecto, muchas gracias por el apoyo y los ánimos recibidos.



# Resumen

Los Sistemas Tutores Inteligentes (ITS) ofrecen un entorno personalizado que guía al estudiante de forma automática, permitiéndole una puesta en práctica de sus conocimientos y habilidades más efectiva que con las clases tradicionales. Por otro lado, esta metodología de enseñanza basada en las facilidades que ofrece Internet, da mayor libertad de estudio a los alumnos, pudiendo elegir cuándo y dónde se conectan a las clases. Además, los Sistemas de Gestión de Aprendizaje (LMS) o plataformas que alojan los ITS almacenan las unidades didácticas del curso y la información sobre el aprendizaje de los alumnos de forma estandarizada permitiendo su reutilización para otros cursos.

Dentro de este contexto, en el presente trabajo se propone un nuevo enfoque para el diseño de un ITS adaptativo aplicado a los estudios de Ingeniería de Sistemas y Automática y a las prácticas de laboratorios virtuales y remotos. El modelo del sistema se obtiene de la aplicación de la teoría básica de control (realimentación) en combinación con técnicas de minería de datos que extraen y analizan la información de las interacciones del alumno con la plataforma o LMS en su proceso de enseñanza y aprendizaje.

También en este trabajo se desarrolla un nuevo algoritmo de agrupamiento basado en el algoritmo *fuzzy k-prototypes*. Esta técnica de minería de datos permite el agrupamiento de datos mixtos (cualitativos y cuantitativos) correspondientes a los atributos que definen a los tipos de alumnos (estilo de aprendizaje, motivación, autoestima, notas, tiempos de estudio, rendimiento, etc.). Con la idea de diseñar un sistema de control adaptativo rápido, con baja complejidad computacional, se ha optado por agrupar a los alumnos en función de sus características o atributos y así no tener que proponer tantas rutas de aprendizaje distintas como alumnos se tenga en el curso.

Con el objetivo de evaluar el agrupamiento obtenido, se ha implementado una estrategia de validación de clusters para datos mixtos basada en el criterio de varianza de Sugeno y Yasukawa. Para medir la efectividad del algoritmo propuesto, se han realizado experimentos de agrupamiento supervisado utilizando conjuntos de datos del repositorio UCI.



# Índice General

1	<b>Introducción y objetivos.....</b>	<b>1</b>
2	<b>Introducción a los Sistemas Tutores Inteligentes .....</b>	<b>3</b>
2.1	Arquitectura de los Sistemas Tutores Inteligentes.....	3
2.2	Metodología.....	5
2.3	Conclusiones.....	6
3	<b>Introducción a la Minería de Datos .....</b>	<b>7</b>
3.1	Aplicación de la minería de datos a los ITS .....	7
3.2	Definición de los objetivos y metodologías de la minería de datos .....	8
3.3	Estado del arte de la minería de datos en los ITS .....	11
	Redes de Bayes .....	11
	Árboles de decisión .....	11
	Razonamiento basado en casos .....	12
	Redes neuronales .....	12
	Algoritmos genéticos.....	13
	Lógica borrosa .....	13
	Sistema multiagentes .....	14
	Patrones secuenciales .....	15
	Reglas de asociación.....	15
	Agrupamiento .....	16
3.4	Herramientas informáticas para la minería de datos.....	16
3.5	Conclusiones.....	17
4	<b>Modelo del Alumno .....</b>	<b>19</b>
4.1	Estilos de aprendizaje .....	19
4.2	Nivel de Conocimiento .....	22
4.3	Conclusiones.....	24
5	<b>Laboratorios Virtuales y Remotos .....</b>	<b>27</b>
5.1	Características.....	27
5.2	Laboratorios Remotos.....	29
5.3	Laboratorios Virtuales .....	30
5.4	Conclusiones.....	31
6	<b>Entorno de trabajo .....</b>	<b>33</b>
6.1	Plataformas de código abierto .....	33
	Moodle.....	33
	Dokeos.....	34
	ATutor .....	35
	Ilias .....	36
	Sakai CLE.....	36
	.LRN .....	37
	Drupal .....	38
6.2	Situación actual en la Web .....	39
	Web 2.0: Entornos colaborativos.....	39
	Web semántica.....	40

6.3	Conclusiones.....	41
<b>7</b>	<b>Extracción de información.....</b>	<b>43</b>
7.1	Archivos log.....	43
	Archivos log del servidor .....	43
	Archivos log del cliente.....	44
7.2	Preprocesado de los datos.....	45
7.3	Bases de Datos.....	45
	MySQL.....	45
	PostgreSQL.....	46
	Comparativa y conclusiones.....	46
7.1	Análisis de los Datos .....	47
	Información personal o perfil del usuario.....	48
	Resultados académicos .....	49
	Datos de interacción del usuario con el sistema.....	50
7.4	Conclusiones.....	50
<b>8</b>	<b>Propuesta de metodología para la implementación del ITS adaptativo.....</b>	<b>51</b>
8.1	Selección del LMS.....	51
	Plataforma.....	51
	Modelo del Alumno.....	56
	Modelo del Dominio.....	64
	Modelo del Tutor .....	69
8.2	Bases de Datos.....	74
	Tablas: Definición de las interrelaciones .....	74
	Tablas: Definición de campos .....	79
8.3	Minería de datos .....	81
	Minería de datos 1 (Modelo Alumno Estático).....	83
	Minería de datos 2 (Modelo Alumno Dinámico) .....	84
	Minería de datos 3 (Modelo Alumno Dinámico).....	85
	Minería de datos 4 (Modelo Dominio).....	86
	Minería de datos 5 (Modelo Tutor) .....	88
8.4	Sistema de control: objetivos/resultados .....	90
8.5	Conclusiones.....	91
<b>9</b>	<b>Aplicación del Sistema de Control.....</b>	<b>93</b>
9.1	Estado del arte de las técnicas de agrupamiento para el análisis de datos cuantitativos y cualitativos .....	94
9.2	Fuzzy k-prototypes .....	94
9.3	Fuzzy k-prototypes modificado .....	97
9.4	Resultados experimentales.....	98
9.5	Agrupamiento de los datos de los alumnos de la UPV.....	100
	Índice de validación de los clusters .....	101
	Resultados.....	103
	Sistema borroso experto .....	103
9.6	Conclusiones.....	104
<b>10</b>	<b>Conclusiones y trabajo futuro .....</b>	<b>107</b>
<b>11</b>	<b>Anexos.....</b>	<b>109</b>

1.	Cuestionario sobre el Índice de Estilos de Aprendizaje de Felder y Soloman (IEA) [47] .....	109
2.	Cuestionario de Metas para Adolescentes (CMA) [86] [87] .....	114
3.	Corrección de los factores [86].....	117
4.	Escala de autoestima de Rosenberg (RSES) [87].....	118
5.	Conjunto de datos de los alumnos (pre-procesados 1) .....	119
6.	Conjunto de datos de los alumnos (pre-procesados 2); <b>Error! Marcador no definido.</b>	
12	<b>Bibliografía</b> .....	<b>135</b>



# Capítulo 1

## Introducción y objetivos

El desarrollo en los últimos años de los sistemas informáticos facilita las tareas de cálculo numérico y algebraico en los estudios científicos y de ingenierías pero exige a cambio, que los alumnos estén mejor preparados, adquiriendo la capacidad de familiarizarse y aprender rápidamente estos conocimientos informáticos para poder emplearlos en sus estudios.

Por otro lado, se observa en el mercado un aumento de la demanda de productos cada vez más sofisticados y complejos, lo que conlleva una búsqueda de la productividad y mejora de la calidad de los procesos de fabricación por parte de las industrias, que necesitan de ingenieros de sistemas y automática competentes capaces de hacer frente a estos problemas.

Esta idiosincrasia actual reclama un cambio o transición en el modelo de enseñanza universitaria, de forma que las instituciones educativas puedan ofrecer todos los contenidos del plan de estudios (conceptuales, actitudinales y procedimentales) y también permita a los estudiantes disponer de un aprendizaje adaptativo, más flexible a sus necesidades, ritmo de aprendizaje y aptitudes.

Además, la situación personal o profesional de parte del alumnado les impide el poder asistir a las clases: tanto teóricas como prácticas en el laboratorio, bien por problemas geográficos (de movilidad al centro) como temporales (los horarios fijos preestablecidos de las lecciones y prácticas).

Asimismo las instituciones disponen de un presupuesto ajustado con el que deben hacer frente a la adquisición y mantenimiento de equipos instrumentales y de control para la realización de las prácticas. El coste de los mismos obliga a compartirlos, estableciendo grupos de prácticas con diferentes horarios que dividen a los alumnos por parejas, normalmente.

Con el fin de poder satisfacer las dificultades de asistencia de los alumnos e igualmente, ofrecerles una educación efectiva, es decir, que combine materiales teóricos reforzados con ejercicios prácticos, se introducen los sistemas de enseñanza a distancia basados en Web o sistemas de *e-learning*.

La tecnología del control y gestión a distancia está perfectamente implantada y desarrollada en los sectores industriales, sanitarios, militares, exploración espacial y comercio electrónico, conocido con el nombre de *e-commerce*, pero está en fase de investigación en lo que se refiere a la educación, comenzando a despegar sus aplicaciones y estudios en los últimos años, como demuestran la aparición de congresos y de artículos en revistas y periódicos científicos, lo que pone de relieve su importancia.

Gracias a la difusión de los equipos informáticos y del acceso a Internet, tanto en los hogares como en los centros educativos, permite que la implantación de este recurso sea una realidad, asimismo, los desarrolladores de software y de comunicaciones prevén esta necesidad de comunicación, adaptando la conectividad de las redes mediante pasarelas y protocolos para que los equipos se puedan comunicar a través de Internet.

Por otro lado, destacar que el campo de la educación a distancia de la ingeniería de sistemas y automática, y más concretamente, su práctica en los laboratorios, es un sector idóneo para su desarrollo. Se han observado una serie de problemáticas, en los trabajos desarrollados hasta el momento, que se intentarán salvar con las soluciones propuestas en este trabajo de investigación.

Se ha detectado una falta de estandarización de los laboratorios basados en Web, sobre todo en relación con las conexiones de los equipos, a las que se han buscado soluciones específicas para el caso concreto de estudio. Por otro lado, se han desarrollado proyectos ad-hoc, consistentes en un software de instalación y ejecución en los equipos de los usuarios finales, es decir, no se ha empleado un sistema o plataforma de enseñanza basada en Web genérica, de acceso global para todos los usuarios. Por último, falta el interés de los desarrolladores de plataformas (tanto libres como comerciales) para desarrollar herramientas que faciliten la adaptabilidad de los recursos educativos que ofrecen.

Los sistemas de educación a distancia eran originariamente cursos consistentes en un conjunto de páginas Web estáticas a través de las cuales navegaban los estudiantes. Estos evolucionaron a los Sistemas de Gestión del Contenido de Aprendizaje (LMS), que son plataformas en las que existen diversos canales y espacios de trabajo en los que se facilita el intercambio de información y la comunicación entre los alumnos del curso, produciendo materiales, realizando pruebas, promoviendo discusiones y permitiendo el aprendizaje colaborativo (foros, charlas, áreas de almacenamiento de archivos, servicios de noticias, etc.).

La última tendencia son los Sistemas Educativos Basados en Web Inteligentes y Adaptativos (AIWBES), que van más allá al construir un modelo de los objetivos, preferencias y conocimiento de cada alumno individual y al emplearlo en la interacción del estudiante durante su proceso de aprendizaje, con la intención de adaptarlo a sus necesidades. AIWBES es el resultado de la unión de los Sistemas Tutores Inteligentes (ITS) y los Sistemas Hipermedia Adaptativos (AHS).

Las líneas de investigación actuales aplican las técnicas de minería de datos a los sistemas de enseñanza a distancia basados en Web que acabamos de citar. Estas herramientas inteligentes utilizan métodos de extracción de conocimiento o minería de datos para descubrir información útil sobre los estudiantes y sus patrones de estudio, con el objetivo de guiarlos durante su proceso de aprendizaje con el fin de optimizarlo.

Los objetivos de este trabajo de investigación son dos: por una parte, desarrollar una nueva generación de ITS adaptativos aplicados a laboratorios virtuales (en los que se proponen simulaciones de sistemas reales) y remotos (los que proporcionan el control remoto de sistemas reales). Y por otro lado, a partir de los métodos y técnicas de minería de datos, crear un nuevo algoritmo de agrupamiento borroso (*fuzzy clustering*), que resuelva los inconvenientes actuales: no se está teniendo en cuenta debidamente la incertidumbre de los datos y las conclusiones que proporciona el sistema a los estudiantes y tutores, no son todo lo buenas que se desearía.

Este trabajo está estructurado de la siguiente manera: en primer lugar se introducen las características y componentes que integran los sistemas tutores inteligentes (modelo pedagógico, campo de aplicación: laboratorios y modelo del alumno). A continuación se describen las metodologías de minería de datos (redes de bayes, sistemas borrosos expertos, redes neuronales artificiales, etc.), así como los objetivos de la minería de datos (clasificación, agrupamiento, estimación, modelado de dependencias, visualización y descubrimiento de reglas). En el capítulo 4 se detallan los métodos y técnicas para desarrollar el modelado del alumno y de su comportamiento de aprendizaje. En el siguiente capítulo se citan las características de los laboratorios virtuales y remotos y a continuación, se describen las diferentes plataformas disponibles en el mercado para implementar el ITS. Seguidamente, se propone la metodología que se seguirá para la implementación del sistema presentado. Por último, se finaliza el trabajo con las conclusiones y las líneas del trabajo futuro, así como con una sección que incluye la bibliografía consultada y los anexos.

## Capítulo 2

# Introducción a los Sistemas Tutores Inteligentes

En este capítulo se presentan los componentes que constituyen la arquitectura de los Sistemas Tutores Inteligentes (ITS) y se detallan las fases a seguir o la metodología que define los pasos para crear un ITS.

La tutorización personalizada es el modo más efectivo para enseñar, pero resulta una tarea laboriosa y costosa, sin embargo, los sistemas tutores inteligentes son sistemas informáticos de aprendizaje personalizado, que no requieren la intervención de tutores humanos y reducen el coste, al automatizar la selección de los materiales del curso, su presentación y la evaluación de los estudiantes.

Los ITS son sistemas de inteligencia artificial, más concretamente, sistemas expertos que simulan las características de un tutor humano. Aparecieron a finales de la década de 1970, pero no fue hasta los años 90 que creció su popularidad y con ella surgieron congresos y publicaciones relacionados con esta temática (*International Conference on Intelligent Tutoring Systems (ITS)*, *World Conference on Educational Multimedia, Hypermedia and Telecommunications (ED-MEDIA)*, *International Conference on Artificial Intelligence in Education (AIED)*, etc.)

### 2.1 Arquitectura de los Sistemas Tutores Inteligentes

Los sistemas tradicionales de enseñanza presentan hechos y conceptos a los estudiantes seguidos por pruebas con una serie de cuestiones. Estos métodos son efectivos para exponer grandes cantidades de información y verificar su adquisición. Sin embargo, hay conocimientos que los estudiantes pueden memorizar pero no emplear correctamente. Los ITS suplen esta carencia al utilizar simulaciones y otros entornos de aprendizaje que requieren que los alumnos pongan en práctica sus conocimientos y habilidades. Estos entornos activos y personalizados les ayudan a retener y aplicar los conocimientos y destrezas de forma más efectiva que los sistemas clásicos de enseñanza.

Los sistemas ITS, con el fin de proporcionar ayuda, guía y realimentación a los estudiantes, se basan normalmente en tres tipos de conocimiento, organizado en tres módulos o modelos independientes [1]:

- **Modelo dominio o experto:** contiene la descripción del conocimiento o comportamiento que representa el dominio o campo de la enseñanza, es decir, los recursos educativos del curso (representado por un sistema experto o modelo cognitivo). Permite al ITS comparar las acciones y elecciones del estudiante con las del experto, con el objetivo de evaluar lo que el usuario conoce y desconoce.
- **Modelo alumno:** evalúa la información o actuación de cada alumno para determinar su conocimiento, habilidad perceptual y capacidad de razonamiento. El modelo del alumno también incluye sus carencias y malas concepciones de aprendizaje.

- **Modelo tutor:** codifica los métodos de enseñanza que son apropiados para el dominio objetivo y el estudiante. Es el motor de ejecución del sistema adaptativo. En función del conocimiento, nivel de experiencia y estilos de aprendizaje de los alumnos, el modelo tutor selecciona la intervención educativa más adecuada. Ésta se realiza comparando el modelo del estudiante con los resultados esperados del modelo dominio, las discrepancias observadas son señaladas al modelo tutor que toma una acción correctiva acorde.

Hay autores que amplían esta clasificación, añadiendo dos módulos más que complementan e interaccionan con los anteriores [2]:

- **Módulo interfaz:** suministra el medio para la interacción del estudiante con el ITS, normalmente a través de un interfaz gráfico de usuario y a veces, mediante una simulación del dominio con la tarea que el estudiante está aprendiendo (i.e. el control de una planta industrial)
- **Modelo pedagógico:** proporciona un modelo o conjunto de reglas del proceso de enseñanza que colabora con el modelo tutor (i.e. da información sobre cuándo hacer un repaso, presentar un nuevo tema, cómo presentarlo, etc.)

En la figura 2.1 se muestra el sistema de control de un ITS, siendo el proceso a controlar el aprendizaje del alumno. El alumno presenta un estilo de aprendizaje y unos conocimientos iniciales que se modifican dinámicamente a partir de la interacción con el curso. Esta información del alumno se extrae y se almacena en una base de datos (BBDD) para su posterior análisis mediante técnicas de minería de datos. La minería de datos realimenta al controlador del ITS, adaptando los modelos que conforman el Sistema de Gestión del Aprendizaje (LMS) o plataforma que almacena y gestiona el curso. El LMS está estructurado en tres modelos o módulos: Dominio (que incluye al módulo interfaz), Alumno y Tutor (que incluye el modelo pedagógico). Por su parte, el profesor supervisa el proceso y define los contenidos o currículum del curso. Por último, se verifica y controla el buen funcionamiento del ITS y la adecuación del mismo al alumno, mediante la comparación de los resultados obtenidos por el alumno con respecto a los objetivos que se plantean en el curso.

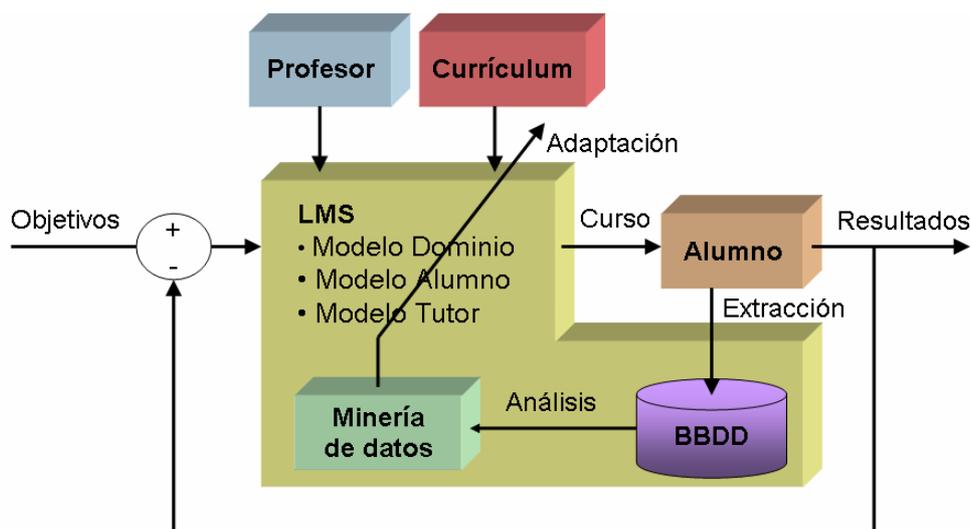


Figura 2.1: Componentes de un Sistema Tutor Inteligente

## 2.2 Metodología

Se ha modificado la metodología que se propone en [2] para el desarrollo de los sistemas tutores inteligentes con el fin de adecuarla al sistema ITS definido en la figura 2.1. Como se puede apreciar es una metodología cíclica, que consta de las siguientes etapas:

- **Construcción del curso:** El profesor o diseñador construye el Curso Adaptativo Hipermedia proporcionando la información del modelo dominio, el modelo pedagógico y el módulo interfaz. Se emplea una herramienta autor para facilitar esta tarea (que puede ser una herramienta de tipo comercial como *ToolBook*, *Director*, etc. o una herramienta específica diseñada para un curso en concreto y que normalmente viene proporcionada por la plataforma). La información restante, modelo del tutor y modelo del alumno normalmente vienen dadas o adquiridas por el propio sistema. Una vez que el profesor y la herramienta autor finalizan la elaboración del curso, entonces, el contenido del mismo puede ser publicado en el servidor Web para que los alumnos puedan utilizarlo de forma remota.
- **Ejecución del curso:** El estudiante se conectará al sitio que alberga la aplicación, usando un navegador Web, para realizar el curso. Durante esta etapa, se recoge la información utilizada, de forma transparente para el alumno y se almacena en el servidor en una base de datos (ficheros históricos o *logs*).
- **Aplicación de técnicas de minería de datos:** Consiste en la aplicación de técnicas de descubrimiento de conocimiento sobre los datos de utilización recogidos en la etapa anterior de ejecución del curso. Esta etapa se añadiría a la etapa de evaluación del curso de la metodología clásica. El objetivo es procesar toda la información almacenada en el sistema, previo preprocesado y colocación en un sistema de gestión de bases de datos que garantice una manipulación más rápida de dicha información. Una vez transferidos los datos, el diseñador del curso puede aplicar los algoritmos de minería de datos y descubrir relaciones importantes entre éstos.
- **Mejora del curso:** El sistema de control, ayudado por la información que se le suministra en forma de relaciones importantes descubiertas entre los datos, lleva a cabo las modificaciones que crea más apropiadas para mejorar el rendimiento del curso (orientado a solventar carencias, corregir problemas o mejorar determinados aspectos del mismo). Estas modificaciones pueden afectar al contenido del curso, su estructura, el interfaz gráfico, etc. y se realizan con la ayuda de la herramienta autor.

El proceso de ejecución, aplicación y mejora se repite continuamente mientras dure el curso.

Se diferencia de la metodología propuesta en [2] en el agente encargado de mejorar el curso, dónde se ha sustituido la tarea manual de actualización del profesor por un sistema de control adaptativo autónomo bajo la supervisión del profesor.

## 2.3 Conclusiones

De la metodología para la construcción de los ITS, investigada y desarrollada hasta el momento, se puede observar que el modelo del dominio es limitado y estructurado, debiéndose modificar (en el proceso de mejora) por el propio profesor o diseñador manualmente, por lo que falla la automatización de la adaptabilidad del sistema en este aspecto, siendo uno de los puntos de interés de este trabajo de investigación.

Se observa que en los ITS el proceso de aprendizaje no está predefinido por el diseñador, sino que se forma y construye a partir de la información que se captura de los estudiantes y que se almacena en bases de datos. El análisis de las bases de datos y la toma de decisiones asociada (modelo alumno y modelo tutor), con el fin de ayudar al alumno a obtener óptimos resultados, se realiza gracias a la ayuda de las técnicas de descubrimiento de conocimiento o minería de datos, que es el objetivo del siguiente capítulo.

# Capítulo 3

## Introducción a la Minería de Datos

Se presenta a continuación el uso que tiene la minería de datos para los sistemas tutores inteligentes. Cada uno de estos objetivos de la minería de datos se puede llevar a cabo mediante diversos métodos. Seguidamente se clasifican y describen los métodos más empleados en minería de datos y se presenta el estado de la investigación en esta área aplicada al campo de los ITS. A continuación se muestran las herramientas informáticas disponibles para lograr los objetivos de la minería de datos y se finaliza el capítulo con las líneas más actuales de investigación, que son las que seguirá este trabajo de investigación.

### 3.1 Aplicación de la minería de datos a los ITS

Los algoritmos para minería de datos llevan a cabo tareas descriptivas, como el descubrimiento de relaciones o el reconocimiento de patrones, o tareas predictivas, como la clasificación o el ajuste de modelos que permiten predecir el comportamiento [3].

El conocimiento descubierto se filtra para la toma de decisiones y se emplea para realimentar el ciclo del proceso de enseñanza y aprendizaje del sistema educativo a distancia, como se aprecia en la figura 3.1.

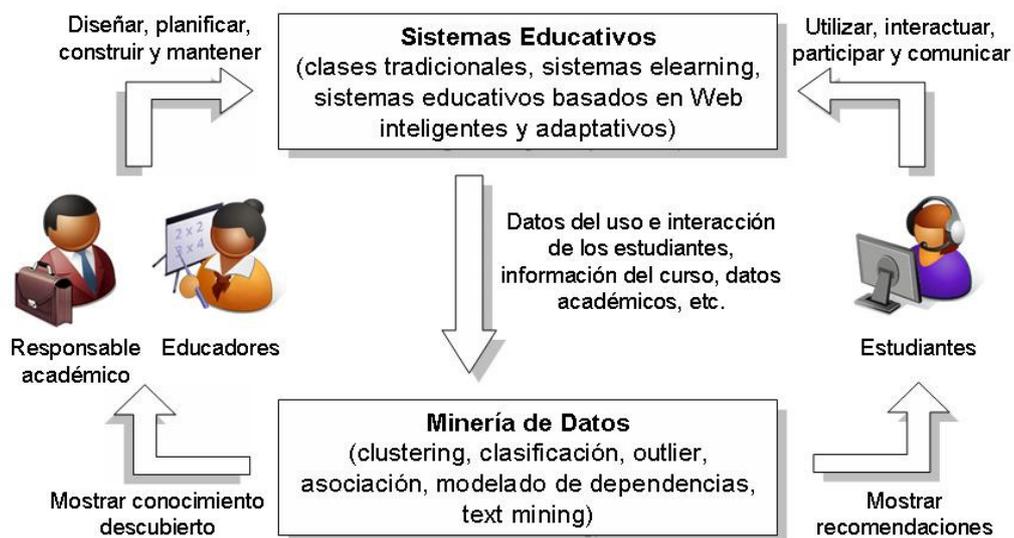


Figura 3.1: El ciclo de aplicación de la minería de datos a los sistemas educativos

La aplicación de técnicas de minería de datos en educación se puede ver desde dos puntos de vista u orientaciones distintas [4], en función de si va orientado hacia los tutores o desarrolladores de los ITS, el objetivo es ayudarlos a mejorar el funcionamiento o rendimiento de los sistemas a partir de la información de utilización de los alumnos, o si en cambio, está encaminado hacia los alumnos, el objetivo es ayudarlos en su proceso de aprendizaje.

Las principales aplicaciones se pueden ver sintetizadas en la tabla 3.1 con las dos visiones u orientaciones que se han expuesto.

Tutores / Desarrolladores	Estudiantes
<ul style="list-style-type: none"> <li>- Obtener una mayor realimentación de la enseñanza</li> <li>- Conocer el proceso de aprendizaje</li> <li>- Evaluar a los estudiantes por sus patrones de navegación</li> <li>- Reestructurar los contenidos del curso para personalizarlo</li> <li>- Clasificar a los estudiantes en grupos</li> <li>- Detectar comportamientos irregulares de aprendizaje</li> </ul>	<ul style="list-style-type: none"> <li>- Sugerir buenas experiencias de aprendizaje</li> <li>- Adaptación del curso según el proceso de aprendizaje</li> <li>- Ayudar a los estudiantes con sugerencias y atajos</li> <li>- Recomendar caminos más cortos y personalizados</li> </ul>

*Tabla 3.1:* Aplicaciones de las técnicas de minería de datos

### 3.2 Definición de los objetivos y metodologías de la minería de datos

La minería de datos es el proceso de descubrimiento de conocimiento para encontrar información no trivial, previamente desconocida y potencialmente útil de grandes repositorios de datos [5]. La minería de datos es en realidad uno de los pasos que comprenden el proceso de Descubrimiento de Conocimiento en Bases de Datos (KDD), que está compuesto por [4]:

- **Preprocesamiento:** Consiste en la recogida o extracción de los datos, limpieza de datos, discretización, selección de los atributos e integración de datos.
- **Minería de datos:** Consiste en la selección de los algoritmos de minería de datos a utilizar y la aplicación de dichos algoritmos sobre los datos.
- **Postprocesamiento:** Consiste en la interpretación, evaluación de los resultados obtenidos y la utilización del conocimiento descubierto.

Los principales objetivos o tareas de la minería de datos son [3]:

- **Análisis de secuencias:** Tratamiento de secuencias de datos, con el fin de determinar patrones y pautas de comportamiento.
- **Clasificación:** Consiste en clasificar los objetos, entre un rango de categorías.
- **Agrupamiento:** Se trata de agrupar los objetos según similitud de características, formando conjuntos o clases. A diferencia de la clasificación, en este caso no hay una división previa del espacio en categorías o clases.
- **Asociación:** El objetivo es el de descubrir relaciones ocultas entre los objetos, o incluso entre los propios atributos de los objetos, de los cuales se pueden extraer una base de reglas, con una estructura condicional (*Si A es B y C es D, Entonces E es F*).

- **Modelado de dependencias:** Describe relaciones significativas de dependencia entre variables. Hay dos tipos de dependencias: la estructural, que especifica las variables que son localmente dependientes y la cuantitativa, que da una medida numérica de cómo de fuertes son las dependencias.
- **Predicción:** Obtención o entrenamiento de modelos de predicción, con el fin de validar hipótesis de comportamiento preconcebidas.
- **Regresión:** A partir de muestras de datos, estima una función o modelo que pueda establecer una relación de dependencia de ciertas variables respecto de otras, con el fin de predecir valores a partir de nuevos datos.
- **Sumarización o agregación:** Permite calcular valores agregados, que no son los datos directos registrados sino datos derivados de ellos, con el objetivo de generar descripciones globales (que pueden ser cualitativas) de conjuntos de datos.
- **Visualización del modelo:** Consiste en adecuar y reinterpretar los datos mediante gráficos para que sean visualmente comprensibles y se puedan extraer conclusiones de los mismos.
- **Análisis exploratorio de datos:** Dado un conjunto de datos del cual se desconocen sus posibles interdependencias y relaciones de similitud, trata de identificar patrones de forma visual y sin ninguna estructura de búsqueda o semejanza preconcebida.

La minería de datos es un área multidisciplinar donde convergen diferentes paradigmas de computación o métodos que tienen la finalidad de llevar a cabo algunos de los objetivos anteriores [3], [6] y [7]:

- **Redes de Bayes:** Consisten en representaciones gráficas de distribuciones de probabilidad junto a posibles dependencias entre ellas, en donde los nodos son los atributos o dimensiones de los objetos y se le asocia una función de probabilidad condicional, que relaciona los nodos con sus antecesores y predecesores.
- **Árboles de decisión:** Son estructuras de búsqueda consistentes en nodos y ramificaciones: cada vez que se alcanza un nodo, se produce una decisión que genera dos o más alternativas o ramas a seguir. De esta forma se consigue una estructuración jerarquizada de los datos. Una vez el árbol está construido, la clasificación de datos consiste en partir de la rama principal e ir decidiendo en cada nodo a qué rama tiene una mayor pertenencia el objeto o dato en cuestión.
- **Razonamiento basado en casos:** Son métodos que tratan de resolver un problema haciendo uso de experiencias y resultados anteriores y conocidos. Todos los casos de distintos problemas planteados son descritos y almacenados, cuando un nuevo problema se plantea, se buscan situaciones similares y se aplica la solución almacenada, añadiendo a la base de casos la descripción y solución de este problema para futuras referencias.
- **Métodos bayesianos:** Son métodos para realizar inferencias a partir de los datos, induciendo modelos probabilísticos que después serán usados para razonar sobre nuevos valores observados. El teorema de Bayes permite actualizar la creencia que tenemos en un suceso o conjunto de sucesos a la luz de los nuevos datos u observaciones.
- **Redes neuronales:** Están basadas en el modelo de computación paralela del cerebro, de forma que el valor de salida depende de una función matemática que toma como valor de entrada la suma ponderada de todas las entradas. En función del objetivo a conseguir se usan distintos tipos de funciones matemáticas, como funciones suma, producto, exponenciales o de activación (perceptrón, sigmoidales

o de tangente hiperbólica), cuyo valor de salida varía entre dos estados en función de si la combinación de los valores de entrada sobrepasa cierto límite.

- **Máquinas de vectores soporte (SVM):** Inducen separadores lineales o hiperplanos en espacios de características de muy alta dimensionalidad (introducidos por funciones de núcleo o kernel) con un sesgo inductivo muy particular (maximización del margen). El aprendizaje de las SVM representa un problema de optimización con restricciones que se puede resolver usando técnicas de programación cuadrática.
- **Algoritmos genéticos:** Son algoritmos que imitan los mecanismos de selección natural para una mejor adaptación al entorno. Partiendo de una solución, unas relaciones o un modelo, se generan aleatoriamente posibles soluciones y se computa su eficiencia en la resolución del problema. Las mejores respuestas son seleccionadas y se combinan entre sí (reproducción) para obtener nuevas combinaciones, que mejorarán los resultados anteriores. El proceso finaliza cuando no se detectan mejoras o hasta que se cumple cierto valor de umbral de eficacia, intercalando recombinaciones y mutaciones aleatorias para evitar estancamientos.
- **Lógica borrosa:** Es una herramienta para la interpretación de datos, que pueden ser imprecisos, intervalares, no numéricos, inexactos o con incertidumbres, y que sin embargo aportan información de forma intuitiva. La lógica borrosa traduce estos datos a un conjunto de pertenencias borrosas, en donde cada dato pertenece a todas las áreas o rangos de pertenencia definidos en diferentes porcentajes.
- **Sistema multiagente:** Consiste en múltiples entidades autónomas inteligentes (agentes) que interactúan teniendo distinta información y/o intereses, con el fin de resolver problemas que resultarían difíciles o imposibles para un único agente.
- **Patrones secuenciales:** Se trata de la extracción de patrones frecuentes relacionados con el tiempo u otro tipo de secuencia.
- **Reglas de asociación:** Son métodos que descubren relaciones entre los datos, produciendo como resultado una base de reglas antecedente-consecuente. La lógica borrosa se usa mucho en estos métodos, consiguiendo reglas que son más comprensibles. También se usan combinaciones con redes neuronales para extracción de reglas y validación.
- **Modelos de regresión:** Consiste en explicar el modelo de una variable a partir del conocimiento de otras. Hablamos de modelo de regresión cuando la variable de respuesta y las variables explicativas son todas ellas cuantitativas.
- **Agrupamiento:** Se trata de encontrar grupos entre un conjunto de individuos. Los métodos más populares son el algoritmo K medias y los métodos jerárquicos. En el primero se realiza un agrupamiento por vecindad en el que se parte de un número determinado de prototipos y de un conjunto de ejemplos a agrupar sin etiquetar. Las regiones se definen minimizando la suma de las distancias cuadráticas entre cada vector de entrada y el centro de su correspondiente clase, representado por el prototipo correspondiente. Los métodos jerárquicos se basan en la construcción de un árbol (llamado dendrograma) en el que las hojas son los elementos del conjunto de ejemplos, y el resto de los nodos son subconjuntos de ejemplos que pueden ser utilizados como particionamiento del espacio.

### 3.3 Estado del arte de la minería de datos en los ITS

La meta de aplicar las técnicas de minería de datos a los sistemas e-learning es la de poder obtener clasificaciones y agrupamientos, por ejemplo, agrupar a los usuarios por su comportamiento de navegación, agrupar las páginas por su contenido, tipo o acceso, agrupar los comportamientos de navegación similares, etc.

A continuación se describe el estado del arte de los trabajos de aplicación de minería de datos a los sistemas ITS y AIWES, clasificando las investigaciones realizadas en función de la técnica de minería de datos que se ha aplicado.

#### Redes de Bayes

En [9] se analizan los datos obtenidos de cursos basados en sistemas e-learning y utilizan técnicas de agrupamiento similares al modelo probabilístico de Bayes para descubrir patrones que reflejan comportamientos de los usuarios y dar soporte a la tutoría en comunidades de aprendizaje virtual.

Siguiendo con el mismo enfoque, en [10] se propone un modelo híbrido que combina las técnicas de minería de datos (la categorización y agrupamiento de los estudiantes) y de aprendizaje de máquinas para la construcción de una red bayesiana para describir el proceso de aprendizaje de los estudiantes. Su objetivo es clasificar a los estudiantes para poder ofrecerles diferentes itinerarios dependiendo de sus habilidades y otras características.

En el estudio [11] se detecta en línea irregularidades en perfiles de procesos de aprendizaje a partir de los tiempos de respuesta de los contenidos de sistemas e-learning. El método de detección está basado en distribuciones predictivas bayesianas, con el fin de asistir a los alumnos que presentan alguna dificultad.

En [12] se utiliza una red bayesiana de correlaciones entre actitudes, percepciones y comportamientos de los estudiantes, para poder inferir las actitudes positivas y negativas de los estudiantes a partir de sus comportamientos, con el objetivo de poder responder de forma adaptada a las necesidades de los estudiantes.

#### Árboles de decisión

En [13] se presenta la arquitectura de un ITS que incorpora árboles de decisión y emparejamiento borroso como estrategia educativa. El ITS contiene los siguientes componentes: sistema de extracción de información, sistema de análisis del perfil del alumno, sistemas de emparejamiento y actualización de la estrategia, registro de archivos log Web (número de alumnos, páginas visitadas, opciones del menú y banners clicados, las rutas de navegación más comunes, los archivos más descargados, etc.), base del modelo del alumno y base de la estrategia educativa. El sistema analiza los modelos de aprendizaje y aplica diferentes estrategias educativas para satisfacer los requisitos personales.

En [14] se presenta una herramienta de visualización para el seguimiento de las deficiencias de comprensión del alumnado. Se recurre a un árbol con tipo de datos abstractos (ADT), construido de los conceptos tratados en un laboratorio, clase o curso. Cada nodo puede asociarse con diferentes entidades como el comportamiento del alumno, la actuación en clase o el desarrollo del laboratorio. Con el uso de la herramienta, el profesor puede ayudar al alumno a descubrir conceptos que necesitan

mayor apoyo y el alumno puede observar conceptos a los que debe dedicar mayor tiempo de estudio.

En [15] se estudia cómo estimar la motivación del alumno en el proceso de aprendizaje a partir de los archivos log que registran la interacción del alumno en un entorno de aprendizaje basado en Web. Se construyen árboles de decisión para la minería de datos. Los resultados sugieren que el tiempo empleado en la lectura es un factor importante para la predicción de la motivación así como el resultado de los controles, que es un indicador relevante del nivel de motivación.

En [16] se presenta un entorno de aprendizaje educativo basado en Web, llamado Lecomps, que automatiza la construcción de rutas de aprendizaje personalizadas al estilo de aprendizaje del alumno y adaptadas a su conocimiento, en base al mapeado de grafos and-or cuyos vértices son Contenidos de aprendizaje (LC) y Objetivos de aprendizaje (LO). Se construye el curso personalizado como una selección de LC basada en el conocimiento meta y en el de partida del alumno. La optimización de las distancias del grafo entre los LC que se deben producir y cada uno de los LO que se consume, da lugar al orden de presentación y la distribución de los contenidos.

### **Razonamiento basado en casos**

En [17] se utilizan algoritmos para clasificar a los estudiantes en clases dependiendo de sus acciones y de descubrir reglas de asociación entre diferentes puntos de conocimiento, utilizan razonamiento basado en casos para personalizar la interacción. Los casos son concretamente un conjunto de preguntas y sus correspondientes respuestas, de forma que ante una pregunta de un alumno, se seleccionan las N preguntas más similares y se le proporciona las respuestas al alumno.

En [18] se propone la generación de la estructura de enlaces de un entorno de aprendizaje adaptativo basado en Web utilizando los ficheros log de la interacción de los estudiantes. En concreto descompone los ficheros log en episodios de aprendizaje y utiliza razonamiento basado en casos para ofrecer a los alumnos una ayuda contextual con una estructura de enlaces adaptados del curso.

En [19] se define un sistema e-learning personalizado, con un enfoque que incluye el razonamiento basado en casos para analizar la evaluación del alumno en el curso y algoritmos genéticos que se emplean para construir el itinerario de aprendizaje, secuenciando el currículo de forma personalizada, en base a los requisitos individuales del alumno, con el fin de ayudarlo con un aprendizaje más efectivo.

En [20] se propone un enfoque de razonamiento basado en casos para sistemas educativos basados en Web, con la idea de adaptar la experiencia de aprendizaje de cada alumno a otros alumnos con las mismas características. Se emplea la lógica borrosa para adaptar los contenidos y contextos del e-learning en función de los estilos de aprendizaje y las necesidades individuales de los alumnos y agentes pedagógicos para el guiado y la propuesta de actividades y recursos a los alumnos en el proceso de aprendizaje.

### **Redes neuronales**

En [21] se presenta un modelo de redes neuronales que evalúa automáticamente el nivel de conocimiento meta cognitivo del alumno observando su comportamiento online de navegación. El modelo está implementado a través de una red neuronal multicapa realimentada (MLFF) formada por tres capas, que clasifica los estudiantes a partir de

datos de navegación (páginas visitadas, glosarios y transcripciones consultadas y tiempo empleado).

En [22] se describe un sistema multiagente que busca la adaptabilidad mediante el uso de redes neuronales. La estructura propuesta consta de un sistema gestor del curso y de un sistema inteligente multiagente (MIS) que implementa todos los módulos del ITS. El modelo del estudiante usa lógica borrosa para determinar el grado de conocimiento adquirido por el estudiante en un tiempo dado. Para determinar la evaluación del alumno se emplean reglas borrosas entre la evaluación del conocimiento (calidad de las respuestas y tiempo empleado) y la evaluación del comportamiento (tiempo por página, movimiento de la barra de desplazamiento, participación en chats y foros). Se emplean redes neuronales, concretamente Cuantificación de Vectores de Aprendizaje (LVQ) para implementar la clasificación de patrones (los alumnos) de modo que los agentes pedagógicos puedan adaptar correctamente el currículum. Para facilitar la tarea de los agentes, es básica la representación del conocimiento. Se propone un módulo de conocimiento compuesto de una base de datos de conocimiento (que incluye todo tipo de materiales instructivos) y una base de metadatos (con el fin de indexar el material instructivo, dándole significado y estableciendo las relaciones).

### **Algoritmos genéticos**

En [23] se propone la utilización de algoritmos evolutivos multiobjetivo para el descubrimiento de relaciones importantes a partir de datos de utilización (tiempos de acceso a páginas, aciertos y fallos, y niveles de conocimiento) de los estudiantes. Con el objetivo de poder realizar mejoras tanto del contenido, como de la estructura de los cursos y de su adaptación.

El estudio propuesto en [24] también emplea algoritmos evolutivos para realizar un análisis de asociación de los estudiantes para predecir su rendimiento. Las características de los grupos se extraen de la información de los archivos log del sistema. Se demuestra que una combinación de múltiples clasificadores mejora el proceso de clasificación y la ponderación de los vectores de características en los algoritmos genéticos optimiza la precisión de la predicción.

Siguiendo con el mismo enfoque, en [25] se utiliza agrupamiento de recursos Web valorados y descubrimiento de reglas de asociación mediante algoritmos genéticos para la optimización de la minería de datos con el objetivo de clasificar a los estudiantes basándose en las características extraídas de los ficheros log.

### **Lógica borrosa**

En [26] se desarrolla un ITS basado en la teoría de respuesta a un ítem borroso, donde se recomiendan materiales del curso con niveles de dificultad acordes a las respuestas de realimentación del alumno inciertas o borrosas sobre la comprensión y la dificultad de los materiales propuestos.

En [27] se propone un sistema de diagnóstico de aprendizaje inteligente para un modelo de aprendizaje temático basado en Web, consistente en buscar contenidos relacionados con el tema en Internet y discutir los resultados en un entorno colaborativo. A partir de los archivos log que registran el comportamiento del alumno (tiempo total en la plataforma, la frecuencia de registro de las sesiones, los materiales descargados por los alumnos, los artículos colgados o respondidos, el tiempo utilizado en los grupos

de discusión online, etc.), un sistema de diagnóstico inteligente basado en lógica borrosa, ofrece una guía de aprendizaje para asistir a los alumnos y clasifica la participación online de la clase para el tutor. El sistema también puede predecir los resultados de los alumnos en la evaluación final.

Con un enfoque distinto, en [28] se propone un estudio de reconocimiento de imagen para la captura de imágenes del rostro de los estudiantes mientras están aprendiendo, y el análisis de sus características faciales con el fin de evaluar sus estados afectivos y el grado de atención mediante integrales borrosas y detectar los malos comportamientos de aprendizaje (i.e. somnolencia o distracción).

El sistema de diagnóstico de aprendizaje inteligente descrito en [29] permite, a partir de los archivos log del comportamiento de aprendizaje de los alumnos, el guiado y la asistencia de los alumnos para mejorar su aprendizaje y la evaluación de su participación en clase, prediciendo el éxito de los alumnos. El sistema tiene dos partes: un sistema experto borroso que aconseja al alumno en el proceso de aprendizaje y evalúa su participación en cada actividad en función de su perfil y de un clasificador compuesto que predice el éxito del alumno en las notas finales. El clasificador compuesto está formado por tres clasificadores independientes (los  $k$  vecinos más próximos, métodos bayesianos y máquinas de vectores soporte) cuya salida representa una decisión más fiable y precisa que los clasificadores simples.

### **Sistema multiagentes**

En [30] se presenta un sistema ITS capaz de adaptarse al ritmo de aprendizaje, ofreciendo las áreas en las que el alumno destaca y la ratio de contenidos a medida que se adquiere mayor conocimiento. Los agentes se modelan para realizar tareas pedagógicas y distribuir eficientemente las lecciones a los alumnos. Se introduce un entorno de trabajo, IMAPS, que captura la creación dinámica de conocimiento en el dominio de interés y el contenido de conocimiento aprendido por el alumno a lo largo del tiempo (tiempo empleado por lección, tiempo por sesión, número de revisiones, frecuencia de peticiones de ayuda, tiempo medio por página). Se han empleado técnicas de agrupamiento para agrupar características y métodos estadísticos para detectar procesos de aprendizaje irregulares.

Siguiendo con el mismo enfoque, en [31] se presenta el sistema multiagente MARS que recomienda materiales educativos online (*learning objects*) a los usuarios en función de sus intereses de aprendizaje. La arquitectura del sistema está compuesta por un sistema multiagente, reglas de recomendación, un mapeado ontológico, funciones emergentes y analizadores de metadatos, implementados con herramientas basadas en Java. También discute la sincronización cooperativa de los agentes pedagógicos en el sistema.

En [32] se presenta el sistema ALLEGRO formado por un ITS y un ambiente colaborativo de aprendizaje (CSCL). La arquitectura multiagente dispone de dos tipos de agentes: humanos y de software que funcionan gracias a la técnica de razonamiento basado en casos, de modo que se emplea la experiencia almacenada de la solución exitosa de problemas similares pasados. El sistema está formado por un agente tutor que guía el proceso de aprendizaje decidiendo las acciones pedagógicas a realizar, el modelo del aprendiz que contempla el estilo de aprendizaje, la comprensión de los temas, las limitaciones y el nivel de conocimientos, el agente colaborativo que busca a otros aprendices que estén tratando el mismo tema para establecer comunicación sincrónica o asincrónica y ofrecer colaboración al aprendiz y por último, el interfaz que

es el puente entre los agentes humanos y los de software. Se fundamenta el paradigma instruccional en tres modelos pedagógicos: conductismo, cognitvismo e histórico-social (cognición distribuida y aprendizaje basado en problemas).

### **Patrones secuenciales**

En [33] se propone un algoritmo de minería de patrones secuenciales extendido combinado con técnicas de descubrimiento de reglas de asociación (las reglas IGB son correlaciones entre premisas mínimas y conclusiones máximas) para extraer patrones temporales y relaciones entre agentes humanos ejecutando tareas procedimentales. Se ha aplicado en CanadarmTutor cuya actividad es el movimiento de un brazo robot de 7gdl de una configuración a otra, siendo los parámetros a definir la selección de la cámara, el aumento/disminución del plano/inclinación/zoom de la cámara y la selección de un valor de rotación de la articulación del brazo. El agente guía al estudiante determinando las posibles acciones que puede tomar, para ello compara las acciones hasta el momento actual con las de los patrones frecuentes y le asigna un nivel de conocimiento (novel, intermedio, avanzado).

En [34] se aplica el algoritmo de patrones secuenciales generalizados (GSP) para identificar secuencias significativas de actividad (indicativas de problemas o patrones que sean marcas de éxito) en los datos de interacción de un grupo de alumnos (por sesiones y un análisis de la distribución de las sesiones) con el fin de poder actuar a tiempo o detectar una mejora durante el proceso de aprendizaje. Los estudiantes colaboran a través del sistema TRAC y se comunican escribiendo páginas Web llamadas Wiki. TRAC dispone de un sistema de seguimiento que recoge las acciones o eventos de los alumnos: la creación de una página Web (el texto escrito), la generación de un comentario o la modificación de un archivo (las líneas añadidas o borradas).

En [35] se describe un algoritmo de minería de patrones secuenciales para el descubrimiento de información sobre uso del curso por los alumnos (la ruta más empleada) que el tutor puede usar para mejorar y personalizar el curso hipermedia adaptativo (recomendando enlaces automáticamente a los nuevos alumnos).

### **Reglas de asociación**

En [36] se propone utilizar agentes recomendadores para sugerir actividades de aprendizaje en línea en un curso Web basándose en los historiales de acceso. Concretamente utiliza minería de reglas de asociación para entrenar al agente recomendador y construir un modelo que representa el comportamiento de acceso.

En [37] se analizan los ficheros log<sup>1</sup> de entornos de aprendizaje Web utilizando técnicas de minería de reglas de asociación y filtrado colaborativo, para descubrir patrones de navegación útiles y proponer un modelo de navegación, consistente en las relaciones de asociación y las relaciones de secuencia entre documentos.

La utilización de métodos heurísticos como regresión lineal en combinación con reglas de asociación es propuesta en [38] para buscar modelos de transferencia de aprendizaje de estudiantes a partir de los ficheros log de interacción del ITS. El modelo de transferencia relaciona preguntas con los componentes del conocimiento necesarios para responderlas correctamente. Las reglas de asociación guían el proceso de búsqueda para encontrar modelos de transferencia que predican el éxito de los estudiantes.

---

<sup>1</sup> Un fichero log es un registro de las comunicaciones entre un sistema y los usuarios de dicho sistema.

En [39] se utilizan reglas de asociación borrosas para descubrir relaciones entre patrones de comportamiento de los estudiantes, incluyendo el tiempo de acceso, números de páginas leídas, preguntas contestadas, mensajes leídos y enviados, etc. Mediante monitorización y análisis, el sistema indica, mediante umbrales de alarma, los resultados a los estudiantes y profesores.

En [40] se presenta un sistema educacional llamado WELSA cuya función es el seguimiento del estudiante (monitorizando su interacción con el sistema), identificando sus preferencias de aprendizaje e infiriendo mediante reglas su pertenencia a una dimensión de estilo de aprendizaje particular. Una vez determinado el modelo dinámico del estudiante, el sistema le adapta los recursos educativos y la navegación gracias a la definición con metadatos del material instructivo y a un sistema inteligente basado en reglas.

### **Agrupamiento**

En [41] se propone un entorno de aprendizaje colaborativo basado en un sistema inteligente de hormiguero que estructura las oportunidades para una colaboración efectiva y el aprendizaje de forma dinámica. Los algoritmos de agrupamiento basados en hormigas muestran algoritmos dinámicos y descentralizados con una estructura jerárquica. Las hormigas o agentes pueden ajustar su comportamiento de acuerdo a diferentes entornos de agrupamiento, cambios de métricas de distancia y atributos de objetos, sin afectar a su actuación, dado que su estado se determina por su información local y se transmite al resto de agentes a través de un efecto cooperativo, formando clusters.

Siguiendo con el mismo enfoque, en [42] se presenta un sistema de minería de patrones secuenciales y agrupamiento mediante el algoritmo de k medias implementado en WEKA con medida de la distancia euclídea, para ver aspectos relevantes o factores de éxito en la colaboración de grupos que pueden ser asociados con salidas positivas o buenas prácticas o bien negativas y que son indicativas de problemas.

El objetivo de [43] es agrupar mediante el algoritmo de k medias y de c medias borroso, los datos de resultados de exámenes disponibles de un ITS y visualizar usando un sistema de coordenadas paralelo, el resultado del análisis de datos multidimensional. El agrupamiento y la visualización de las puntuaciones en los niveles de conceptos dan información significativa del trabajo realizado en el curso y de los conceptos que presentan mayor dificultad para los alumnos.

## **3.4 Herramientas informáticas para la minería de datos**

Para agilizar el proceso de aplicación de la minería de datos existen varias herramientas informáticas que pueden obtener los datos del LMS, procesarlos y extraer conclusiones sobre patrones de comportamiento, agrupamientos, relaciones, etc. que posteriormente empleará el LMS para actualizar los modelos dominio, del alumno y del tutor y generar así el itinerario de aprendizaje más adecuado al mismo. En la tabla 3.2 se muestra una selección de las aplicaciones más utilizadas y una breve descripción de las mismas.

Herramienta	Distribución	Funciones de Minería de Datos
Clementine	Comercial	Asociación (GRI, Apriori), clasificación (árboles de decisión, redes neuronales), agrupamiento (k medias), regresión lineal y logística y modelos combinados.
Oracle Data Mining	Comercial	Clasificación (árboles de decisión, redes neuronales, vecinos más próximos, aprendizaje bayesiano), regresión (redes neuronales), y agrupamiento (k medias y o agrupamiento)
SAS Enterprise Miner	Comercial	Clasificación (árboles de decisión, redes neuronales), regresión lineal y logística y modelos combinados
DB2 Intelligent Miner	Comercial	Agrupamiento, asociaciones, patrones, Clasificación, predicción y análisis de series temporales
Kepler	Comercial	Clasificación (árboles de decisión, redes neuronales, vecinos más próximos), regresión no lineal y estadística.
Weka	Código abierto	Asociación (Apriori), clasificación (árboles de decisión, vecinos más próximos, SVM), agrupamiento (K medias, EM, Cobweb) y modelos combinados.
Keel	Código abierto	Regresión, clasificación, agrupamiento y minería de patrones
RapidMiner	Código abierto	Regresión, clasificación y agrupamiento
Xelopes	Código abierto	Clasificación (árboles de decisión, SVM, redes neuronales), agrupamiento (k medias, jerárquico) y reglas de asociación

*Tabla 3.2:* Herramientas informática para la minería de datos

### 3.5 Conclusiones

A lo largo de este capítulo se han descrito las principales técnicas de minería de datos aplicadas a entornos de enseñanza basada en Web y se han presentado las investigaciones y trabajos desarrollados más relevantes.

Como se ha podido observar, en la mayoría de estos trabajos las técnicas de minería de datos realizan las funciones de clasificación y agrupamiento, descubrimiento de reglas de asociación o secuencias de patrones. Al final del capítulo se han citado algunas herramientas informáticas de código abierto o comerciales capaces de realizar estas funciones de minería de datos.

La minería de datos aplicada a sistemas tradicionales de e-learning suele utilizar solamente la información proporcionada por los ficheros log capturados por los servidores Web y las preferencias personales de los estudiantes. Por otro lado, la minería de datos aplicada a ITS o AIWBES, suele disponer de mayor información en los ficheros log como es la interacción entre el estudiante y el sistema sobre las actividades realizadas, los aciertos y fallos, las notas y niveles de conocimiento, etc.

Todos estos factores que determinan las características del alumno en el proceso de enseñanza y aprendizaje se van a estudiar en el siguiente capítulo que describe las teorías pedagógicas para modelar el Modelo del Alumno.



## Capítulo 4

# Modelo del Alumno

La característica principal de los Sistemas Tutores Inteligentes es que todos los alumnos puedan aprender, a pesar de sus diferentes estilos de aprendizaje. Para ello es necesario detectar de qué modo aprenden los alumnos, y en función de ello presentar los materiales de enseñanza (contenidos, tareas, cuestionarios, etc.) acordes al mismo.

Los estilos de aprendizaje designan una combinación de características cognitivas, afectivas y otras psicológicas que sirven como indicadores estables relativos al modo en que el alumno percibe, interactúa y responde al entorno de aprendizaje.

En los sistemas e-learning se han propuesto, principalmente, dos métodos para determinar los estilos de aprendizaje [44]:

- explícito: el alumno rellena un cuestionario psicológico especializado
- implícito: se analiza e interpreta la interacción del alumno con el sistema

Además de los estilos de aprendizaje, el modelo del alumno lo conforma también el nivel de conocimiento, los intereses y los objetivos del estudiante.

El modelo del alumno contiene, por tanto, dos sub-modelos distintos, uno para representar el estado de conocimiento del alumno y otro para representar las características cognitivas y preferencias de aprendizaje del estudiante. Se realiza esta distinción porque el estado de conocimiento del alumno se actualiza constantemente en base a las interacciones del alumno con el ITS y condiciona la adaptabilidad de los contenidos presentados por el ITS. Y por otro lado, las características cognitivas y los estilos de aprendizaje son rasgos más estáticos, que se mantienen durante un periodo de tiempo significativo, y que condicionan la adaptabilidad de la presentación del material multimedia.

El capítulo comienza describiendo el principal modelo pedagógico para determinar los estilos de aprendizaje del alumno y muestra algunos estudios y proyectos que lo han implementado para el modelo del alumno. A continuación se presentan los métodos para determinar el nivel de conocimiento previo del alumno y las investigaciones que los están desarrollando y por último, se cierra el capítulo con las conclusiones.

### 4.1 Estilos de aprendizaje

Un estilo de aprendizaje se define como unas preferencias o resistencias características en el modo en que el alumno adquiere y procesa la información [45], y ello determina la única forma de aprender que tiene cada alumno.

En el modelo de Felder [45], se definen diferentes dimensiones relativas a cómo la gente procesa la información (descritas en la tabla 4.1), de modo que cada dimensión tiene dos posibles valores:

- **Procesamiento:** activo/reflexivo
- **Percepción:** sensorial/intuitiva
- **Entrada:** visual/verbal
- **Comprensión:** secuencial/global

Dimensión	Estilo	Descripción
Procesamiento	Activo	Tiende a hacerlo mejor cuando puede trabajar en ello y en realidad experimenta o manipula cosas manualmente
	Reflexivo	Prefiere pensar las cosas a fondo antes de actuar
Percepción	Sensorial	Se orienta hacia hechos y datos concretos
	Intuitiva	Prefiere lo conceptual y teórico a lo concreto
Entrada	Visual	Prefiere ver lo que está aprendiendo a través de gráficos, diagramas e imágenes
	Verbal	Tiene mayor éxito cuando oye o lee la información con palabras.
Comprensión	Secuencial	Prefiere disponer la información de forma lineal y ordenada
	Global	Prefiere ver la imagen de conjunto primero

**Tabla 4.1:** Comportamiento de los estudiantes de acuerdo con su aprendizaje

En general, para cada dimensión, todo el mundo tiene a veces un comportamiento y otras veces el contrario, pero, frecuentemente, existe una preferencia (fuerte o moderada) para una categoría o la otra. En cualquier caso, cuando una preferencia por una categoría es fuerte, el proceso de aprendizaje puede mejorar en efectividad con una instrucción adaptada a este estilo de aprendizaje.

En el proyecto ADAPTAPlan propuesto en [46] se definen clusters para cada una de las cuatro dimensiones de Felder, con la intención de separar claramente la preferencia de los distintos estudiantes.

Se obtiene el estilo de aprendizaje directamente al recopilar la información del cuestionario sobre el Índice de Estilos de Aprendizaje [47] que pasa el alumno. El objetivo de este cuestionario es establecer el estilo de aprendizaje dominante de cada alumno. El cuestionario está formado por 44 preguntas, de modo que hay 11 cuestiones para cada una de las 4 dimensiones (distribuidas desordenadamente en el formulario). Existen dos posibles respuestas para las cuestiones, cada una definiendo un valor distinto de la dimensión.

La Figura 4.1. muestra un ejemplo del resultado que obtiene el alumno al pasar el formulario, que puede estar situado en un extremo de la escala o bien en la zona intermedia. Para facilitar el procesado de los estilos de aprendizaje, los diferentes valores cuantitativos de la escala (11, 9, 7, 5, 3, 1) se agrupan usando modificadores cualitativos o clusters (fuerte, moderado, equilibrado), como se describe en la tabla 4.2.

ACT	11	9	7	5	3	X	1	1	3	5	7	9	11	REF
							<-- -->							
SEN	11	9	7	5	3	X	1	1	3	5	7	9	11	INT
							<-- -->							
VIS	11	9	7	5	3	1	1	X	3	5	7	9	11	VRB
						<-- -->								
SEQ	11	X	7	5	3	1	1	3	5	7	9	11	GLO	
						<-- -->								

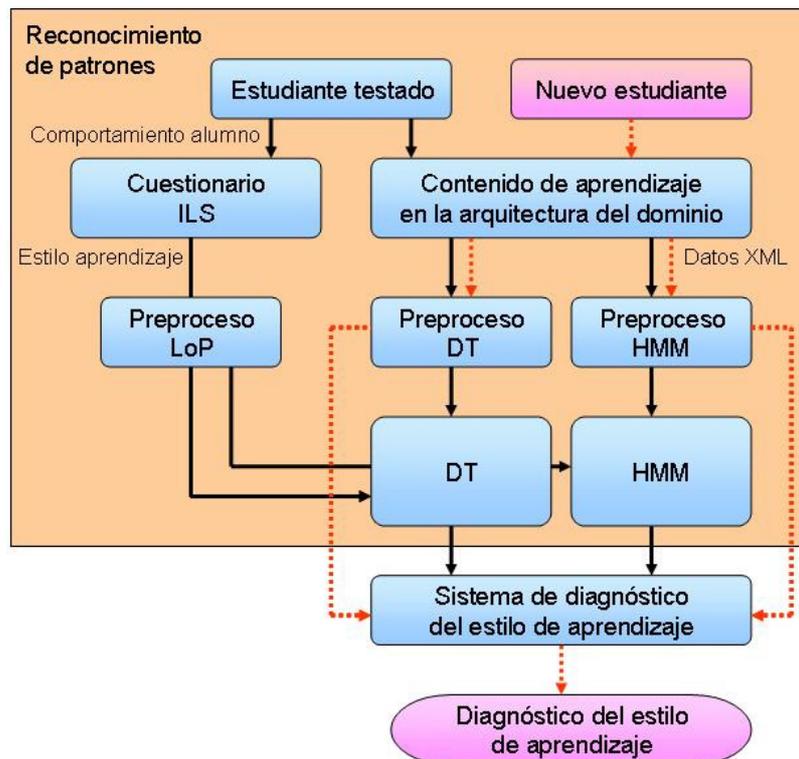
**Figura 4.1:** Resultado para las cuatro dimensiones de los Estilos de Aprendizaje de Felder

Cluster	Valor	Descripción del Estilo
Equilibrado	1s, 3s / -3i, -1i	Sensorial / Intuitivo
Moderado	5s, 7s / -7i, -5i	Sensorial / Intuitivo
Fuerte	9s, 11s / -9i, -11i	Sensorial / Intuitivo

**Tabla 4.2:** Clusters para los Estilos de Aprendizaje de Felder (para la dimensión Percepción)

En [48] han desarrollado un ITS que diagnostica los estilos de aprendizaje de Felder a través de patrones de comportamiento del estudiante al interactuar con el interfaz de usuario, en lugar de extraer el estilo de aprendizaje usando el cuestionario sobre el Índice de Estilos de Aprendizaje. En base al resultado, el sistema personaliza esta interfaz para adaptarla a los estilos y preferencias específicos del individuo. El contenido de aprendizaje de este proyecto se ha implementado con Macromedia Flash.

El enfoque seguido en el estudio se muestra en la Figura 4.2. Para construir el modelo para los estilos de aprendizaje, se han recogido las actuaciones del estudiante con la interfaz (en archivos XML) aplicando posteriormente dos técnicas de minería de datos: se han analizado los datos con Árboles de Decisión (DT), que producen las reglas de clasificación a partir del reconocimiento y clasificación de patrones y a continuación se han modelado estos datos mediante el método estadístico de los Modelos de Markov Ocultos (HMM), para tener en consideración la secuencia con la que se han realizado las acciones del alumno.



**Figura 4.2:** Diagrama de Flujo del Diagnóstico del Estilo de Aprendizaje

En [49] se propone una metodología para determinar el modelo del alumno de un ITS para un entorno distribuido usando el Razonamiento Basado en Casos (CBR) con la ayuda de agentes.

El enfoque se basa en encontrar, a medida que el alumno interactúa con el sistema, un caso que sea similar al dominio de aprendizaje de un antiguo alumno en un entorno distribuido. Cada alumno tiene un único agente personal (perfil de alumno) que incluye el conocimiento previo, el estilo de aprendizaje, intereses, cursos matriculados, etc. Los otros dos agentes del sistema son el agente profesor y el agente curso, que gestionan el material del curso, las técnicas de enseñanza y la estrategia de modelado del alumno.

En el trabajo propuesto en [50] se detecta el estilo de aprendizaje de un estudiante evaluando las redes Bayesianas, cuya misión es modelar los diferentes aspectos del comportamiento de un estudiante mientras éste trabaja en el sistema educativo basado en Web. Para ello se construye un grafo que contiene las variables de interés (tres de las dimensiones de Felder: percepción, procesamiento y comprensión), las relaciones entre las mismas y se asigna una distribución de probabilidad para cada nodo del grafo que indica la importancia de las relaciones modeladas. Se ha descartado la dimensión de entrada en el trabajo, dado que los videos o simulaciones no forman parte de los cursos Web ofrecidos.

A continuación, se infiere el estilo de aprendizaje de acuerdo con el modelo de comportamiento obtenido, a partir de los valores de probabilidad marginal calculados para los nodos. Se ha constatado en este estudio diferentes niveles de precisión en el modelo para las distintas dimensiones de los estilos de aprendizaje de Felder.

## 4.2 Nivel de Conocimiento

El modelo de conocimiento del alumno más generalizado entre los trabajos e-learning, es el que se basa en la taxonomía de Bloom [51], que considera seis niveles de conocimiento (conocimiento, comprensión, aplicación, análisis, síntesis y evaluación), en orden creciente de competencia, como se observa y describe en la tabla 4.3.

Objetivo Bloom	Descripción	Nivel
Conocimiento	Recuerda un hecho sin una comprensión real del significado	Principiante
Comprensión	Capta el significado del material	
Aplicación	Puede usar el material aprendido en situaciones nuevas y específicas	Medio
Análisis	Puede dividir un problema complejo en diferentes partes	
Síntesis	Puede unir diferentes partes con el objetivo de crear nuevas entidades	Experto
Evaluación	Puede sopesar valores de una materia con un propósito específico	

*Tabla 4.3:* Niveles de conocimiento según Bloom

El alumno adquiere estos niveles a través del proceso de aprendizaje, por medio del estudio de los objetos de aprendizaje de las materias del curso y de la actuación con las actividades asociadas.

Mientras que el estilo de aprendizaje es algo inherente al estudiante, el nivel de conocimiento es el conocimiento adquirido por el estudiante con respecto a una competencia o un objetivo didáctico. De modo que el nivel de conocimiento puede ser adquirido dinámicamente a través de las interacciones del usuario con el sistema.

En el enfoque propuesto en [46], el nivel de conocimiento de un estudiante de un objetivo se describe por medio de un atributo que toma uno de los posibles valores: principiante, medio o experto (según la tabla 4.3). Se almacenan estos valores en el Perfil de Estudiante, de acuerdo con la especificación del estándar IMS-LIP2 y se definen los niveles de competencia a través de la herramienta IMS-RDCEO. Es tarea del autor del curso la descripción semántica de las competencias genéricas del curso y de las actividades propuestas, y la realiza por medio de Objetos de Aprendizaje con formato de metadatos, según los estándares IMS-MD/IEEE-LOM. El proyecto define las relaciones entre el modelo del alumno (estilos de aprendizaje de Felder, nivel de competencia basado en la taxonomía de Bloom y nivel de competencia colaborativa) y los recursos meta-datos, con el fin de planificar las rutas de aprendizaje, mediante la herramienta IMS-LD.

Como se acaba de comentar para la definición del modelo del alumno, también se considera el nivel de competencia colaborativa, que define aspectos importantes en la forma de actuar cooperativa y colaborativa del alumno, según se explica en la tabla 4.4, donde se detallan los seis valores de colaboración que a su vez pueden agruparse en tres niveles, según [46], que son bajo, medio y alto.

Valor de Colaboración	Descripción	Nivel
Estudiante no colaborativo	Actúa como si no existieran facilidades para la colaboración	Bajo
Estudiante comunicativo	Comparte información con otros estudiantes usando las herramientas de comunicación disponibles	Medio
Estudiante participativo	Interactúa frecuentemente en el curso	
Estudiante con iniciativa	Empieza las actividades propuestas sin esperar las contribuciones de otros alumnos	
Estudiante perspicaz	Hace contribuciones y comentarios en actividades de otros alumnos con el fin de recibir mayor puntuación	Alto
Estudiante útil	Hace comentarios y contribuciones que son consideradas por otros alumnos	

**Tabla 4.4:** Nivel para la Tabla de Competencia Colaborativa

Siguiendo con un enfoque similar, en [52] se presenta un método de modelado del conocimiento gráfico (MOT), en el que se representan las estructuras o dominios del conocimiento por la unión o la relación entre objetos tipo (conceptos, procedimientos y principios), mediante mapas conceptuales, organigramas, árboles de decisión y otros tipos de modelos. Se incluye también un editor gráfico MOT+LD que exporta los modelos, añadiendo las restricciones necesarias, a formatos estándar como son XML, IMS-LD y OWL.

La taxonomía de conocimientos propuesta en este proyecto, combina y adapta otras taxonomías, como se aprecia en la tabla 4.5, de modo que las habilidades cognitivas se ven bajo tres perspectivas: como un proceso de resolución de un problema genérico (recibir, reproducir, crear...), como un acto de procedimiento de meta-conocimiento (reconocer, integrar, especificar...) o como un objetivo de conocimiento relacionado con

<sup>2</sup> Los estándares son un modo de almacenamiento de la información del alumno (incluyendo progreso y calificaciones), creados con el fin de permitir la transferencia de la misma a través de distintas aplicaciones de software.

una tarea de procesamiento del conocimiento (reconocer, identificar, memorizar, ilustrar, discriminar...).

Niveles de taxonomía de las habilidades cognitivas			Meta-conocimiento Activo (Pitrat)	Problemas genéricos (KADS)	Objetivos cognitivos (Bloom)	Ciclo de habilidades (Romiszowski)
1	2	3				
Recibir	1. Reconocer					Atención
	2. Integrar	2.1 Identificar 2.2 Memorizar			Memorizar	Discriminación y agudeza perceptiva
Reproducir	3. Especificar	3.1 Ilustrar 3.2 Discriminar 3.3 Explicar	Almacenamiento y búsqueda de conocimiento		Entender	Interpretación
	4. Transponer/ traducir					Memoria procedimental Memoria esquemática
	5. Aplicar	5.1 Usar 5.2 Simular	Expresión del uso del conocimiento		Aplicar	
Crear	6. Analizar	6.1 Deducir 6.2 Clasificar 6.3 Predecir 6.4 Diagnosticar	Descubrimiento de conocimiento	Predicción, supervisión, clasificación, diagnóstico	Analizar	Análisis
	7. Reparar			Reparar		Síntesis
	8. Sintetizar	8.1 Inducir 8.2 Diseñar 8.3 Modelar/ construir		Planificación, diseño, modelado	Sintetizar	
Reinvertir	9. Evaluar		Adquisición de conocimiento		Evaluar	Evaluación
	10. Auto-administrar	10.1 Influenciar 10.2 Auto-controlar				Iniciación, continuación, control

**Tabla 4.5:** Taxonomías de las Habilidades Cognitivas

En el estudio propuesto en [52] se presentan unos índices de seguimiento capaces de describir la actitud del estudiante durante el proceso de aprendizaje y un modelo matemático para implementarlo. El seguimiento del alumno vigila cuánto está aprendiendo y qué temas le son más difíciles. Para ello controla el tiempo de estudio, es decir, el tiempo que el alumno emplea cuando se encara con un Objeto de Aprendizaje, el nivel de preparación y el interés por determinados tipos de recursos (contabilizando las veces que se repite una actividad, qué secuencia de recursos con una dificultad dada se han consultado y la nota que obtiene en el control final). Los parámetros están definidos en meta datos estándar para la descripción del usuario como IMS-LIP.

### 4.3 Conclusiones

El modelo del alumno lo conforman dos tipos de informaciones: estática y dinámica, en función de si se ven o no modificadas por la interacción del alumno con el sistema. La información estática la determina el estilo de aprendizaje del alumno y se obtiene a través de patrones de comportamiento, razonamiento basado en casos o redes baye-

sianas. Esta ontología se integra en el estándar IMS LIP con el fin de utilizar la información para los otros modelos que conforman el ITS y en conjunto presentar y adaptar el itinerario de aprendizaje. La información dinámica en cambio, se obtiene de la interacción del usuario y valora el nivel de conocimiento que va adquiriendo. Para ello se emplean los archivos log que registran todas las acciones del usuario en el navegador Web (tiempo medio de respuesta, intentos para resolver una cuestión, páginas consultadas, etc.) así como los resultados obtenidos en los cuestionarios que evalúan cada objeto de aprendizaje. Todos los parámetros obtenidos después de aplicar técnicas de minería se definen en meta datos estándar para la descripción del usuario como IMS-LIP.



## Capítulo 5

# Laboratorios Virtuales y Remotos

En este capítulo se presentan los laboratorios electrónicos (e-laboratorios) que forman parte del entorno del *e-learning*. Se distinguen dos tipos: los laboratorios remotos, que realizan el control a distancia de sistemas físicos reales y los laboratorios virtuales, que proponen simulaciones de sistemas de control.

Los e-laboratorios se emplean en la enseñanza a distancia de los cursos de ingeniería, concretamente se emplean en las áreas de control industrial y de automatización, con el fin de poder realizar experimentos o prácticas. En ellos se plantean situaciones reales a los alumnos, permitiéndoles descubrir el comportamiento de los sistemas, practicar el uso de instrumentos de medida, verificar teorías científicas...

La introducción de los e-laboratorios en la educación a distancia permite el acceso a cualquier hora y desde cualquier lugar a los laboratorios, agilizando también el aprendizaje, dado que se pueden analizar los problemas con mayor profundidad, al disponer de mayor tiempo de dedicación y de unos materiales de consulta más exhaustivos.

Por otro lado, la creación de e-laboratorios supone aumentar el número de laboratorios tradicionales, al compartir instrumentos y equipos sofisticados y caros entre instituciones.

Sin embargo, el diseño e implementación de e-laboratorios presenta el problema de la falta de estándares industriales y protocolos de comunicación comunes. Así como el hándicap de la reutilización, ya que la mayor parte de los e-laboratorios desarrollados hasta el momento se han construido para un curso específico de una institución académica concreta y no permiten su adaptación o ampliación para otro uso.

A continuación se muestran las características de un e-laboratorio y un estado del arte de los laboratorios remotos y virtuales propuestos para el campo de los sistemas de control y de automatización industrial. Por último, se analizan las tendencias que siguen estas investigaciones en el apartado de conclusiones.

### 5.1 Características

Los e-laboratorios, tanto los remotos como los virtuales, siguen una estructura como la que se muestra en la figura 5.1 [53]. Presentan una arquitectura centralizada, en la que la máquina del cliente (alumno) se conecta a través de Internet con la máquina del servidor (laboratorio).

En el caso de los laboratorios remotos, el servidor está a su vez conectado al sistema de control (i.e. tarjetas de adquisición de datos) o a través de redes: buses de campo industriales (i.e. Profibus, DeviceNet, Inter-bus, Ethernet) o por cable serie RS232, en los procesos de control más sencillos.

En la máquina del cliente se emplea un navegador Web como interfaz gráfico de usuario, para monitorizar los parámetros de control y mostrar los resultados de los experimentos.

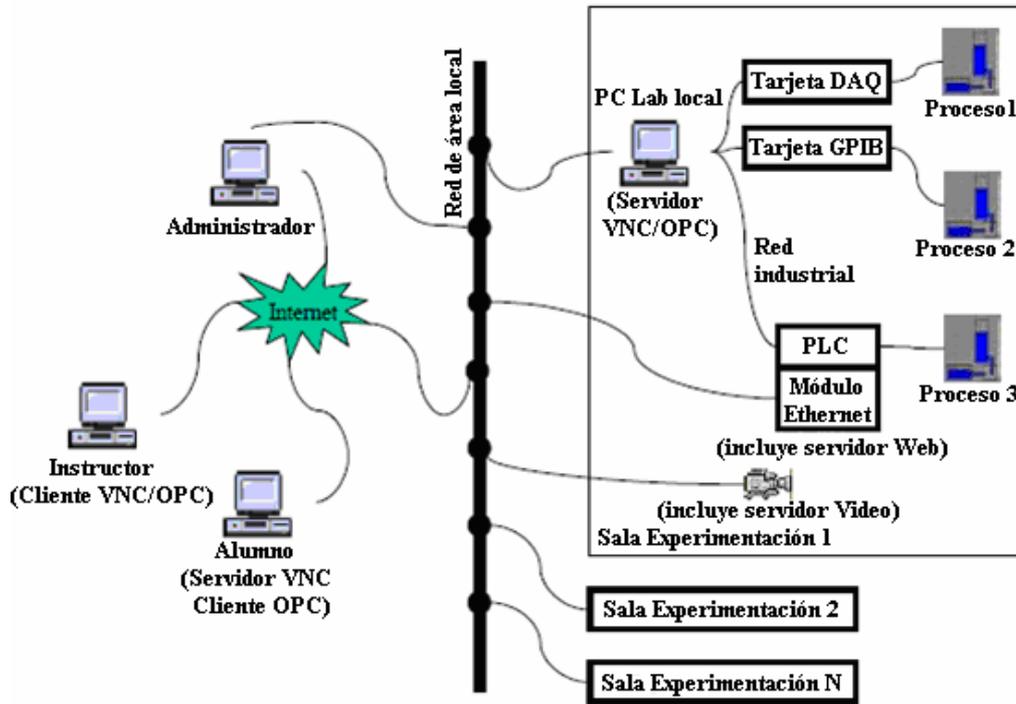


Figura 5.1: Estructura de un laboratorio electrónico

Este interfaz gráfico debe estar integrado en el Sistema de Gestión de Aprendizaje (LMS) que es el encargado de realizar las tareas de seguimiento del proceso de enseñanza-aprendizaje del alumno, como se aprecia en la figura 5.2. En el lado del servidor debe estar ejecutándose un software de control, análisis y medida, que puede ser un sistema de desarrollo integrado comercial (i.e. LabView o Matlab con Simulink), o bien un enfoque basado en la Programación Orientada a Objetos (i.e. C++ o .NET) o una combinación de ambos. En el caso del laboratorio virtual se sustituye el proceso físico a controlar por un motor de simulación que modela el sistema.

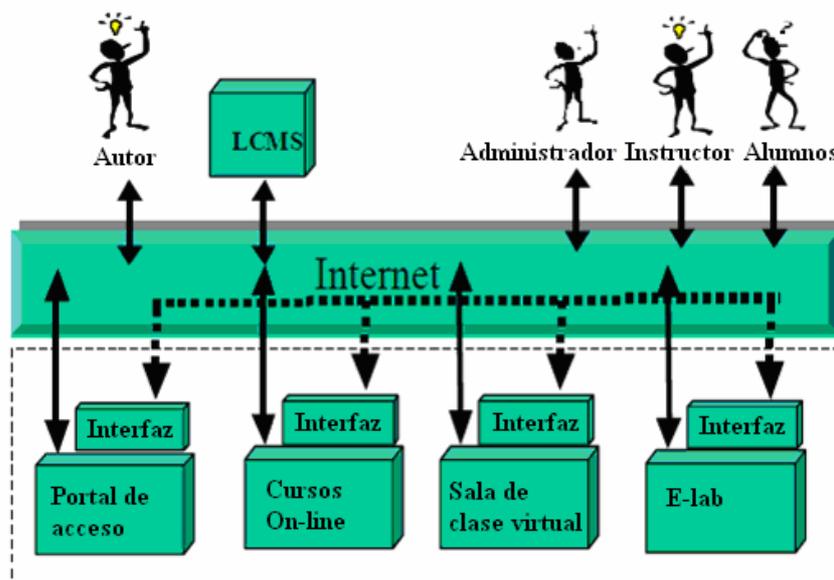


Figura 5.2: Integración del e-Laboratorio en el entorno de aprendizaje global

## 5.2 Laboratorios Remotos

Existen diferentes estudios que plantean laboratorios remotos en diferentes ramas del conocimiento, en este apartado se van a exponer aquellos que se centran en el campo de la ingeniería de control. A continuación se presenta un estado del arte de los mismos, agrupándolos en función del sistema controlado y analizando la arquitectura de conexión que se ha implementado. La mayoría de estudios desarrollan únicamente la parte técnica del laboratorio (protocolos de comunicación, el software del servidor y del cliente), sin incluir un entorno de tutorización inteligente adaptativo que guíe, adapte y verifique la adquisición de conocimientos por parte del alumno.

En [54] se propone un entorno para el manejo de un sistema de control real a través de Internet. El sistema es un péndulo invertido de la marca Quanser. El software empleado para la elaboración del laboratorio de control virtual es Matlab con Simulink.

La experiencia [55] propone el acceso remoto a los siguientes procesos dinámicos: un motor de corriente continua, un tanque con un nivel de control, un sistema de levitación magnética y un simulador de helicóptero con dos grados de libertad. El acceso se realiza a través de un navegador Web. Es necesario instalar el software Matlab con Simulink en la máquina del cliente, para que éste pueda ejecutar sus propios algoritmos de control.

En [56] el sistema permite a los alumnos el acceso a equipos de medida a través de Internet. La arquitectura del sistema está formada por una máquina servidor conectada a los diferentes equipos (sensores, actuadores, motores, fuentes de alimentación) y el equipo de control se realiza a través de la aplicación LabView. En este sistema, se limita el acceso al alumno a algunos parámetros por motivos de seguridad.

Siguiendo con el mismo enfoque, en [57] se aplica el acceso remoto a un laboratorio de teoría de fluidos formado por un sistema con varias cañerías. Los estudiantes pueden monitorizar y controlar el sistema, creado con LabView. Permite explorar conceptos de ingeniería, la resolución de problemas y el análisis de datos y errores experimentales. Además, un interfaz para Matlab permite a los estudiantes validar sus propios códigos.

Un laboratorio remoto para sistemas lineales basado en iLab (herramienta que proporciona un marco de trabajo para el desarrollo de laboratorios remotos) se propone en [58]. El acceso al laboratorio se da a través de software desarrollado con Java. Los estudiantes envían parámetros al servidor que conecta con el analizador de señales digitales, ejecuta las tareas necesarias y devuelve las respuestas solicitadas. Los estudiantes pueden entonces visualizar los diagramas de Bode, Nyquist y Nichols relativos al experimento realizado.

Dentro del departamento de Ingeniería Eléctrica de la Universidad de Dublín, el sistema [59] implementa cuatro tipos de lazo de control (control manual, PID, algoritmo de espacio de estados y lógica borrosa). Para este proyecto, se ha usado el software LabView, Java, HTML, Java Script y VB script. El sistema dispensa imágenes y sonidos del laboratorio a través de una cámara fijada en una plataforma móvil. El software empleado para el envío de esta información es Microsoft NetMeeting, siendo el sistema a controlar unos tanques acoplados.

En [60] se describe el desarrollo y la evaluación experimental, mediante una gráfica de puntuación, de una plataforma de e-laboratorios en el campo de la robótica. La arquitectura del sistema dispone de una interfaz gráfica de usuario basada en Java que incorpora un panel con video en tiempo real, basado en RTP e implementado usando JMF. El panel de control interactivo es una emulación del manual del robot, un

manipulador de tipo SCARA con cuatro grados de libertad, que permite al alumno programar el robot usando su lenguaje de programación V+, para realizar operaciones de ensamblado y tareas de paletizado. La interfaz se ejecuta como un applet en cualquier navegador estándar Web, permitiendo a los usuarios conectarse vía Internet. El servidor del robot comunica con el controlador del robot por cable serie RS232.

La plataforma Tipy propuesta en [61] muestra un sistema de automatización, consistente en una réplica a pequeña escala de un almacén automático vertical Schneider. Dispone de un PLC conectado por el bus Ethernet al motor y sensores así como a una interfaz hombre-máquina. La interfaz Web consiste en páginas PHP que funcionan por el servidor Web APACHE hipervinculadas por un servidor de base de datos MySQL. La representación 3D del sistema se realiza por un applet JAVA. El alumno se conecta como invitado (sin tener que descargar el software) a los dispositivos remotos mediante un software comercial (Symantec PCAnywhere, Microsoft NetMeeting, ExpertCity GotoMyPC, NetSupport, etc.)

En [62] se propone un entorno de experimentación remota con un sistema de ayuda inteligente que apoya a los estudiantes en el uso del laboratorio. La arquitectura consiste en un servidor (administrador del laboratorio) conectado por Internet a un número de estaciones de trabajo, en las que se ejecuta el software de la aplicación con LabView y las interfaces de los circuitos, con Macromedia Flash. Los equipos de instrumentación se controlan por un bus de campo de propósito general, y las placas experimentales, por RS232 o por cable paralelo. Para proporcionar una realimentación visual, se emplean Webcams que se integran al sistema mediante el servidor Macromedia Flash Communication. El usuario accede como cliente al laboratorio, a través de Microsoft.NET Technology. El sistema de gestión de los usuarios (permisos, reservas, etc.) se almacena en una base de datos MySQL. La herramienta de ayuda inteligente está basada en la metodología de razonamiento basado en casos (CBR) y utiliza el paquete Caspian.

### 5.3 Laboratorios Virtuales

La sustitución de sistemas de control reales por simulaciones informáticas en procesos didácticos y de investigación es una práctica ampliamente consolidada, sin embargo, su extensión a los sistemas de educación a distancia ha empezado a ser objeto de estudio de la última década. A continuación se realiza un estado del arte de los laboratorios virtuales del campo de los sistemas de control y automatización.

Una herramienta que simula un laboratorio de procesamiento de señales digitales se describe en [63]. El laboratorio tiene ejercicios que cubren la teoría de la transformada en Z, respuesta en frecuencia, transformada de Fourier y otros métodos de control. La herramienta tiene páginas Web que soportan programas en Java para el desarrollo de los sistemas de simulación de los alumnos.

Siguiendo con la enseñanza de bases teóricas de sistemas de control, en [64] se desarrolla una herramienta basada en el lenguaje de programación Ch (intérprete C/C++) que soporta la notación de los modelos en espacio de estados o en funciones de transferencia. El estudiante, a través de un navegador Web elige el proyecto y los métodos de análisis del controlador, especificando los parámetros de control. Los datos son enviados al servidor para el cálculo numérico y los resultados son devueltos al cliente con el uso de una CGI (*Common Gateway Interface*).

En [65] se proponen experimentos en línea que permiten comparar el comportamiento de diferentes controladores PD, PID y RST en situaciones en las que cambia el punto de funcionamiento del sistema y se ven afectados por perturbaciones.

En el campo de la robótica móvil, se presenta un laboratorio con un asistente virtual inteligente que guía en el uso del laboratorio y ofrece ayuda y propuestas de actividades personalizadas en función de los resultados previos obtenidos [66]. Una interfaz de usuario (un sistema invitado de acceso) visualiza los experimentos del laboratorio (cinemática, sensorización, planificación de trayectorias, algoritmos de control y programación en Java). La representación del modelo de estudiante está basada en modelos relacionales probabilísticos (redes Bayesianas). Los resultados se emplean en el módulo tutor que decide la mejor acción pedagógica a partir de un conjunto de reglas.

## 5.4 Conclusiones

El motivo por el que han crecido estos e-laboratorios es la necesidad en aumento que tienen los alumnos de ingeniería de practicar las lecciones para una asimilación mejor de la teoría. Se ha apreciado que la demanda de uso de laboratorios y equipos de instrumentación supone un problema, de ahí que se priorice el uso simuladores y laboratorios remotos, generando un mayor comodidad en los estudiantes y una difusión más eficaz de la información.

Los laboratorios virtuales y remotos son complementarios, es decir, realizar una simulación previa como entrenamiento en un laboratorio virtual ayuda a evitar problemas de seguridad en el uso de los dispositivos físicos de los laboratorios remotos. Por otro lado, las simulaciones tienen la ventaja de permitir a los alumnos la observación de fenómenos físicos con un comportamiento difícilmente reproducible en un sistema a pequeña escala o bien, modelar sistemas de gran complejidad. Pese a que las simulaciones son excelentes para la asimilación de los conceptos teóricos, éstas no pueden sustituir la experimentación real (o remota) en un laboratorio.



## Capítulo 6

# Entorno de trabajo

Una plataforma o sistema de gestión para la enseñanza (LMS) regula el acceso a los contenidos del curso, administra los estudiantes, sus puntuaciones y la verificación de los resultados y los objetos alcanzados y dispone de herramientas de comunicación para fomentar el trabajo colaborativo o en grupo de los estudiantes. El contenido de los cursos puede ser texto simple, documentos, archivos multimedia o contenido interactivo como cuestionarios, applets, servlets o portlets. El estudiante accede al LMS a través de un navegador Web que aloja el contenido del curso.

El capítulo se divide en dos apartados, en el primero se describen las características que ofrecen las plataformas de código abierto más utilizadas hoy en día para implementar un ITS y a continuación se informa sobre la situación actual de la Web, definiendo en qué consiste la Web 2.0 y los entornos colaborativos así como la última tendencia en entornos educativos, denominados Sistemas Educativos basados en Web semántica (SWBES).

### 6.1 Plataformas de código abierto

Las plataformas de libre distribución permiten la reducción de costes de mantenimiento y adquisición que suponen las plataformas comerciales, y ofrecen además la posibilidad de personalización para un uso en concreto. Se distribuyen dentro de los proyectos Open Source o de código abierto, que se pueden copiar, distribuir y modificar bajo los términos de la licencia GNU General Public License (GPL).

A continuación se describen las plataformas o LMS que ofrecen mayores prestaciones y que tienen una mayor difusión en el mundo de la enseñanza universitaria.

#### **Moodle**

Moodle es un sistema de gestión de cursos, de distribución libre, que ayuda a los educadores a crear comunidades de aprendizaje en línea. Su diseño está basado en las ideas del constructivismo en pedagogía que afirman que el conocimiento se construye en la mente del estudiante en lugar de ser transmitido sin cambios a partir de libros o enseñanzas y en el aprendizaje colaborativo. La base de usuarios registrados incluye más 29 millones, distribuidos en 51.000 Web en todo el mundo y está traducido a 75 idiomas.

#### **Herramientas disponibles**

Está organizado en los siguientes módulos:

- Cuestionarios: base de datos con preguntas que pueden crearse en HTML con imágenes o importarse de archivos de texto, posibilidad de límite de tiempo y de varios intentos.
- Tareas: el tutor especifica el tiempo disponible y la calificación y los alumnos pueden subir archivos de cualquier tipo al servidor. Proporciona diferentes rutas

de aprendizaje pero no existe un modelo de alumno para poder implementar la adaptabilidad.

- Interacción: Foros (de profesores, noticias o abiertos), Consultas (recoge las votaciones de los alumnos), Diario (información privada entre el estudiante y el profesor motivada por preguntas abiertas).
- Recursos: contenido digital, Word, PowerPoint, Flash, video, sonidos y enlaces con aplicaciones Web para transferir datos.
- Encuestas: existen encuestas ya preparadas (COLLES, ATTLS), generando informes que pueden incluir gráficos y exportables con formato de hoja de cálculo Excel o archivo de texto CSV.
- Autenticación vía LDAP, Shibboleth o varios métodos estándar (IMAP)
- Permite la exportación de preguntas o cursos a Blackboard.
- Dispone de herramientas (MySQL Admin o phpMyAdmin) para administrar las bases de datos MySQL.

### **Estándares**

El código de Moodle está escrito en PHP. Permite la importación y exportación de concursos y preguntas en diversos formatos como GIFT (el formato propio de Moodle), IMS QTI, XML y XHTML (NB, aunque consigue exportar con buenos resultados, la importación no está completada actualmente) y de recursos usando IMS Content Packaging, SCORM, AICC (CBT), LAMS.

### **Dokeos**

Dokeos es un entorno de e-learning, una aplicación de administración de contenidos de cursos y también una herramienta de colaboración. Es software libre y está bajo la licencia GNU GPL, el desarrollo es internacional y colaborativo. También está certificado por la OSI y puede ser usado como un sistema de gestión de contenido (CMS) para educación y educadores. Está traducido a 37 idiomas, tiene una comunidad de usuarios que alcanza un millón y se emplea por más de tres mil organizaciones.

### **Herramientas disponibles**

- Itinerario de aprendizaje en SCORM: permite añadir contenido, cuestionarios y actividades, generando automáticamente una secuencia SCORM, una navegación y un informe.
- Producción de documentos basados en plantillas (diagramas, video, flash...).
- Cuestionarios: opción múltiple, llenado de espacios en blanco, cotejar alternativas, preguntas abiertas, hotspots.
- Interacción: agenda, foros, chats, videoconferencia, respuesta abierta a cuestiones y tareas de realimentación.
- Video conferencia: vía Web.
- Conversión de presentaciones en PowerPoint e Impress a cursos en SCORM.
- Informes con calificaciones, progresión, tiempo por módulo, etc. Exportables a formato Excel.
- Encuestas.
- Autenticación vía LDAP y OpenID.
- Gestión de usuarios, cursos y sesiones.

## Estándares

El código de Dokeos está escrito en PHP, usando MySQL como base de datos. También soporta la importación de archivos en SCORM. Los datos de los usuarios pueden ser importados al sistema usando archivos en formato CSV o XML. Dokeos puede añadir información de usuarios y validar sus datos de usuario y clave usando LDAP. Para la versión 1.6, el equipo de desarrolladores de Dokeos ha puesto énfasis en cumplir los estándares de la W3C en cuanto a XHTML y CSS. Sin embargo, el uso de JavaScript es requerido, y la utilización de SCORM requiere el uso de marcos (frames) en el módulo de itinerario de aprendizaje (lecciones a partir de la versión 1.8.7).

## ATutor

Es un Sistema de Gestión de Contenidos de Aprendizaje de Código abierto basado en la Web y diseñado con el objetivo de lograr accesibilidad y adaptabilidad. Los educadores pueden rápidamente ensamblar, empaquetar y redistribuir contenido educativo, y llevar a cabo sus clases online. Contiene herramientas para la gerencia de administración de alumnos, tutores, cursos y evaluaciones en línea, una herramienta de autoría y una de colaboración. Está traducido a 32 idiomas y cuenta con más de mil instalaciones del sistema en todo el mundo.

## Herramientas disponibles

- Accesibilidad conforme a los estándares 1.6.2 IMS/ISO AccessForAll permite a los alumnos configurar el entorno, el contenido, los controles de navegación (global, jerárquica o secuencial) y las herramientas de aprendizaje a sus necesidades. Comprobador de accesibilidad integrado en el editor de contenido para verificar el acceso a personas discapacitadas que usan tecnología asistida.
- Interacción: Inbox de correo electrónico, foros, chats, blogs, RSS, avisos, votaciones, trabajo en grupos y almacenamiento de documentos individual y de grupos.
- Creación de contenido en HTML o en texto, con la inserción de imágenes y notación matemática en formato Latex, desde el editor visual en JavaScript del sistema o importado de un editor local.
- Descarga de paquetes de contenido reproducibles con visores off-line.
- Encuestas (elección múltiple, verdadero/falso, ordenación, relación, arrastre, cuestiones abiertas y bases de datos con preguntas) que pueden asociarse a contenidos para su rápido acceso. Paquetes IMS QTI importables y exportables.
- Glosarios, bases de datos de enlaces Web y motor de búsqueda de contenidos en los cursos.
- Seguimiento del alumno del contenido de las páginas visitadas, que permite la extracción de gráficas que identifican posibles vacíos de contenido y las tendencias de aprendizaje de cada alumno.
- Realimentación de las acciones del estudiante mediante mensajes de éxito, avisos o errores.
- Copia de seguridad del curso almacenada en el servidor de ATutor o en una máquina local.
- Creación de múltiples cuentas de administradores con la especificación de privilegios.

### **Estándares**

ATutor es un programa diseñado en PHP, Apache, MySQL, que trabaja sobre plataformas Windows, Linux, Unix y Solaris. Incorpora las especificaciones de empaquetado de contenido IMS/SCORM, permitiendo a los diseñadores que creen contenido reutilizable. El contenido creado en otros sistemas conforme a IMS o SCORM se puede importar en ATutor, y viceversa. La conformidad con las especificaciones de accesibilidad de W3C WCAG 1.0 en el nivel AA+, permite el acceso a todos los usuarios usando tecnologías asistidas.

### **Ilias**

Es un sistema de gestión para la enseñanza, desarrollado en código abierto bajo la licencia GPL GNU y puede ser utilizado sin ninguna restricción. Debido a esta característica, ILIAS puede ser fácilmente adaptado a los requerimientos específicos de cada organización. Está traducido a 22 idiomas.

### **Herramientas disponibles**

- Entorno personal individualizado
- Control de acceso a los recursos basado en roles
- Contenido de aprendizaje en formato estándar XML, SCORM, AICC
- Gestión del progreso de aprendizaje: creación de itinerarios de aprendizaje basados en la verificación del conocimiento del alumno. Monitorización del progreso en función del tiempo empleado en el aprendizaje de los materiales, resultado de controles, artículos publicados en foros, etc.
- Cuestionarios, valoraciones y encuestas
- Interacción: Chat, foros, correo, RSS y podcasting.
- Autenticación vía LDAP, Shibboleth, CAS, Radius, SOAP

### **Estándares**

Los siguientes sistemas operativos Unix, Linux, Mac OS y Windows pueden usarse como servidores de Ilias con MySQL y PHP. Cumplimiento del estándar LOM, SCORM 1.2, SCORM 2004, IMS-QTI y AICC. Obtención de los certificados SCORM 1.2 RTE nivel 3 y SCORM 2004 3ª edición.

### **Sakai CLE**

El objetivo del Proyecto Sakai es crear un entorno de colaboración y aprendizaje (CLE) para la educación superior, con una arquitectura basada en Java y de código abierto, que pueda competir con sus equivalentes comerciales Blackboard / WebCT y que mejore otras iniciativas de Código Abierto como Moodle. El Software Sakai posee múltiples funcionalidades de comunicación entre profesores y alumnos, RSS, distribución de material docente, realización de exámenes, gestión de trabajos, etc. Para gestionar el Proyecto se ha creado la Fundación Sakai, a la que pertenecen más de 160 instituciones y Universidades.

### Herramientas disponibles

- Plantillas prediseñadas para el entorno.
- Interacción: anuncios, correo, editor wiki, blog, calendario, Chat, foro, glosario, RSS y enlaces con páginas Web. Permite compartir archivos.
- Editor para la creación de contenidos, estructura y permisos. Posibilidad de diseñar lecciones secuenciales.
- Libro de notas: da realimentación al alumno de sus resultados
- Encuestas y concursos
- Informes sobre la actividad relativa a la plataforma exportables, mediante el seguimiento de los usuarios y el uso del sistema.
- Servidor de Base de Datos: MySQL, Oracle e IBM DB2 (próximamente)

### Estándares

Como un sistema basado en Web, Sakai CLE se ajusta a las especificaciones http y HTML mantenidas por el W3C. La configuración y la descripción de los archivos de datos se adhieren a las recomendaciones de W3C relativas a la creación de un formato XML válido. Se han implementado varias guías o especificaciones IMS: Content Packaging, IMS-QTI y IMS-TI. Soporta también SCORM 1.2.

## .LRN

Es un software de construcción de comunidades y e-learning de código abierto, desarrollado en el MIT. Está construido con OpenACS, un conjunto de herramientas para la construcción de aplicaciones Web orientada a comunidades escalable hasta más de 100.000 usuarios. Un gran número de organizaciones y universidades en todo el mundo apoyan el proyecto.

### Herramientas disponibles

- Accesibilidad conforme al nivel AA de la WCAG 1.0 definida por W3C.
- Interacción: noticias, foros, Chat, editor de texto Wiki, mensajes de correo masivos, almacenamiento de documentos individuales y de grupos, calendario de eventos y blogs.
- Creación de contenido de aprendizaje bajo los estándares IMS-CP y SCORM.
- Seguimiento del alumno mediante estadísticas de las visitas a los diferentes módulos del LMS.
- Cuestionarios. Permite su importación y exportación en formato IMS-QTI.
- Informes de calificaciones exportables a formato CSV para su manipulación con herramientas como Excel.
- Tareas: permite a los estudiantes subir tareas al servidor y a los profesores crearlas, categorizarlas, asignar porcentajes de evaluación y crear informes.
- Editor Web integrado en la plataforma para la creación de contenido en formato HTML.
- Base de datos: Oracle y PostgreSQL

### Estándares

Cuenta con soporte a diversos estándares como IMS-MD, IMS-CP, SCORM, IMS-QTI, IMS-LD, IMS Enterprise, obteniendo de esta forma un fácil camino hacia la interoperabilidad de sistemas.

## Drupal

Drupal es un sistema de gestión de contenido modular y muy configurable. Es un programa de código abierto, con licencia GNU/GPL, escrito en PHP, desarrollado y mantenido por una activa comunidad de usuarios. Su diseño es especialmente idóneo para construir y gestionar comunidades en Internet. No obstante, su flexibilidad y adaptabilidad, así como la gran cantidad de módulos adicionales disponibles, hace que sea adecuado para realizar muchos tipos diferentes de sitio Web.

### Herramientas disponibles

- Ayuda on-line y motor de búsqueda del contenido indexado del curso.
- Módulos que proporcionan funcionalidades como 'página de categorías', autenticación mediante Jabber, mensajes privados, bookmarks, grupos de noticias, blog, foros de discusión, wiki, multimedia, galería de fotos, calendario de eventos, encuestas, libro colaborativo, sistema de comentarios enlazados, informes, etc.
- Personalización: el contenido y la presentación pueden ser individualizados de acuerdo con las preferencias definidas por el usuario.
- Gestión de usuarios: se pueden registrar los usuarios e iniciar la sesión de forma local o utilizando un sistema de autenticación externo como Jabber, Blogger o LiveJournal. Para su uso en una intranet, Drupal se puede integrar con un servidor LDAP.
- Permisos basados en roles e historial de login
- Gestión de contenido: El sistema de control de versiones permite seguir y auditar las sucesivas actualizaciones del contenido: qué se ha cambiado, la hora y la fecha, quién lo ha cambiado, etc. así como recuperar una versión anterior.
- Objetos de Contenido: El contenido creado en Drupal es, funcionalmente, un objeto (Nodo) lo que permite un tratamiento uniforme de la información.
- Plantillas (Templates): El sistema de temas de Drupal separa el contenido de la presentación permitiendo controlar o cambiar fácilmente el aspecto del sitio web. Se pueden crear plantillas con HTML y/o con PHP.
- Sindicación del contenido: exporta el contenido en formato RDF/RSS
- Base de datos: MySQL y PostgreSQL, aunque permite incorporar soporte para otras bases de datos.
- Multiplataforma: Puede funcionar con Apache o Microsoft IIS como servidor Web y en sistemas como Linux, BSD, Solaris, Windows y Mac OS X. Por otro lado, al estar implementado en PHP, es totalmente portable.
- Múltiples idiomas: el texto puede ser traducido utilizando una interfaz Web, importando traducciones existentes o integrando otras herramientas de traducción como GNU ettext
- Análisis, Seguimiento y Estadísticas: informes sobre referrals (enlaces entrantes), popularidad del contenido, o de cómo los usuarios navegan por el sitio.

### Estándares

Dispone de un módulo en desarrollo para exportar contenido en el estándar IMS-CP. Otro módulo, actualmente un prototipo, incluye el estándar SCORM 2004, con la implementación SCORM CMI data model para la comunicación entre los objetos de contenido (SCO) y el LMS. Los objetivos futuros de este módulo son incluir informes de los SCO empleados por un usuario en una sesión de aprendizaje y la implementación de la secuenciación y navegación del los SCO.

## 6.2 Situación actual en la Web

### Web 2.0: Entornos colaborativos

El concepto de Web 2.0 surgió en el año 2004 y fue acuñado por Tim O'Reilly, propietario de una empresa editorial de libros de informática. Se refiere a una segunda generación en la historia de la Web basada en comunidades de usuarios y una gama especial de servicios, como las redes sociales, los blogs, los wikis o las folcsonomías, que fomentan la colaboración y el intercambio ágil de información entre los usuarios.

Las ventajas que ofrecen los sistemas basados en Web son [67]:

- Interactividad: hacen posible la comunicación bidireccional y multidireccional y la interacción tanto síncrona como asíncrona.
- Aprendizaje colaborativo: propician el trabajo en grupo y el cultivo de actitudes sociales, permitiendo el aprendizaje entre iguales a través del intercambio de ideas y tareas, desarrollado estos aprendizajes de forma más o menos guiada (cooperativo).
- Libertad de edición y difusión: dado que todos pueden editar sus trabajos y difundir sus ideas. La posibilidad de creación existía en la Web 1.0 pero estaba limitada a los desarrolladores Web y la información quedaba centralizada.

El arquetipo de la Web 2.0 está teniendo un fuerte impacto en las metodologías y tecnologías e-learning, reforzando la participación de los estudiantes en la creación y el intercambio de materiales y recursos. Las plataformas o LMS permiten el aprendizaje colaborativo con foros, chats, áreas de almacenamiento de archivos, servicio de noticias, etc. A continuación se describen algunos trabajos en el campo del e-learning que potencian e incluyen estos servicios colaborativos.

En [68] se revisa el impacto del concepto Web 2.0 en el desarrollo y uso de los sistemas e-learning, con aplicaciones que animan una participación activa del usuario en la creación, intercambio y estructuración de la información. Para ello la Universidad de Oxford realizó una encuesta con el fin de conocer el nivel de uso y las áreas de aplicación de los servicios Web 2.0 (marcadores sociales, calendarios, intercambio de imágenes, vídeos, archivos, autoría colaborativa, redes sociales, weblogs, herramientas de comunicación y juegos o espacios sociales) distribuidas por grupos según las edades. Los resultados muestran un elevado uso de Wikipedia (70-80%), la lectura de Weblogs es moderada (50-60%) en cambio sólo los menores de 24 años escriben sus propios Weblogs, obteniéndose una distribución similar para otras herramientas de redes sociales. Pese a la incorporación de los aspectos Web 2.0 a los sistemas e-learning, algunos de ellos sólo se emplean por una minoría de alumnos, lo que puede deberse a que el grupo de edad más predispuesto a estas tecnologías aun no ha alcanzado el nivel universitario o que hace falta una mayor integración de estas aplicaciones en los entornos e-learning.

Siguiendo con esta línea de investigación, en [69] se estudia si los nativos digitales ya han llegado al nivel universitario, mediante un cuestionario sobre el uso de las herramientas Web 2.0 en el ámbito académico. Del estudio se extrae que no se puede hablar de nativos digitales como una nueva generación de estudiantes dado que el uso que le dan a las herramientas colaborativas es puramente lúdico-social. Por tanto, en la actualidad, los estudiantes no exigen una revolución técnico-pedagógica que se adapte a las características y estilos de aprendizaje de esta hipotética generación digital.

En [70] se propone una herramienta e-learning 2.0, diseñada para apoyar a los usuarios en la edición de recursos educativos y la composición de contenidos multimedia a través del trabajo colaborativo.

En [71] se describe una plataforma diseñada para el apoyo de videos colaborativos, llamada DIVER, donde el video es el contenido y se establecen conversaciones colaborativas alrededor del mismo mediante marcas temporales de las secuencias. Existen otras aplicaciones online que incluyen comentarios o anotaciones pop-up a los videos, como son YouTube, Google Vídeos, Flickr, Nico Nico Douga, etc. Sin embargo, estas últimas no disponen de mecanismos para la gestión de las conversaciones. En el trabajo, se caracterizan cinco diseños de patrones de colaboración (CDPs): interpretación colectiva, diseño distribuido, realimentación de la actuación, codificado de datos distribuidos y la motivación basada en vídeos. Se propone una matriz 3D que incluye los patrones observados y puede servir para expandir las funcionalidades de los videos colaborativos. Los aspectos de la matriz son: estilo del discurso (formal/informal), relación con la fuente del vídeo (realizador/observador) y resultado del objetivo (diseño, síntesis, evaluación y análisis).

En [72] se presenta un software de aprendizaje social Web 2.0, llamado eLogbook que actúa como un contenedor genérico de aplicaciones Web. Se basa en un modelo 3As donde Actores (usuarios o software de laboratorio) generan Activos (recursos) dentro de una Actividad específica (tópicos de discusión o tarea). Se han integrado prácticas de laboratorios remotos externos en dicho marco de trabajo, proporcionando servicios colaborativos a comunidades de aprendizaje en línea. El caso de estudio consiste en el análisis de la conducta dinámica en velocidad y posición de un motor de corriente continua de la UNED. Las aplicaciones cliente y servidor se han programado usando Easy Java Simulations (*Ejs*) y LabVIEW, respectivamente. Esta última realiza la adquisición de datos y el control de lazo cerrado y se incluye realimentación visual del equipo distante en la aplicación cliente mediante una cámara IP que enfoca el motor.

En [73] se define una estructura de colaboración de tipo flujo de trabajo, denominada flujo de aprendizaje. Se añade dicha funcionalidad en la plataforma Moodle, mediante la integración del motor de flujo de trabajo jBPM en el LMS. La idea es mejorar la efectividad del aprendizaje colaborativo, estructurando las interacciones de los usuarios y la información que intercambian. Permite asignar diferentes actividades a diferentes participantes, determinar el final de una actividad y secuenciar las actividades.

## Web semántica

La última tendencia combina los sistemas e-learning con los sistemas de inteligencia artificial aplicados a la enseñanza, dando lugar a los Sistemas Educativos basados en Web semántica (SWBES). La Web semántica se basa en la idea de añadir metadatos semánticos y ontológicos a las páginas Web con el fin de ampliar la interoperabilidad. Estas informaciones adicionales describen el contenido, el significado y la relación de los datos y pueden ser evaluadas automáticamente mediante minería Web<sup>3</sup>.

En [74] se presenta un modelo computacional, llamado Mathema, para desarrollar un SWBES centrado en facilitar la tarea del desarrollador y del autor del curso. El modelo tiene tres capas: el marco de trabajo, que contiene la ontología del sistema (que define los modelos pedagógico, del alumno y del dominio), la ontología de inferencia (que

---

<sup>3</sup> Es un caso particular de la minería de datos, sirve para extraer conocimiento a partir de datos de la Web [8]. Se pueden distinguir tres tipos: minería de contenidos Web, minería de estructura Web y minería de utilización Web.

configura los mecanismos de inferencia mediante el razonamiento basado en casos y el razonamiento basado en reglas) y la ontología de interacción (que se encarga de la interacción de los agentes inteligentes con la metodología GAIA). La segunda capa es la aplicación, dónde se definen los requisitos del alumno para personalizar el ITS y por último, la capa autor, que proporciona un interfaz amigable para la creación del curso.

Siguiendo el mismo enfoque, en [75] se define una arquitectura para el desarrollo de un SWBES usando la tecnología de sistemas multiagente integrada con los servicios Web. Los agentes inteligentes están diseñados considerando las características de los estudiantes y los cursos, estiman las habilidades relevantes y el conocimiento de los alumnos, formulan objetivos y resultados de aprendizaje, seleccionan los objetos de aprendizaje apropiados y evalúan la actuación de los alumnos. La propuesta presenta el problema de no ser extensible con la adición de nuevos agentes a la arquitectura.

En [76] se propone el uso de servicios descritos en tecnología Web semántica para la construcción de sistemas e-learning. Se presentan ejemplos de servicios de búsqueda en bancos de datos de cuestiones, que mediante el razonamiento basado en reglas facilitan la declaración de nuevos servicios, propiciando la interoperabilidad de los metadatos.

En [77] se diseña un marco de trabajo orientado a los servicios de semántica Web para el desarrollo de entornos de aprendizaje adaptativos. El enfoque está basado en modelos de aprendizaje abstractos y reutilizables que describen el proceso de aprendizaje semánticamente como una composición de objetivos de aprendizaje, que se consiguen gracias a la descripción semántica de los servicios Web. Para ilustrar el enfoque, se presenta una implementación prototipo basada en el marco de trabajo WSMO.

En [78] se presenta una plataforma de aprendizaje en línea, denominada INES que incluye un LCMS y un ITS. El sistema comprende varias herramientas y tecnologías: un agente conversacional para ayudar a los alumnos, un agente inteligente basado en la tecnología BDI<sup>4</sup> (creencias, deseos e intenciones) que es el centro del sistema, un motor de inferencia basado en JESS (un motor de reglas para la plataforma Java) y una ontología que proporciona el apoyo semántico para los usuarios del sistema, sus actividades y los contenidos de aprendizaje. La metodología empleada en la ontología crea tres subontologías: del usuario (basada en LIP desarrollado por el consorcio IMS), del contenido (con un modelo LOM que es un estándar IEEE) y del seguimiento (almacena los cursos en los que está inscrito el alumno y la actividad que realiza: objetos de aprendizaje consultados, tiempo empleado, número de accesos y fecha del último acceso)

### 6.3 Conclusiones

La mayoría de las plataformas e-learning populares aun no se han aprovechado de las ventajas de la adaptabilidad (ver tabla 6.1), posiblemente porque el beneficio esperado no justifica aun el alto esfuerzo que supone la implementación y la creación de herramientas de autor para los cursos adaptativos. Sin embargo, si que introducen herramientas colaborativas que permiten el uso de aplicaciones Web 2.0 para establecer

---

<sup>4</sup> En la tecnología BDI, el modelo de agentes está basado en la teoría de Bratman, de modo que las intenciones juegan un papel principal en el razonamiento hacia las acciones. De hecho, las intenciones se suponen estables y condicionan futuras deliberaciones, controlan la conducta e influyen en las creencias sobre el futuro.

redes sociales entre los alumnos y permitir la comunicación, cooperación y la creación o intercambio de conocimiento entre ellos.

Cabe destacar, que en el campo del e-commerce comienzan a aparecer plataformas que incorporan el indexado semántico, como es el caso de NetBase que acaba de lanzar una plataforma de Contenido Inteligente que se aplicará al grupo editorial Elsevier para dar valor añadido a las búsquedas de contenido en material médico y científico.

La tabla 6.1 presenta un resumen de las herramientas o características que soportan las plataformas de código abierto más representativas que se han descrito en este capítulo. Esta tabla sirve como base para el estudio y selección de la plataforma más adecuada para la implementación del ITS adaptativo (ver capítulo 8.1).

		Moodle	Dokeos	ATutor	Ilias	Sakai	.LRN	Drupal
Aprendi- zaje	Encuestas	X	X	X	X	X	X	X
	Almacenado evaluación	X	X	X	X	X	X	X
	Glosario	X		X		X		X
	Motor de búsqueda	X	X			X		X
Cooperación	Foro	X	X	X	X	X	X	X
	Noticias	X	X	X	X	X	X	X
	Correo	X	X	X	X	X	X	X
	Gestión de archivos	X	X	X	X	X	X	X
	Calendario	X	X			X	X	X
	Votaciones	X	X	X		X		X
	Blog	X	X	X		X	X	X
	Wiki	X	X	X	X	X	X	X
Colabo- ración	Chat	X	X	X	X	X		X
	Videoconferencia		X					X
	Pizarra			X				X
	Compartir escritorio							
Gestión de enseñanza	Trabajo en grupo	X	X	X	X	X	X	X
	Creación de contenido	X	X	X	X	X	X	X
	Importación de contenido	X	X	X	X	X	X	X
	Almacenado del Modelo de Alumno							
	Personalización	X	X	X	X	X	X	X
Adaptabi- lidad	Guiado adaptativo			X				
	Personalización del entorno	X				X	X	X
	Arquitectura extensible	X	X			X		X
	Seguimiento del estudiante	X		X		X	X	X
	Apoyo adaptativo							
Estándar	IMS CP	X	X	X	X	X	X	
	IMS QTI	X	X	X	X	X	X	
	SCORM	X	X	X	X	X	X	
	IEEE LOM RDF	X	X		X	X		
BDD	MySQL	X	X	X	X	X		X
	PostgreSQL	X					X	X
	Oracle	X				X	X	

**Tabla 6.1:** Características de las plataformas e-learning

# Capítulo 7

## Extracción de información

Un sistema educativo basado en Web maneja gran cantidad de información relativa a la interacción del alumno con el sistema. Esta información debe ser extraída, preprocesada y almacenada en bases de datos para su posterior análisis mediante la aplicación de técnicas de minería de datos. La inferencia de conocimiento a la información permite su uso para la adaptación dinámica del sistema LMS y de los modelos que conforman el ITS, al ritmo y estilo de aprendizaje del alumno.

En este capítulo se describen cuáles son los archivos o informes de dónde se puede extraer la información y se resumen cuáles son los datos más relevantes para el ITS, de los que se han descrito a lo largo de los capítulos anteriores, con el fin de sintetizar cuál es la información útil para poder modelar el proceso de aprendizaje del alumno. A continuación se describe la estructura de almacenamiento de la información en las bases de datos de los sistemas LMS y se finaliza el capítulo con una conclusión de los puntos tratados en el mismo.

### 7.1 Archivos log

Los sistemas educativos basados en Web pueden registrar los accesos del estudiante en archivos log que proporcionan un seguimiento básico de la navegación del alumno en la página Web.

Hay varios tipos de archivos log [4]:

- **Archivo log del servidor:** constituye la fuente de información más ampliamente empleada para la minería de datos, conteniendo información básica como el tiempo, la ruta y la respuesta a entradas.
- **Archivos log del cliente:** consiste en un conjunto de archivos log, uno por alumno, y contiene información sobre las interacciones del usuario con el sistema.
- **Archivo log Proxy:** consiste en un conjunto de archivos log, que proporcionan una caché entre los navegadores del cliente y los servidores Web. Esta información complementa a la del archivo log del servidor.

#### Archivos log del servidor

Normalmente, hay un único archivo log para todos los estudiantes y existen varios formatos [79]:

- el formato log común NCSA (Microsoft)
- el formato log extendido W3C (Microsoft Internet Information Server 4.0 y 5.0) dispone de una información más detallada.

La información que proporciona cada formato de archivo log se puede consultar en la tabla 7.1.

NCSA Common Log File	W3C Extended Log
<ul style="list-style-type: none"> <li>- dirección IP del cliente</li> <li>- nombre o ID del usuario</li> <li>- fecha y hora</li> <li>- página visitada</li> <li>- código de estado de la consulta http</li> <li>- método de consulta</li> <li>- número de bytes enviados</li> </ul>	<ul style="list-style-type: none"> <li>- dirección IP del cliente</li> <li>- nombre o ID del usuario</li> <li>- fecha y hora</li> <li>- página visitada</li> <li>- código de estado de la consulta http</li> <li>- método de consulta</li> <li>- número de bytes enviados</li> <li>- puerto del servidor</li> <li>- parámetros de la consulta (cookies)</li> <li>- navegador Web del cliente</li> <li>- sistema operativo del cliente</li> </ul>

**Tabla 7.1:** Información de los formatos de archivo log

## Archivos log del cliente

Se pueden implementar por un agente remoto (como Javascripts, Java Applets), modificando el código fuente de un navegador existente, o empleando cookies.

Según lo que establece la Ley Orgánica 15/1999 de 13 de diciembre de Protección de Datos de Carácter Personal (LOPD) se debe informar al usuario que, con la descarga de la aplicación, se va a proceder a un seguimiento de su actividad y al tratamiento de estos datos con fines educativos.

En [80] se desarrolla un esquema denominado metadatos de atención contextualizada (CAM) para capturar las observaciones del usuario, basado en attentionXML, que es una especificación de código abierto para capturar datos de cómo la gente usa la información en los navegadores, la páginas Web, los feeds de noticias y blogs. Estos datos son más tarde analizados y proporcionan información estadística interesante sobre los intereses y actividades de los usuarios a lo largo del tiempo, como se aprecia en la tabla 7.2, además se pueden emplear para predecir tendencias o comportamientos, identificar patrones, etc.

CAM XML Format
<ul style="list-style-type: none"> <li>- dirección IP del cliente</li> <li>- tareas realizadas en un documento (lectura, edición, actualización, escucha, etc.)</li> <li>- fecha y hora de acceso</li> <li>- contexto (curso y tema)</li> <li>- duración del trabajo</li> <li>- ID de la sesión</li> <li>- datos creados por el usuario (ranking, anotaciones, tags)</li> <li>- indicador de preferencias (votación, información gusta-disgusta, etc.)</li> <li>- agrupación de URLs seguidas por el usuario</li> <li>- seguimiento de la relación social entre el autor del documento o en un chat y el lector o participante (elemento xfn)</li> </ul>

**Tabla 7.2:** Información del formato CAM XML

## 7.2 Preprocesado de los datos

Una vez que el servidor registra y almacena la información en un archivo log, se deben analizar los datos para obtener información útil, por ejemplo, en el archivo log se realiza un seguimiento de las páginas visitadas o los archivos consultados pero no de los usuarios.

El preprocesado consiste en el proceso de limpieza y preparación de los datos del archivo log para su posterior análisis mediante minería de datos.

El objetivo es importar los datos a una base de datos relacional, asignando a cada registro una clave o identificador único, limpiar los datos (eliminando los datos innecesarios) y calcular las métricas de interacción estándar que servirán como base para futuros análisis.

A continuación se enumeran las tareas que se deben realizar en esta fase [4]:

- **Identificación del usuario:** asociación de las páginas referenciadas con el usuario conectado
- **Identificación de la sesión:** considera todas las páginas referenciadas con un usuario dado y un curso en una sesión, que se define cuando el intervalo entre dos sucesivas transacciones en el curso supera los 30 minutos.
- **Identificación de transacciones:** se dividen las sesiones en unidades menores, referidas a transacciones o episodios.
- **Transformación de datos:** consiste en el cálculo de nuevos atributos, la conversión de atributos numéricos a nominales, el dar significado a referencias contenidas en el archivo log, etc.
- **Integración de datos:** es la integración y sincronización de la información a partir de fuentes heterogéneas.
- **Reducción de información:** trata de eliminar dimensiones a los datos

## 7.3 Bases de Datos

Para permitir el análisis de la información de los archivos log, estos archivos de texto se importan a bases de datos relacionales compatibles con las del sistema LMS.

A continuación se citan las características de los sistemas gestores de bases de datos (DBMS) relacionales libres de código abierto más populares y potentes, que funcionan sobre múltiples plataformas.

### MySQL

Es un software de bases de datos distribuido bajo licencia GNU GPL. Actualmente existen más de 6 millones de copias de MySQL funcionando, siendo la serie en desarrollo de MySQL Server la 6.0.

Es un sistema cliente/servidor que se compone de un servidor SQL multihilo, varios programas clientes y bibliotecas, herramientas administrativas y una gran variedad de interfaces de programación (APIs).

El tamaño máximo de las bases de datos soportado, viene determinado por el sistema operativo no por los límites internos de MySQL, por ejemplo Windows FAT32 y Linux (2GB/4GB) o Windows NTFS y MacOS (2TB). Se permiten 64 índices por tabla (cada índice puede consistir de 1 a 16 columnas) y 50 millones de registros.

Las características que presenta son integridad referencial, transacciones distribuidas, procedimientos almacenados, disparadores, cursores, vistas actualizables, modo strict,

soporte para SSL (protocolo de capa de conexión segura), query catching, subselects, búsqueda e indexación de campos de texto, MySQL Cluster, planificación de eventos y funciones XML.

## PostgreSQL

Es un sistema de gestión de base de datos relacional orientada a objetos de software libre distribuido bajo licencia BSD<sup>5</sup> clásica. Lo utilizan en la actualidad un centenar de empresas, siendo la serie actual en desarrollo la 8.4.

Soporta bases de datos con un tamaño superior a los 4TB y cientos de millones de registros. El tamaño máximo de un campo es de 1GB y el número de columnas por tabla oscila entre 250 y 1600.

Las características que presenta son transacciones, subselects, disparadores, vistas, claves foráneas, integridad transaccional, tipos de datos definidos por el usuario, herencia de tablas, normas, y control de concurrencia multi-versión (MVCC) lo que permite que mientras un proceso escribe en una tabla, otros accedan a la misma sin necesidad de bloqueos.

## Comparativa y conclusiones

Ambos sistemas gestores de bases de datos ofrecen estabilidad, flexibilidad y buena actuación. La elección de uno de ellos dependerá del uso que se vaya a realizar.

De la tabla 7.3 se desprende que PostgreSQL ofrece más características (lo que va en detrimento de su velocidad) útiles para aplicaciones de bases de datos tradicionales, donde se exijan transacciones y claves de referencia foráneas, mientras que MySQL se centra en aplicaciones basadas en Web, dado que tiene una actuación más rápida y está diseñado para un buen funcionamiento con servidores basados en Web.

En este proyecto se va a diseñar un sistema tutor inteligente que trabaja con datos obtenidos de sistemas educativos basados en Web, por lo que la elección más idónea es la base de datos MySQL que presenta las ventajas de ser más rápida, tener un diseño más simple (lo que implica facilidad de uso) y un sistema de duplicación ampliamente testado. Además, en las últimas versiones, se han incluido características adicionales que antes sólo ofrecía PostgreSQL, como son los procedimientos almacenados, los tipos de datos GIS, las vistas, los disparadores (triggers) o los cursores.

	MySQL	PostgreSQL
Almacenamiento de datos	- InnoDB - BDB	- Sistema de almacenamiento Postgre
Integridad de los datos	- Acuerdo ACID <sup>6</sup> - Bloqueo a nivel de fila	- Acuerdo ACID - MVCC - Bloqueo a nivel de fila
Características avanzadas	- Disparadores (triggers) - Vistas - Procedimientos almacenados - Cursores - División de datos	- Disparadores (triggers) - Vistas - Herencia de tablas - Secuencias - Procedimientos almacenados - Cursores

<sup>5</sup> La licencia BSD, al contrario que la GPL, permite el uso del código fuente en software no libre.

<sup>6</sup> El acuerdo ACID (Atomic, Consistent, Isolated, Durable) es un requisito que asegura la integridad de los datos.

	MySQL	PostgreSQL
Velocidad	- Más rápido	- Más lento
Índices	- Columna única - Multi-columna - Único - Clave primaria - Campos de texto	- Columna única - Multi-columna - Único - Clave primaria - Claves foráneas - Campos de texto
Tipos de datos	- Estándar - GIS <sup>7</sup>	- Estándar - Definidos por el usuario - GIS - Network-aware (reconocen los datos Ipv4 e Ipv6)
Duplicación (replication)	- Maestro único/ multiesclavo	- Maestro único/ multi-esclavo - Multi-maestro/ multi-esclavo
Plataforma	- Windows, Linux, FreeBSD, MacOSX - Modelo multi-hilo	- Windows, Linux, FreeBSD, MacOSX - Modelo non-threaded <sup>8</sup>
Métodos de interfaz	- ODBC <sup>9</sup> y JDBC - Acceso vía C/C++, Java, Perl, Python y PHP	- ODBC y JDBC - Acceso vía C/C++, Java, Perl, Python y PHP
Copia de seguridad	- Comprobación de la consistencia de los datos	- Sistema Write Ahead Logging

**Tabla 7.3:** Tabla comparativa con las características más destacadas de los DBMS

## 7.1 Análisis de los Datos

Con el preprocesado se consigue preparar los datos de los ficheros log, tanto del servidor como del usuario para su posterior análisis. Sin embargo la información que proporcionan estos ficheros log de la interacción del alumno con el sistema no es suficiente para la creación y actualización del modelo de alumno y modelo del dominio del sistema ITS y se debe completar con cuestionarios externos o formularios explícitos, en los que se solicite información directamente al alumno sobre sus opiniones del proceso de aprendizaje o aspectos del sistema que no se pueden extraer por métodos indirectos. Esta información adicional, se deberá incluir en la base de datos y complementará los datos anteriores.

Con el fin de definir los campos necesarios para la base de datos relacional es preciso establecer un listado en el que aparezcan las actividades del curso en las que ha participado el alumno, su nivel de conocimiento sobre las mismas, la forma de interactuar, sus preferencias, etc. Esta tarea se facilita si clasificamos la información en función de los datos que interesan a cada modelo que conforma el ITS:

- información personal o perfil del usuario (modelo del alumno)
- resultados académicos (modelo del alumno)
- datos de interacción del usuario con el sistema LMS (modelo del dominio)

<sup>7</sup> GIS (Geographic Information System) es un sistema que captura, almacena, analiza, gestiona y presenta datos relacionados con una localización geográfica.

<sup>8</sup> En el modelo multi-hilo (threaded) todos los usuarios se conectan a una única base de datos, mientras que en el modelo non-threaded cada nueva conexión consigue un nuevo proceso de la base de datos.

<sup>9</sup> Los estándares ODBC (Open Database Connectivity) y JDBC (Java Database Connetivity) permiten a las aplicaciones el acceso a los datos de una amplia variedad de sistemas de gestión de bases de datos.

A continuación se desarrollan los listados con la información relevante que se quiere obtener para cada uno de los modelos descritos. En este apartado únicamente se describe cuál es la información útil y de qué fuente se va a obtener (archivos log de servidor, de cliente o cuestionarios), dejando la implementación de la base de datos (con la definición de los tipos de datos, número de registros, tamaño de almacenamiento, relaciones, etc.) para el capítulo 8.2.

Los listados creados (tablas 7.5, 7.6 y 7.7) han tenido en consideración y han ampliado los datos que se muestran en la tabla 7.4, en la que se puede observar un resumen de la información obtenida después de aplicar técnicas de minería de datos a los sistemas educativos basados en Web propuestos por diferentes artículos que se han ido comentando a lo largo del trabajo de investigación en los capítulos anteriores.

Información	Ref. Bibliográfica
Tiempo de estudio (acceso a páginas, por sesión o tema)	[39], [23], [30], [27], [21], [52]
Tiempo medio por página	[30], [22]
Tiempo empleado en los grupos de discusión online	[27]
Tiempo empleado en responder una cuestión o tarea	[11], [22], [52]
Frecuencia de registro de las sesiones	[27]
Frecuencia de peticiones de ayuda	[30]
Páginas visitadas (revisiones)	[30], [13], [21]
Número de páginas leídas	[39]
Preguntas contestadas	[39]
Mensajes leídos y enviados	[39]
Participación en chats y foros	[22]
Aciertos y fallos	[23], [19], [22]
Recursos más interesantes (demandados o descargados)	[13], [27], [52]
Artículos colgados o respondidos	[27]
Número de alumnos	[13]
Opciones del menú y banners clicados	[13], [26], [22]
Rutas de navegación más habituales	[13], [52]

**Tabla 7.4:** Información útil de la interacción del alumno con los LMS

### Información personal o perfil del usuario

El perfil del usuario lo conforma el estilo de aprendizaje del alumno (las características cognitivas y preferencias de aprendizaje del estudiante), el nivel de conocimiento previo, historial de los cursos matriculados, la disponibilidad temporal (si existen cargas familiares o laborales), los intereses y los objetivos del estudiante.

Información	Fuente de información
Estilo de aprendizaje de Felder	Cuestionario sobre el Índice de Estilos de Aprendizaje [47]
Nivel de conocimiento previo	Test inicial (basado en la taxonomía de Bloom [51]) Cursos matriculados y notas obtenidas
Disponibilidad temporal	Cuestionario personal Archivo log del servidor (tiempo de estudio) Archivo log del servidor (frecuencia de registro)
Intereses y objetivos	Cuestionario personal

**Tabla 7.5:** Información personal o perfil del usuario

## Resultados académicos

Representa el estado de conocimiento del alumno con respecto a una competencia o un objetivo didáctico. Para ello se pretenden conocer las actividades del curso en las que ha participado: la lectura y/o escritura de documentación (en páginas Web, blogs, wikis,...), descarga de archivos, realización de tests y cuestionarios, la actuación o resultado obtenido en las tareas o talleres, la participación en foros y chats, la comunicación con los compañeros, etc.

Información	Fuente de información
Lectura de documentación	Informe del LMS (tiempo empleado en una tarea) Archivo log del servidor Páginas visitadas Tiempo empleado en visitar una página Web Número de páginas leídas (visitadas) Archivo log del cliente: Lectura de documentos (feed de noticias o blogs)
Realización de tareas	Informe del LMS: Número de tareas realizadas Tiempo empleado en una tarea Archivo log del servidor Páginas visitadas (revisiones de contenidos) Archivo log del cliente: Tareas realizadas en un documento (lectura, edición, actualización, escucha, etc.) Subida de archivos al servidor
Descarga de archivos	Archivo log del servidor (páginas visitadas) Archivo log del cliente (descarga de documentos)
Realización de test	Informe del LMS: Número de tests realizados Tiempo empleado en los tests
Resultados de las actividades	Informe del LMS (nota final obtenida en el curso)
Resultados de los test	Informe del LMS: Número de tests aprobados y suspendidos Número de preguntas contestadas Aciertos y fallos
Participación en foros y chats	Informe del LMS: Número de mensajes publicados en chats y foros Número de mensajes leídos en foros Tiempo empleado en foros y chats
Comunicación con otros alumnos	Archivo log del cliente: Número de mensajes enviados a otros alumnos Número de mensajes leídos de otros alumnos Seguimiento de la relación social del alumno
Comunicación con el profesor solicitando ayuda	Informe del LMS: Número de mensajes enviados al profesor Archivo log del cliente: Frecuencia de petición de ayuda
Secuenciación de las tareas	Archivo log del cliente: Agrupación de URLs seguidas por el usuario

**Tabla 7.6:** Información sobre los resultados académicos

## Datos de interacción del usuario con el sistema

Está íntimamente relacionado con los otros apartados, dado que el modo de interactuar con el sistema es una consecuencia del estilo de aprendizaje del alumno y al mismo tiempo, es el medio mediante el cuál el alumno adquiere los conocimientos del curso al ir realizando las actividades.

Al recoger esta información lo que se pretende es mostrar las preferencias del usuario en el uso del sistema LMS, subrayando qué aplicaciones o herramientas disponibles en el LMS (aprendizaje, cooperación o colaboración) presentan un mayor uso o dedicación por parte del alumno y dentro de las mismas, qué tipo de actividades le han gustado más.

Información	Fuente de información
Preferencias del usuario	Cuestionario de preferencias Archivo log del cliente: Datos creados por el usuario (anotaciones, tags y rankings) Indicador de preferencias (votación, información gusta/disgusta, etc.)
Herramientas más utilizadas	Archivo log del servidor Páginas visitadas (wiki, blog, foro, chat, etc.)
Materiales más demandados	Archivo log del servidor Descarga o lectura de documentos Páginas visitadas (contenidos)
Actividades mejor valoradas	Informe del LMS: Número de tests aprobados y suspendidos Notas obtenidas en las actividades
Rutas de navegación habituales	Archivo log del cliente: Agrupación de URLs seguidas por el usuario
Opciones del menú y banners clickados	Archivo log del cliente: Agrupación de URLs seguidas por el usuario

*Tabla 7.7:* Información de la interacción del alumno con el sistema

## 7.4 Conclusiones

En este capítulo se han enumerado las fuentes que aportan información sobre el proceso de aprendizaje del alumno y su interacción con el sistema LMS en la realización del curso. Las fuentes son tres: los informes sobre los alumnos que proporciona el sistema LMS, los archivos log del servidor y del cliente y por último, los cuestionarios externos, diseñados con el fin de complementar la información que define el modelo del alumno y el modelo del dominio.

Se han descrito las fases de preprocesado, almacenamiento y gestión de los datos en bases de datos relacionales y el análisis de los datos con técnicas de minería de datos (explicadas en los capítulos anteriores). La información útil obtenida se ha clasificado en función del modelo del ITS que la va a utilizar y se ha detallado en tablas resumen que se emplearán en el próximo capítulo para definir los campos y tipos de datos necesarios para la base de datos relacional del LMS.

## Capítulo 8

# Propuesta de metodología para la implementación del ITS adaptativo

Después de haber realizado, en los capítulos previos, el estado del arte de los sistemas tutores inteligentes, narrando las técnicas empleadas para implementarlos y describiendo las características de los elementos o módulos que los componen, se pasa, en este capítulo, a detallar cuáles serán los componentes óptimos que conformarán el sistema FITSLAB.

Para ello se elige en primer lugar la plataforma más idónea para albergar el sistema LMS en base al cumplimiento de una serie de requisitos (características y disponibilidad de herramientas).

A continuación se describen las características de los módulos que constituyen el LMS, que son el modelo del alumno (evalúa el conocimiento y la actuación del alumno en el sistema), el modelo del tutor (selecciona la intervención educativa más adecuada al alumno) y el modelo del dominio (contiene los recursos educativos del curso).

Seguidamente se selecciona la base de datos que almacenará y gestionará los registros de la interacción del alumno con el sistema con el fin de aportar información útil que realimente el sistema de control actualizando los modelos del ITS que se acaban de definir.

El análisis de los datos mediante técnicas de minería de datos, constituye la parte más importante y la aportación de este trabajo de investigación. En este apartado se profundizará y desarrollará la técnica o combinación de técnicas de minería de datos que mejor se adapta y resuelve el problema planteado.

Por último, para realimentar el sistema de control propuesto en la Figura 2.1 es necesario valorar los resultados que obtiene el alumno al finalizar el curso, comparados con los objetivos fijados en el inicio del mismo, con el fin de tomar medidas a largo plazo y adaptar el LMS al proceso de aprendizaje del alumno, lo que se lleva a cabo en el apartado denominado sistema de control.

### 8.1 Selección del LMS

En el proceso de toma de decisiones sobre la plataforma más idónea para gestionar el LMS, se ha recurrido a la teoría de decisión multicriterio discreta (DMD) [81] que constituye una herramienta para poder considerar el conjunto de objetivos o características que debe ofrecer el sistema LMS, buscando un compromiso o equilibrio entre ellos.

#### Plataforma

Con el fin de seleccionar la plataforma óptima para desarrollar el sistema FITSLAB, de las estudiadas en el capítulo 6.1, se han seguido las consignas propuestas en [82] donde se indican una serie de características básicas e imprescindibles que la plataforma LMS debe cumplir.

Se han adecuado y priorizado estos objetivos para adaptarlos a las necesidades de nuestro objeto de estudio, quedando el listado que se muestra a continuación:

1. **Adaptabilidad:** el conjunto de funcionalidades que permiten que el sistema tenga una adaptación automática de la estructura del LMS, conforme a los planes de estudio, los contenidos y los estilos pedagógicos del curso y el proceso de aprendizaje del alumno.
2. **Estandarización:** capacidad de permitir el seguimiento del comportamiento de los estudiantes dentro del curso, utilizar cursos realizados por terceros y garantizar la durabilidad de los cursos evitando que se queden obsoletos.
3. **Gestionabilidad:** Que el sistema pueda obtener y trazar la información adecuada sobre el usuario y el contenido.
4. **Interactividad:** disponer de un conjunto de herramientas que permiten una amplia interacción o participación del usuario con el sistema y con los compañeros del curso.
5. **Tecnología empleada:** en cuanto a la programación, destacan en este orden PHP, Java, Perl y Python como lenguajes Open Source, muy indicados para el desarrollo de webs dinámicas.
6. **Internacionalización:** que la plataforma esté traducida para que los usuarios se familiaricen fácilmente con ella, disponga de documentación para su uso, cuente con el apoyo de comunidades dinámicas de usuarios y con foros de usuarios, desarrolladores, técnicos y expertos.

Sobre el listado anterior se aplicará, para cada uno de los puntos, una matriz de decisión de la teoría DMD. Estas matrices están compuestas de una serie de criterios, es decir, atributos o características que definen cada uno de los objetivos que se acaban de enunciar, de un peso para cada criterio en función de su importancia para la toma de decisión y por último, de las alternativas (las distintas plataformas) que se van a valorar con la teoría DMD.

A la hora de asignar el peso de cada criterio, se ha tenido en consideración las encuestas realizadas por el proyecto GRAPPLE [83] a usuarios (estudiantes, tutores, investigadores y desarrolladores) de siete instituciones europeas sobre las características y dimensiones clave para un sistema de aprendizaje adaptativo. Los resultados obtenidos sobre las características y dimensiones más destacadas por los usuarios, aparecen ordenados por importancia en la tabla 8.1.

Características de la adaptabilidad	<ol style="list-style-type: none"> <li>1. Conocimiento del alumno</li> <li>2. Objetivos/tareas de aprendizaje</li> <li>3. Idioma</li> <li>4. Plataforma</li> <li>5. Intereses del alumno</li> <li>6. Estilo de aprendizaje/cognitivo</li> <li>7. Calificación del estudiante</li> <li>8. Rol del usuario</li> <li>9. Motivación del alumno</li> <li>10. Preferencias personales</li> <li>11. Localización del usuario</li> <li>12. Formación previa</li> <li>13. Factores de la personalidad</li> <li>14. Experiencia en la navegación</li> </ol>	Dimensiones de la adaptabilidad	<ol style="list-style-type: none"> <li>1. Selección de actividades de aprendizaje adaptativas</li> <li>2. Selección de contenido adaptativo (extra, previo y comparativo)</li> <li>3. Apoyo en la resolución de problemas adaptativos</li> <li>4. Agrupación de modelos de alumnos (apoyo colaborativo)</li> <li>5. Presentación adaptativa</li> </ol>
-------------------------------------	---	---------------------------------	--

**Tabla 8.1:** Listado de las especificaciones o requisitos de los usuarios de LMS

		Criterios											
		Guiado adaptativo	Anotación de enlaces	Ocultación de enlaces	Personalización entorno	Arquitectura extensible	Seguimiento estudiante	Tests/notas automáticas	Administración curso	Apoyo creador curso	Plantillas	Librerías o repositorios	Sistema de ontologías
Alternativas	Moodle				X	X	X	X	X	X	X	X	X
	Dokeos					X		X		X	X	X	X
	ATutor	X	X	X			X	X	X	X		X	
	Ilias					Baja		X	X	X	X	X	X
	Sakai CLE				X	X	X	X	X	X	X	X	X
	.LRN				X	Baja	X	X	X	X	X	X	X
	Drupal		X	X	X	X	X	X	X	X	X	X	X
$w_j$		5	4	4	5	5	5	5	4	4	4	4	5

**Tabla 8.2:** Matriz de decisión para valorar la Adaptabilidad de las plataformas

		Criterios				
		Accesibilidad W3C <sup>10</sup>	IMS CP <sup>11</sup>	IMS QTI <sup>12</sup>	SCORM <sup>13</sup>	IEEE LOM RDF
Alternativas	Moodle		X	X	X	X
	Dokeos	X	X	X	X	X
	ATutor	X	X	X	X	X
	Ilias		X	X	X	X
	Sakai CLE	X	X	X	X	X
	.LRN	X	X	X	X	X
	Drupal					
$w_j$		3	4	4	4	5

**Tabla 8.3:** Matriz de decisión para valorar la Estandarización de las plataformas

		Criterios					
		Autenticación usuario	Asignación privilegios	Registro estudiantes	Grupo de trabajo	Autovaloraciones	Modelo del alumno
Alternativas	Moodle	X	X	X	X	X	
	Dokeos	X	X		X	X	
	ATutor	X	X	Baja			
	Ilias	X	X		X	X	
	Sakai CLE	X	X	Baja		X	
	.LRN	X	X	Baja	X	X	
	Drupal	X	X	X	X	X	
$w_j$		3	3	5	4	4	5

Tabla 8.4: Matriz de decisión para valorar la Gestionabilidad de las plataformas

		Criterios												
		Foro	Noticias	Correo	Gestión de archivos	Calendario	Votaciones (polls)	Blog	Wiki	Chat	Videoconferencia	Pizarra	Soporte RSS	Encuestas (surveys)
Alternativas	Moodle	X	X	X	X	X	X	X	X	X			X	X
	Dokeos	X	X	X	X	X	X	X	X	X	X		X	X
	ATutor	X	X	X	X		X	X	X	X		X	X	X
	Ilias	X	X	X	X				X	X			X	X
	Sakai CLE	X	X	X	X	X	X	X	X	X			X	X
	.LRN	X	X	X	X	X		X	X				X	X
	Drupal	X	X	X	X	X	X	X	X	X	X	X	X	X
$w_j$		2	2	2	2	2	2	2	2	2	2	2	3	

Tabla 8.5: Matriz de decisión para valorar la Interactividad de las plataformas

		Criterios									
		S.O. Window	S.O. Linux	Servidor Web Apache	Servidor Web Aolserver	Servidor Web MS ISS	BDD MySQL	BDD PostgreSQL	Lenguaje Prog. PHP	Lenguaje Prog. TCL	Lenguaje Prog. Java
Alternativas	Moodle	X	X	X			X	X	X		
	Dokeos		X	X			X		X		X
	ATutor	X	X	X			X		X		X
	Ilias	X	X	X			X		X		
	Sakai CLE		X	X			X				X
	.LRN		X		X			X		X	
	Drupal	X	X	X		X	X	X	X		
$w_j$		2	1	2	1	1	2	2	2	1	2

**Tabla 8.6:** Matriz de decisión para valorar la Tecnología Empleada de las plataformas

		Criterios				
		Comunidad de Usuarios	Portales o Webs	Idiomas	Documentación	Foros de discusión
Alternativas	Moodle	$29 \cdot 10^6$	$51 \cdot 10^3$	75	X	X
	Dokeos	$1 \cdot 10^6$	$6 \cdot 10^3$	37	X	X
	ATutor		$1 \cdot 10^3$	32	X	X
	Ilias		196	22	X	X
	Sakai CLE		160	2	X	X
	.LRN	$1 \cdot 10^5$	34	3	X	X
	Drupal			42	X	X
$w_j$		1	1	2	2	2

**Tabla 8.7:** Matriz de decisión para valorar la Internacionalización de las plataformas

A continuación se ha seleccionado la ponderación lineal [81] como método de DMD. Su esquema básico consiste en construir una función de valor  $U(A_i)$  para cada alternativa:

$$U(A_i) = \sum_j w_j r_{ij}$$

Siendo  $w_j$  el peso del criterio  $j$  y  $r_{ij}$  la evaluación (rating) de la alternativa  $i$  respecto al criterio  $j$ .

		$U(A_i)$
Alternativas	<b>Moodle</b>	<b>119</b>
	Dokeos	103
	ATutor	104
	Ilias	92
	Sakai CLE	108
	.LRN	106
	<b>Drupal</b>	<b>113</b>

**Tabla 8.8:** Resultados de la Ponderación Lineal para las distintas Alternativas

El resultado de la tabla 8.8 muestra que la plataforma más idónea (por haber obtenido una mayor puntuación) es Moodle.

Con la idea de subsanar las carencias pedagógicas que limitan a Moodle respecto a la presentación y secuenciación de recursos educativos para un aprendizaje adaptativo del alumno, el artículo [84] estudia una integración entre Moodle y el modelo IMS Learning Design (IMS LD), capaz de definir flujos de aprendizaje adaptativos, roles flexibles y contenido adaptativo en las diferentes Unidades de Aprendizaje.

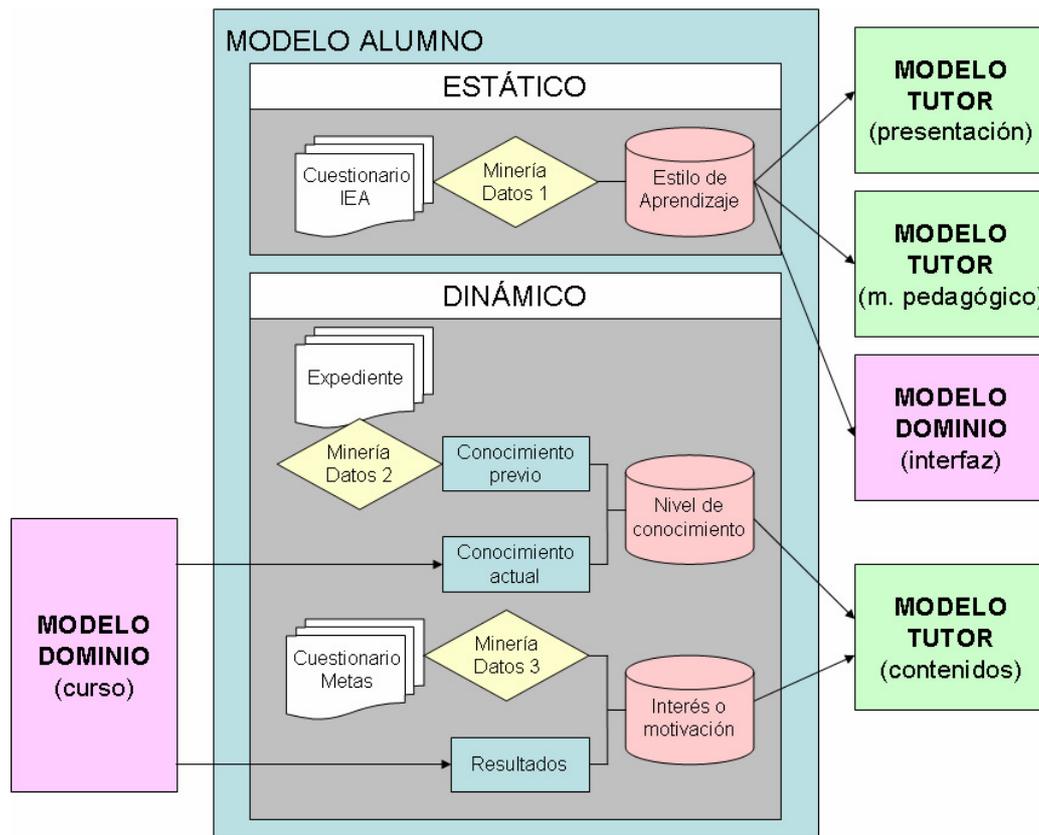
Por otro lado, el trabajo desarrollado en [73] permite la integración del motor de flujo de trabajo jBPM en Moodle con el fin de estructurar las interacciones de los usuarios y la información que intercambian en las herramientas colaborativas (chats y foros). Así se pueden asignar diferentes actividades a diferentes participantes, determinar el final de una actividad y secuenciar las actividades.

## Modelo del Alumno

Evalúa la información o actuación de cada alumno para determinar su conocimiento, habilidad perceptual y capacidad de razonamiento. El modelo del alumno también incluye sus carencias y malas concepciones de aprendizaje.

El modelo está compuesto por dos bloques diferenciados, como se aprecia en la Figura 8.1: uno que modela los rasgos estáticos que no varían a lo largo del tiempo (características cognitivas y preferencias de aprendizaje) y otro que recoge las características dinámicas (estado del conocimiento y del interés o motivación del alumno), que se actualizan constantemente porque dependen de la interacción del alumno con el curso LMS o Modelo del Dominio.

A su vez, los dos submodelos del Alumno, alimentan a varios aspectos o rasgos que caracterizan el Modelo Tutor, definiendo la adaptabilidad del ITS. Con estas relaciones se establece el método pedagógico de enseñanza adaptativo, los contenidos adecuados (secuenciación de los recursos u objetos de aprendizaje) y el modo de presentarlos (tipo de recursos multimedia o colaborativos), todo ello con el fin de planificar la ruta de aprendizaje óptima para el alumno en función de su perfil o modelo de aprendizaje. Por otro lado, también realimentan el Modelo del Dominio, indicando las preferencias del alumno con respecto al aspecto y al conjunto de herramientas que configuran la interfaz de usuario del LMS.



**Figura 8.1:** Esquema de los bloques que conforman el Modelo del Alumno

A continuación se estudia con más detalle cada uno de los submodelos, definiendo los parámetros o variables que los conforman, que se han elegido de la bibliografía analizada en el capítulo 4. Una vez definidos los parámetros, éstos se ajustarán al aplicar el método de minería de datos (ver apartado 8.3) al conjunto del sistema ITS. Los modelos se almacenan en una base de datos relacional según el esquema E/R (ver apartado 8.2), cuya estructura y características se definen seguidamente y se ejemplifican en las figuras 8.2 y 8.3:

1. **Entidades:** representan conjuntos de elementos con existencia propia y que se caracterizan por las mismas propiedades (i.e. Alumno y Curso).
2. **Atributos:** son las características o cualidades propias de cada entidad (i.e. DNI, Nombre, Dirección, e-mail, Teléfono y Nacionalidad). Existe un identificador principal (IP) único para cada ejemplar de la entidad (i.e. DNI).
3. **Dominios:** son el conjunto de valores sobre los que se definen los atributos (i.e. los valores para Nacionalidad: española o extranjera).
4. **Interrelaciones:** representan asociaciones del mundo real entre una o más entidades. Existen varios tipos: de uno a muchos o de muchos a muchos. Para distinguirlas se introducen las propiedades de una interrelación:
  - **Grado:** número de entidades que intervienen en la interrelación (i.e. dos: Alumno y Curso).
  - **Tipo de correspondencia:** número máximo de ejemplares de un tipo de entidad que pueden estar asociados con un ejemplar del otro tipo de entidad (i.e. de muchos a muchos M:N).
  - **Cardinalidad:** el número mínimo y el máximo de ejemplares de un tipo que pueden relacionarse con un elemento de otro tipo de entidad (i.e. un alumno puede realizar de 1 a n cursos (1,n)).

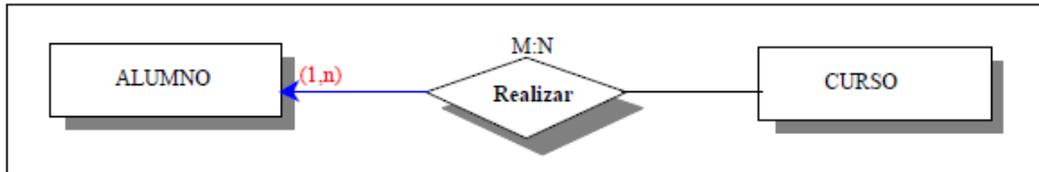


Figura 8.2: Ejemplo de los elementos básicos del modelo E/R

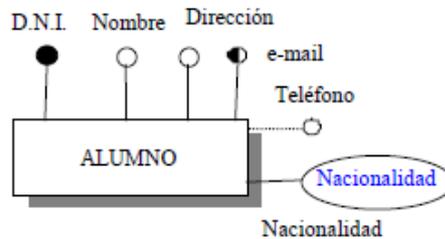


Figura 8.3: Ejemplo de los atributos de una entidad

En primer lugar, en el submodelo estático basado en el trabajo [46], nos encontramos con los datos que proporcionan el cuestionario sobre el Índice de Estilos de Aprendizaje (IEA) (consultar el Anexo 1), que consta de 44 preguntas correspondientes a las 4 dimensiones, cada una con dos respuestas posibles que determinan el estilo. La agrupación en clusters de los resultados del cuestionario determina el estilo de aprendizaje de Felder que se almacena en la base de datos del ITS como la entidad “Estilo de Aprendizaje” que viene caracterizada por tres atributos: dimensión, estilo y cluster, como se aprecia en la Figura 8.4. Los valores que pueden tomar los atributos vienen especificados en la Tabla 8.9, cuya descripción detallada se encuentra en la Tabla 4.1.

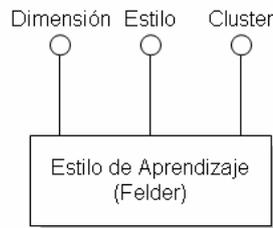


Figura 8.4: Atributos de la entidad Estilo de Aprendizaje

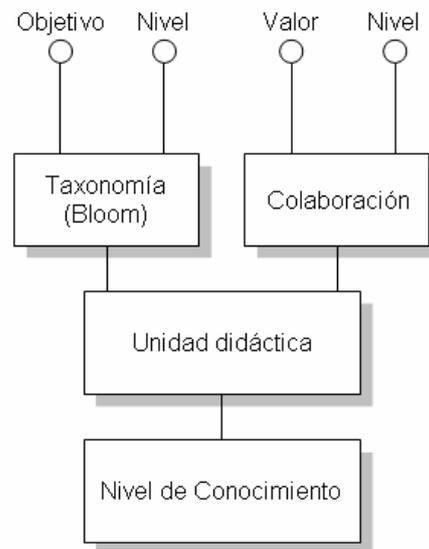
Dimensión	Estilo	Cluster	
Procesamiento	Activo	Fuerte	
	Reflexivo		
Percepción	Sensorial		Moderado
	Intuitiva		
Entrada	Visual	Equilibrado	
	Verbal		
Comprensión	Secuencial		
	Global		

Tabla 8.9: Descripción de los valores de los atributos de la entidad Estilo de Aprendizaje

En el submodelo dinámico existen dos entidades más que se añaden a la base de datos del ITS, la entidad “Nivel de Conocimiento” y la entidad “Motivación”. El Nivel de conocimiento se inicializa con el Nivel de Conocimiento Previo, que se extrae de un cuestionario sobre las materias cursadas anteriormente y que constituyen el expediente académico del alumno. Por otro lado, cuando el alumno comienza el curso, se crea el Nivel de Conocimiento Actual a partir de la actuación del alumno en el curso y que reemplaza al Nivel de Conocimiento Previo. Éste se infiere de la interacción del alumno con el sistema LMS, de su participación en las actividades, de la consulta de recursos y de los resultados obtenidos. Toda esta información que se extrae de los ficheros log (ver capítulo 7.1) y que se almacena en la base de datos, se analiza mediante minería de datos obteniendo unos valores para los atributos que definen la entidad Nivel de Conocimiento (Figura 8.5).

Un curso está estructurado en unidades didácticas, que a su vez están formadas por objetos de aprendizaje definidos con metadatos (LOM). Los LOM pueden ser actividades, talleres, tareas, recursos multimedia, prácticas de laboratorio, etc. La elección y secuenciación de los LOM que conformarán la unidad didáctica se realiza en el Modelo Tutor, en base al currículum del curso almacenado en el Modelo Dominio.

La entidad “Nivel de conocimiento” gestiona los atributos para cada “Unidad Didáctica” que el alumno supera o realiza dentro del curso, dando un valor o resultado global del nivel de conocimiento o progreso del alumno en el curso, en su conjunto. El nivel de conocimiento se basa en la combinación de la taxonomía de Bloom [51] y de la competencia colaborativa [46] determinando unos atributos para cada entidad, como se observa en la Figura 8.5.



**Figura 8.5:** Atributos de la entidad Nivel de Conocimiento

En la tabla 8.10 se resumen los valores que pueden tomar los atributos de la Taxonomía de Bloom y de la Competencia colaborativa, cuyas descripciones detalladas se encuentran en el capítulo 4.2.

Objetivo Bloom	Nivel	Valor de Colaboración	Nivel
Conocimiento	Principiante	Estudiante no colaborativo	Bajo
Comprensión		Estudiante comunicativo	Medio
Aplicación	Medio	Estudiante participativo	
Análisis		Estudiante con iniciativa	
Síntesis	Experto	Estudiante perspicaz	Alto
Evaluación		Estudiante útil	

**Tabla 8.10:** Descripción de los valores de los atributos de la entidad Nivel de Conocimiento

En [84] se realiza un estudio de las metas académicas que persiguen los alumnos y su influencia sobre la motivación. Se define la motivación como el conjunto de procesos implicados en la activación, dirección y persistencia de la conducta. En el trabajo se distingue entre metas de aprendizaje y metas de ejecución o rendimiento, así como los patrones motivacionales respectivos.

Se pueden diferenciar cuatro categorías de metas académicas (relacionadas con la tarea, con la autovaloración o con el “yo”, con la valoración social y con la consecución de recompensas externas) que tienen asociadas motivaciones o miedos, las dos primeras están relacionadas directamente con el aprendizaje o el logro académico y las dos últimas, de forma indirecta, ya que sirven para favorecerlo (ver la tabla 8.11).

Meta	Motivación	Descripción
Tarea	Motivación de competencia	Incrementar la propia competencia
	Motivación intrínseca	Sentirse absorbido por la naturaleza de la tarea
	Motivación de control	Actuar con autonomía y no obligado
Auto-valoración	Motivación de logro	Experimentar el orgullo y satisfacción que sigue al éxito
	Miedo al fracaso	Evitar la experiencia de “vergüenza” o “humillación” que acompaña al fracaso
Valoración social	Aceptación tutores	Experimentar la aprobación de los adultos y evitar su rechazo
	Aceptación iguales	Experimentar la aprobación de los iguales y evitar su rechazo
Recompensa externa	Aceptación social	Conseguir todo lo que signifique premios o recompensas
	Miedo al castigo	Evitar todo lo que signifique castigo o pérdida de situaciones, objetos o posibilidades valoradas

**Tabla 8.11:** Metas de la actividad escolar

En las tablas 8.12 y 8.13 se describen los dos tipos de orientación motivacional (de carácter intrínseco o extrínseco) y los patrones motivacionales, asociados a los dos tipos de metas (de aprendizaje o de rendimiento) lo que condiciona las acciones o tareas a plantear al alumno (Modelo Tutor: planificación de la ruta de aprendizaje).

Meta	Carácter motivacional	Descripción
Aprendizaje	Intrínseca	Deseo de dominio, curiosidad, preferencia por el reto, interés por aprender
Rendimiento	Extrínseca	Obtención de notas, recompensas, juicios positivos, aprobación de padres y profesores y evitación de todo tipo de valoraciones negativas

*Tabla 8.12:* Metas y orientaciones motivacionales

Meta	Confianza	Tarea	Respuestas	Patrón
Rendimiento	Baja	Moderadamente fácil, para evitar muestras de incompetencia	Percepción de fracaso, afecto negativo y atribuciones a la capacidad	Indefensión
Rendimiento	Alta	Moderadamente difícil si las tareas aseguran muestras de competencia	Alguna preocupación por el rendimiento, eficaz solucionando el problema	Dominio atenuado
Aprendizaje	Alta o baja	Moderadamente difícil, de cara a incrementar la competencia	Afecto positivo, eficaz solucionando el problema	Dominio

*Tabla 8.13:* Metas, niveles de confianza y patrones de logro

La forma de obtener las metas de los alumnos y, consecuentemente, inferir su interés o motivación académica, consiste en la realización de un Cuestionario de Metas para Adolescentes (CMA) [86] (ver anexo 2) formado por 79 ítems o preguntas, agrupados en nueve escalas: educativa, profesional, interpersonal, auto imagen, reconocimiento social, emancipativa, riesgo, física y compromiso personal. Este cuestionario ha sido facilitado por Dña. María Luisa Sanz de Acedo Lizarraga, profesora de la Universidad Pública de Navarra.

Del desarrollo y validación de este cuestionario se ha extraído la relación existente entre los factores que definen las metas y sus correlaciones, siendo los factores determinantes para inferir la orientación motivacional los siguientes:

- **Reconocimiento social (Rs):** Busca ofrecer una imagen positiva y competente, obtener éxito y prestigio en cualquier actividad y ser superior a los demás.
- **Educativo (Ed):** Se relaciona con el esfuerzo por aprender, el logro de resultados académicos, la regulación del aprendizaje, el uso de estrategias, la adquisición de nuevos conocimientos y la aprobación del profesor.
- **Interpersonal (In):** Aprecia las habilidades sociales, la amistad, la pertenencia a grupos, la solidaridad, la asertividad y la empatía.

De modo que el factor Reconocimiento social condiciona el carácter motivacional Extrínseco y el factor Educativo el carácter motivacional Intrínseco, según la tabla 8.12. Y por otro lado, el factor Interpersonal condiciona el Valor Colaborativo del Estudiante.

Para determinar qué tipo de motivación tiene el alumno, se realizará el cuestionario CMA con los ítems relativos a los factores que se han descrito (Rs, Ed e In). En total son 39 cuestiones (ver el anexo 3) cada una de ellas con una puntuación de 1 (si el ítem

es nada importante) a 6 puntos (si el ítem es importantísimo). En el estudio realizado por [86], se han obtenido unos valores medios y unas desviaciones estándar en función del sexo y edad de los estudiantes, así como del tipo de centro en el que estudian (ver tabla 8.14).

Escala		Sexo		Edad-años			Tipo de centro	
		Mujer	Hombre	15/16	17	18/19	Público	Concertado
Rs	$\bar{x}$	34,13	38,62	37,42	36,11	34,81	35,89	37,48
	$\sigma$	10,27	11,52	11,48	11,06	10,36	10,96	11,46
Ed	$\bar{x}$	44,67	43,55	45,75	44,11	40,34	43,80	44,58
	$\sigma$	8,23	8,74	8,39	8,22	8,00	8,69	8,19
In	$\bar{x}$	69,76	63,92	66,83	66,88	66,23	66,59	66,94
	$\sigma$	9,10	9,82	9,74	10,00	10,22	10,06	9,67

Tabla 8.14: Medias (M) y desviaciones estándar (DE) por sexo, edad y tipo de centro

Para este trabajo interesan los valores de los alumnos con una edad comprendida entre 18 y 19 años, que son los que estudian en la Universidad. Para considerar que el carácter motivacional es dominante, se ha tomado como límite inferior la media más la desviación típica y como límite superior la puntuación máxima (PM) que se puede obtener en cada escala. Si el factor tiene una graduación, como es el caso del factor Interpersonal, que puede tomar 3 valores colaborativos (pasivo, mixto, activo) se ha considerado  $0 < In < \bar{x} - \sigma$ ,  $\bar{x} - \sigma < In < \bar{x} + \sigma$  y  $\bar{x} + \sigma < In < PM$ , respectivamente. Los resultados se muestran en la tabla 8.15 y 8.16 respectivamente.

Nivel	Carácter motivacional			
	Intrínseca		Extrínseca	
	Dominante	Pasivo	Dominante	Pasivo
Valores CMA	$48 < Ed < 66$	$0 < Ed < 48$	$45 < Rs < 84$	$0 < Rs < 45$

Tabla 8.15: Valores del CMA que determinan el carácter motivacional del alumno

Valores CMA	Valor de colaboración
$77 < In < 84$	Activo
$56 < In < 77$	Mixto
$0 < In < 56$	Pasivo

Tabla 8.16: Valores del CMA que determinan el valor de colaboración del alumno

Después de analizar las tablas anteriores, se definen cuáles son los atributos que caracterizan la entidad “Motivación” (ver figura 8.7) así como los valores que pueden tomar (ver tabla 8.17).

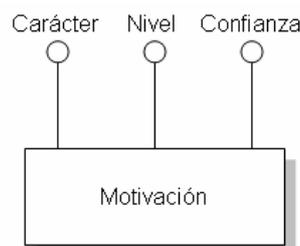


Figura 8.7: Atributos de la entidad Motivación

Carácter	Nivel	Confianza
Intrínseca	Dominante	Alta
Extrínseca	Pasivo	Baja

**Tabla 8.17:** Descripción de los valores de los atributos de la entidad Estilo de Aprendizaje

En la figura 8.7 se obtienen los atributos de la entidad “Motivación” a partir de los datos recopilados del cuestionario CMA, en el cuál se ha realizado una clasificación o agrupación para determinar el carácter motivacional a partir de las metas educativas del alumno y el nivel de dominancia. El valor de los atributos se ve reforzado o complementado gracias a los datos que se extraen de la interacción del usuario con el curso LMS, del cuál se analizan los resultados que obtiene el alumno en las distintas tareas, concluyendo cuál es su rendimiento académico, lo que condiciona o favorece el patrón de logro, la confianza y la respuesta del alumno, según la tabla 8.13.

Para determinar cuál es la confianza inicial del alumno, antes de realizar el curso, se ha acudido a un cuestionario. La Confianza académica es una de las cinco dimensiones de la Autoestima del alumno, junto con la dimensión social, emocional, familiar y física. El estudio realizado por [87] apoya el uso de la escala de autoestima de Rosenberg (RSES) para evaluar la autoestima en el contexto educativo universitario. El RSES consta de 10 ítems que se refieren al autorespeto y la autoaceptación (ver anexo 4), puntuados con una escala de tipo Likert con 4 puntos, asignando el valor de 1 (si está totalmente en desacuerdo) al valor de 4 (si está totalmente de acuerdo). Los ítems 1, 3, 4, 7 y 10 son valorados positivamente (en el orden de 1 a 4 puntos), y los ítems 2, 5, 6, 8 y 9 son valorados negativamente (en el orden de 4 a 1 puntos).

Los resultados extraídos de la realización del cuestionario en [87] muestran los valores medios y las desviaciones estándar en función del sexo de los estudiantes (ver tabla 8.18).

RSES	Mujer	Hombre
$\bar{x}$	31.14	32.53
$\sigma$	4.55	3.92

**Tabla 8.18:** Medias (M) y desviaciones estándar (DE) de la puntuación total por sexo en la RSES

Para este trabajo se considera un índice de confianza tipificado en alto o bajo, para ello se ha seleccionado el valor medio de los resultados de toda la población encuestada, tanto hombres como mujeres, así como la media de la desviación típica, de modo que un valor alto de confianza será aquel comprendido entre la  $\bar{x} + \sigma$  y el valor máximo, y un valor bajo de confianza será aquel comprendido entre el valor mínimo y la  $\bar{x} - \sigma$ . Los valores intermedios (de 28 a 36) implican una confianza normal y no son relevantes a la hora de fijar el nivel de dificultad relativo a la motivación extrínseca. Los valores numéricos para el nivel de confianza se resumen en la tabla 8.19:

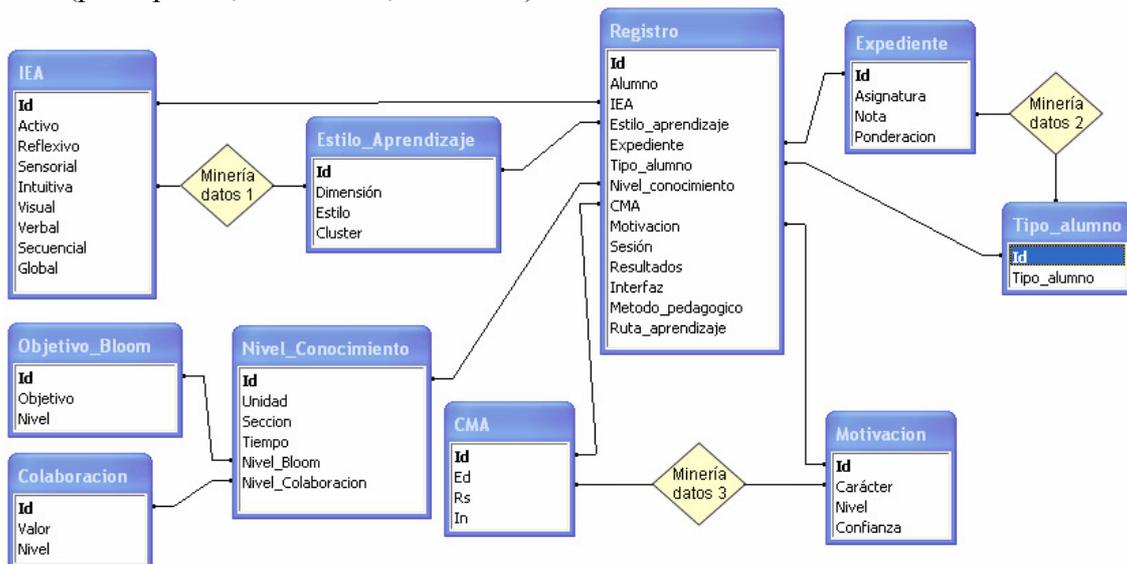
Valores	Nivel de confianza
$36 < RSES < 40$	Alto
$10 < RSES < 28$	Bajo

**Tabla 8.19:** Valores del RSES que determinan el valor de confianza del alumno

En la figura 8.8 se muestra un resumen de las relaciones entre las tablas que se han ido definiendo en este capítulo y que conforman la base de datos del Modelo del Alumno. Estas relaciones se irán completando con las tablas que se definirán en los siguientes modelos (Dominio y Tutor) con el fin de describir el conjunto del modelo LMS. Los rombos con la inscripción “Minería de datos” indican una extracción de información de dicha tabla para, a partir de un análisis de minería de datos que se describe en el capítulo 8.3, rellenar la siguiente tabla con la información obtenida. Por ejemplo, a partir de los datos del índice IEA aplicando la Minería de Datos 1 se obtienen las dimensiones del Estilo de Aprendizaje del alumno.

La tabla Registro muestra aquí únicamente los campos utilizados en las tablas que se han definido hasta el momento, de modo que, en las siguientes representaciones (figuras 8.14 y 8.19) se ampliarán dichos campos a medida que se introduzcan nuevas relaciones. El Registro almacena el número identificador del alumno, su nombre de usuario, el índice IEA, su estilo de aprendizaje, el expediente académico, las competencias iniciales y actuales, que hacen referencia a los objetivos académicos alcanzados, el nivel de conocimiento (los objetivos de Bloom y el nivel de colaboración), el resultado del cuestionario CMA y la motivación.

La tabla Expediente almacena las asignaturas cursadas con anterioridad por el alumno que tienen relación con el curso actual y que establecen el conocimiento inicial del alumno sobre la materia a estudiar. En la tabla, se acumula la nota obtenida y un factor de ponderación, que implanta el profesor en función de la importancia de dicha asignatura sobre el curso a impartir, dando una noción del tipo de alumno del que se trata (principiante, intermedio, avanzado).



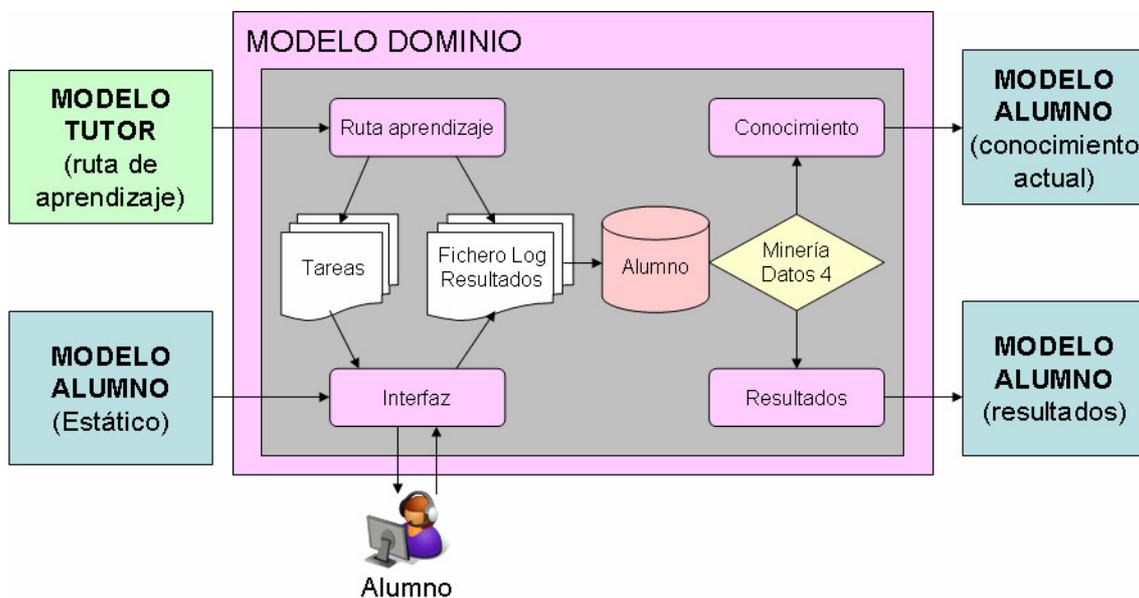
**Figura 8.8:** Relaciones entre las tablas que forman la base de datos del Modelo del Alumno

## Modelo del Dominio

Permite al ITS comparar las acciones y elecciones del estudiante con las del sistema experto, con el objetivo de evaluar lo que el usuario conoce y desconoce. Contiene la descripción del conocimiento o comportamiento que representa el dominio o campo de la enseñanza, es decir, los recursos educativos del curso (representado por un sistema experto o modelo cognitivo).

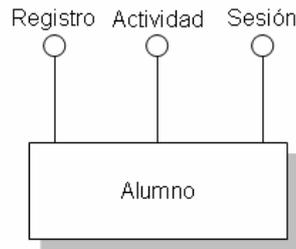
En la figura 8.9 se observa la estructura del Modelo del Dominio, que está compuesto por la Ruta de aprendizaje que viene definida por el Modelo Tutor. Esta Ruta de aprendizaje contiene los LOM y la secuenciación a aplicar para definir la unidad didáctica, que se debe materializar en una serie de tareas (actividades, talleres, recursos, prácticas, etc.) que vienen diseñados y condicionados por el Curso y que se ofrecen al Alumno a través de una Interfaz de Usuario, adaptada a las características del Estudiante, gracias a los datos que facilita el Modelo de Alumno Estático.

La interacción del Alumno con las Tareas en la Interfaz produce una serie de informes, que almacenan esta interacción (ficheros log) así como los resultados de la realización de las Tareas, al comparar las respuestas o acciones del Alumno, con las del Modelo Experto o Cognitivo definido en el Curso. El análisis de esta información (los ficheros log y los resultados) mediante minería de datos proporciona el progreso del alumno, que se puede descomponen en resultados o logros que afectarán a la motivación del alumno y en el nivel de conocimiento actual o aprendizaje adquirido hasta la fecha.



**Figura 8.9:** Esquema de los bloques que conforman el Modelo del Dominio

La base de datos Alumno almacena toda la información relativa a las actividades realizadas durante el proceso de enseñanza/aprendizaje en el sistema LMS: duración de la actividad, recursos consultados, tareas realizadas, resultados obtenidos, etc. Se ha modificado la tabla 7.6 (ver el capítulo 7.4) que contiene la información relativa a los resultados académicos, con el fin de extraer los parámetros que definen los atributos de la base de datos Alumno, que se listan a continuación y que posteriormente se analizarán con minería de datos para lograr la base de datos del Conocimiento y de los Resultados, que realimentarán el sistema de control del proceso de aprendizaje del alumno.



**Figura 8.10:** Atributos de la entidad Alumno

En la figura 8.10 se representan gráficamente los atributos que conforman la entidad Alumno. Por un lado está la tabla Registro que viene definido por la entrada del estudiante en el sistema LMS y que vendrá validado por una clave de acceso para garantizar la identidad del alumno. Por otro lado se tiene el conjunto de actividades que realizará el estudiante durante la sesión de trabajo, que se almacenan en la tabla Actividad, que tiene los campos que se describen en la tabla 8.20. Por último, la tabla Sesión almacena la duración de las sesiones de registro del alumno, y el listado de actividades que ha ejecutado.

Campos	Descripción
Nombre	Indica qué actividad se ha realizado
Acción	Es el tipo de actividad que realiza el alumno: lectura, escritura, descarga o subida de material. Según la dimensión de Procesamiento de Felder, indica si el alumno es activo o reflexivo
Opciones	En el caso de simulaciones, se almacenan las opciones del menú seleccionadas para realizar la actividad.
Interacción_ alumno	En función de la actividad, determina el grado de participación y colaboración del alumno, que puede ser baja, media o alta.
Tiempo_inicio Tiempo_fin	Se almacena el inicio de la actividad y su finalización. Con ello se obtiene la duración de la misma.
Votación	Sirve como realimentación del éxito de la actividad, donde el alumno opina indicando si la actividad le ha gustado o no
Nota	Es la calificación que ha obtenido el alumno (insuficiente, suficiente, bien, notable, sobresaliente). En caso de dejarla en blanco o tratarse de una lectura que se ha leído o no, se asignarán, por defecto los valores “no presentado”, “realizado” o “no realizado” respectivamente.
Aciertos	Contabiliza el número de aciertos y por consiguiente el número de fallos. En función de éstos, se puede sugerir ayuda al alumno.
Petición_ayuda	Queda reflejado si el alumno ha consultado otras fuentes para resolver la actividad, por medio de la URL seguida (páginas Web con recursos, chat, foro, blog, Wiki) o bien ha escrito un correo al profesor o a un compañero

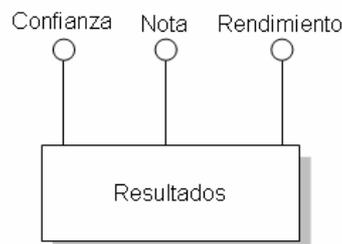
**Tabla 8.20:** Descripción de los campos que conforman la tabla Actividad

Las relaciones entre estos tres atributos de la tabla Alumno se pueden apreciar en la figura 8.11. Además se añade la relación con la tabla Tarea que almacena todos los recursos existentes en la base de datos LOM, definidos por el profesor y basados en lo que establece el currículum. La descripción del significado de los campos que forman la Tarea se puede consultar en la tabla 8.23 del siguiente capítulo, el Modelo Tutor.



**Figura 8.11:** Relaciones entre los campos que forman la tabla Alumno

De los datos recogidos en la base de datos Alumno, se procede a la aplicación de un análisis mediante el método de minería de datos (ver el capítulo 8.3) con el fin de obtener tres bases de datos diferenciadas que realimentan al Modelo del Alumno. Por un lado, se tiene la base de datos que almacena el Nivel de Conocimiento (ver figura 8.5), por otro lado los Resultados (ver figura 8.12) y por último, la Competencia actual que determina el nivel de cumplimiento de las competencias generales y específicas del curso que el alumno tiene en este momento. Los Resultados se definen mediante tres atributos (ver la tabla 8.21): el Rendimiento, que es el cociente entre las notas obtenidas por el alumno y el tiempo invertido en el estudio de curso, la Nota, que son los valores con los que se califica o evalúa la actuación del alumno en las pruebas y la Confianza, que cuantifica el número de intentos o las revisiones que necesita el alumno para superar una tarea.



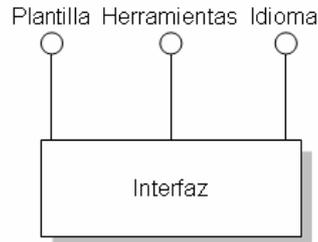
**Figura 8.12:** Atributos de la entidad Alumno

<b>Rendimiento</b>	<b>Nota</b>	<b>Confianza</b>
Bajo	Insuficiente	Baja
Medio	Suficiente	Alta
Alto	Bien	
	Notable	
	Sobresaliente	

**Tabla 8.21:** Descripción de los valores de los atributos de la entidad Resultados.

La interfaz está compuesta de varios atributos como se aprecia en la figura 8.13 cuya descripción se encuentra esquematizada en la tabla 8.22. Por un lado se tiene la Plantilla, que define la apariencia estética del sistema LMS (el fondo del escritorio, los estilos y formatos de los caracteres, la distribución de los menús o portlets y el número

o tipo de barras de herramientas visibles) y el Idioma de trabajo de la aplicación y de los materiales del curso. Ambos parámetros se seleccionarán manualmente por el usuario al comenzar el curso, y se almacenarán de una sesión a la siguiente, pudiéndose cambiar en cualquier momento si el usuario lo desea.



**Figura 8.13:** Atributos de la entidad Interfaz

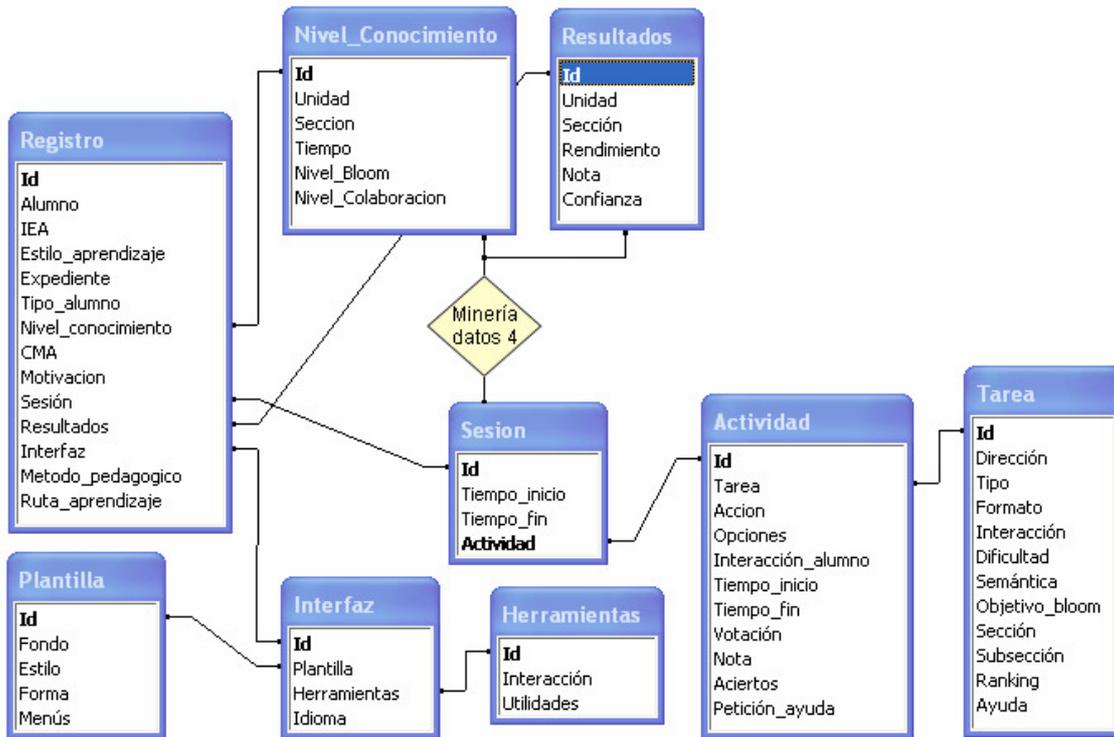
El atributo Herramientas viene condicionado por el Modelo del Alumno Estático, es decir, por su Estilo de Aprendizaje, distinguiéndose dos tipos o grupos de herramientas: de interacción (foro, Chat, diario o blog, wiki, votaciones) o utilidades (agenda o calendario, procesador de texto, presentación, multimedia, vídeo, audio, noticias o feeds, aplicación Web o simulación).

Plantilla	Herramientas	Idioma
Fondo (tema)	Interacción	Español
Estilos	Utilidades	Valenciano
Forma (distribución)		Inglés
Menús		

**Tabla 8.22:** Descripción de los valores de los atributos de la entidad Interfaz

En la figura 8.14 se plasman las relaciones entre las tablas que se emplean para definir el Modelo del Dominio. Como se ha comentado, las nuevas tablas pasan a formar parte, mediante los identificadores, de los campos de la tabla Registro. Así mismo, aplicando minería de datos a la información almacenada en la tabla Sesión, se extraen el nivel de conocimiento, los resultados y competencias actuales del alumno.

La Ruta de Aprendizaje viene determinada por el Modelo del Tutor que se estudia en el siguiente apartado.

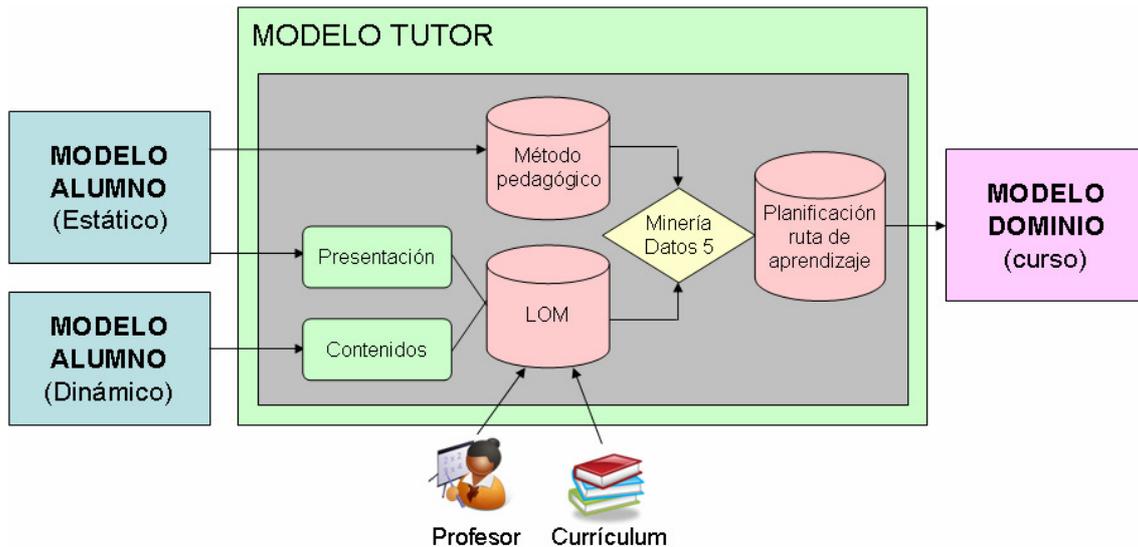


**Figura 8.14:** Relaciones entre las tablas de la base de datos del Modelo del Alumno y del Dominio

## Modelo del Tutor

El Modelo del Tutor codifica los métodos de enseñanza que son apropiados para el dominio objetivo y el estudiante. Es el motor de ejecución del sistema adaptativo. En función del conocimiento, nivel de experiencia y estilos de aprendizaje de los alumnos, el Modelo del Tutor selecciona la intervención educativa más adecuada. Ésta se realiza comparando el Modelo del Alumno Dinámico con los resultados esperados del Modelo del Dominio (Curso), las discrepancias observadas son señaladas al Modelo Tutor que toma una acción correctiva acorde.

En la figura 8.15 se observa la estructura del Modelo del Tutor. Inicialmente se opta por el método pedagógico más adecuado (que responde a la pregunta ¿cómo enseñar?) a partir de los datos proporcionados por el Modelo del Alumno Estático (el estilo de aprendizaje). Por otro lado, estas preferencias del alumno determinan qué tipo de materiales multimedia le gustan, que junto con los contenidos que debe aprender, que vienen definidos por el Modelo del Alumno Dinámico (nivel de conocimiento) permiten seleccionar los LOM que mejor se ajustan a sus necesidades (este apartado responde a la pregunta ¿qué enseñar?). Estos dos bloques sirven para planificar la ruta de aprendizaje que está definida por una secuenciación temporal de los LOM que conforman la unidad didáctica.

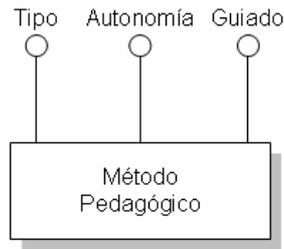


**Figura 8.15:** Esquema de los componentes que conforman el Modelo del Tutor

En [89] se analizan diferentes métodos pedagógicos de enseñanza propuestos por varios autores como son M. Adler, J. Piaget o A. Collins:

- **Instruccional:** se centra en la explicación, se exponen los qué y los porqués de un determinado tema.
- **Orientador:** consiste en asignar prácticas, motivar a los alumnos a reflexionar sobre lo que están haciendo y ofrecer retroalimentación.
- **Socrático:** el profesor controla la claridad de la información suministrada por los alumnos haciéndoles preguntas y alentándolos a examinar la información de manera crítica. Ayuda al alumno a comprender ciertos conceptos por sí mismo y darle la oportunidad de investigar y de aprender cómo hacerlo.
- **Constructivista:** considera al alumno como un agente activo, que trata de comprender concepciones tentativas y las contrasta con nuevos ejemplos, construyendo significados en respuesta a la situación educativa, con la guía del profesor.
- **Cooperativo y colaborativo entre iguales:** utilizan la dinámica de grupos para promover el aprendizaje reflexivo, ya que los estudiantes piensan y discuten juntos los problemas y explotan las motivaciones.
- **Situado en un contexto:** afirma que el verdadero aprendizaje debe situarse en una cultura de necesidades y prácticas que ofrece un contexto, una estructura y una motivación a los conocimientos y habilidades aprendidos. Hay una red social que sustenta el rendimiento y el aprendizaje.

De la descripción de los diferentes métodos pedagógicos anteriores se puede extraer una entidad denominada Método Pedagógico (ver figura 8.16) compuesta de tres atributos (ver tabla 8.23): el Tipo que incluye la clasificación anterior (instruccional, orientador, socrático, constructivista, cooperativo y contextual), la Autonomía, que mide el grado de investigación, búsqueda y autoaprendizaje del alumno, y por último el Guiado, que indica qué actor es el encargado de ofrecer ayuda y consejo en el proceso de aprendizaje, el tutor o profesor o bien, los compañeros del curso o los iguales del alumno.

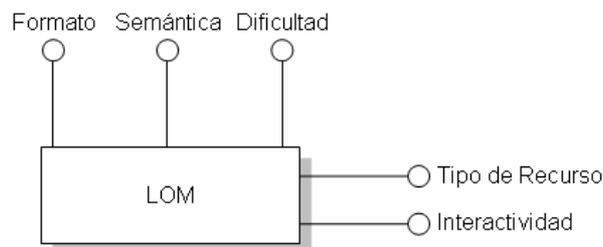


**Figura 8.16:** Atributos de la entidad Método Pedagógico

Tipo	Autonomía	Guiado
Instruccional	Baja	Profesor
Orientador	Media	
Socrático		
Constructivista	Alta	Iguales
Cooperativo		
Contextual		

**Tabla 8.23:** Descripción de los valores de los atributos de la entidad Método Pedagógico

En [46] se definen los atributos de los objetos LOM, como se puede ver en la figura 8.17, así como la descripción y los valores que toman (ver tabla 8.24). Estos atributos se han ampliado para definir la tabla Tarea, que es una modificación de los objetos LOM, en dónde se almacenan todos los recursos y actividades del sistema LMS. El listado de campos que forman la tabla Tarea se puede consultar en la tabla 8.25.



**Figura 8.17:** Atributos de la entidad LOM

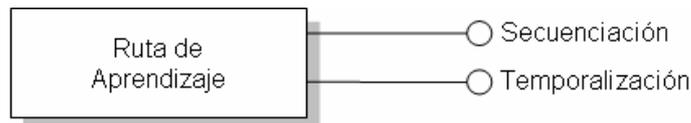
Tipo de Recurso		Formato	Semántica	Dificultad	Interactividad
Ejercicio	Tabla	Texto	Muy baja	Muy fácil	Activo
Simulación	Texto	Multimedia	Baja	Fácil	Mixto
Cuestionario	Examen	Gráfico	Media	Media	Pasivo
Diagrama	Gráfico	Vídeo	Alta	Difícil	
Experimento	Problema	Audio	Muy alta	Muy difícil	
Estadísticas	Auto-evaluación				
Conferencia					
Diapositiva	Índice				

**Tabla 8.24:** Descripción de los valores de los atributos de la entidad LOM

<b>Campos</b>	<b>Descripción</b>
Dirección	Define la ubicación del recurso mediante la dirección URL
Tipo	Indica el tipo de recurso (ejercicio, simulación, cuestionario, diagrama, experimento, estadísticas, conferencia, diapositiva, tabla, texto, examen, gráfico, problema, auto-evaluación, índice)
Formato	Describe la forma del contenido (texto, multimedia, gráfico, vídeo, audio)
Interacción	Describe el grado de participación que puede tener el alumno con la tarea (baja, media, alta)
Dificultad	De la actividad a realizar (muy fácil, fácil, media, difícil, muy difícil)
Semántica	Complejidad de las instrucciones de la actividad (muy baja, baja, media, alta, muy alta)
Objetivo_bloom	El objetivo que se adquiere al realizar la tarea (conocimiento, comprensión, aplicación, análisis, síntesis, evaluación)
Sección	La unidad didáctica a la que pertenece la tarea
Subsección	Subapartado dentro de la unidad didáctica
Asignatura	Curso en el que está incluida la tarea
Ranking	Valoración de la actividad por parte de todos los alumnos del curso, junto con la frecuencia de visualización o de asignación de la tarea
Ayuda	Enlaces con recursos (dirección URL) que pueden servir de ayuda

**Tabla 8.25:** Descripción de los campos que conforman la tabla Tarea

La entidad Planificación está compuesta por la Ruta de Aprendizaje, como se puede observar en la figura 8.18. La Ruta de Aprendizaje se define por los atributos Secuenciación, que es un listado de todas las actividades o tareas planificadas para la Unidad Didáctica, y del atributo Temporalización, que establece el orden y el tiempo necesario para dedicarle a cada tarea.



**Figura 8.18:** Atributos de la entidad Planificación

A continuación, en la figura 8.19 se muestran las relaciones de las tablas que se acaban de definir y que amplían las relaciones existentes que se mostraban en la figura 8.14. En este punto, el Registro ya dispone de todos los campos que lo definen y, aplicando minería de datos junto con la información que proporciona la base de datos que contiene todos los recursos, listados en la tabla Tarea, se es capaz de determinar la ruta de aprendizaje a seguir, que contendrá una serie de actividades a realizar en una temporalización concreta. El hecho de establecer un tiempo para la realización de la unidad didáctica proporciona información del nivel de trabajo del alumno (si es capaz o no de cumplir plazos de entrega) y permite limitar también el acceso al resto de unidades, si éstas, por ejemplo, aun no se han explicado en las clases de teoría o el profesor por el motivo que sea no considera que el material esté disponible hasta una fecha determinada.

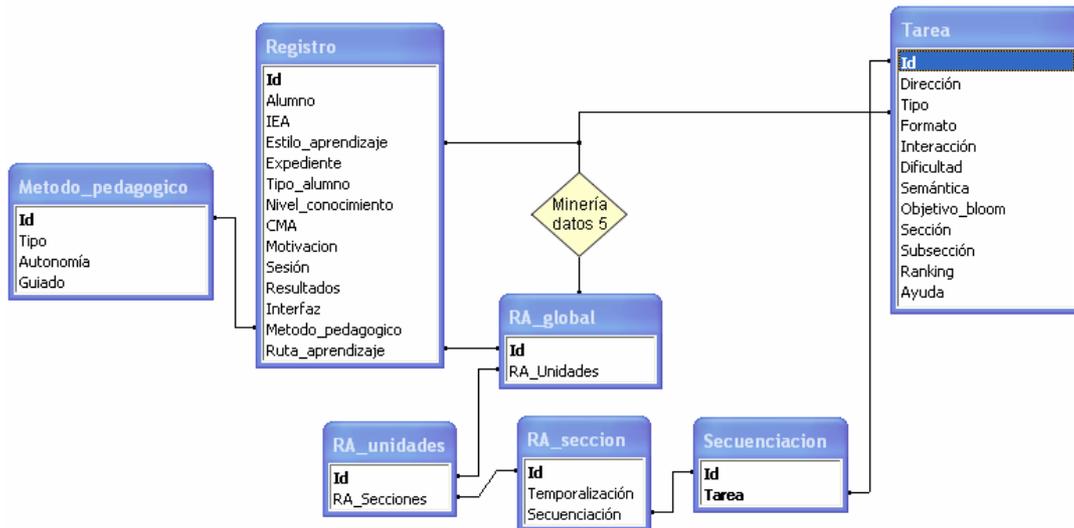


Figura 8.19: Relaciones entre las tablas de la base de datos del Modelo del Alumno y del Tutor

Con el fin de clarificar las ideas expuestas en este capítulo, se integran todos los modelos descritos en un único modelo que constituye el LMS (ver figura 8.20) y que está formado, a su vez, por los modelos del Alumno, Tutor y Dominio, así como de las relaciones y flujos de información entre ellos, que se representan con flechas más gruesas de distinto color, siendo las líneas de color rojo las que sirven para inicializar el sistema LMS y las de color negro las que lo realimentan dinámicamente, actualizándolo en función del proceso de aprendizaje del alumno. Las líneas discontinuas indican la toma de datos iniciales del alumno mediante distintos cuestionarios o bien, examinando el expediente con los cursos previos realizados y las notas obtenidas.

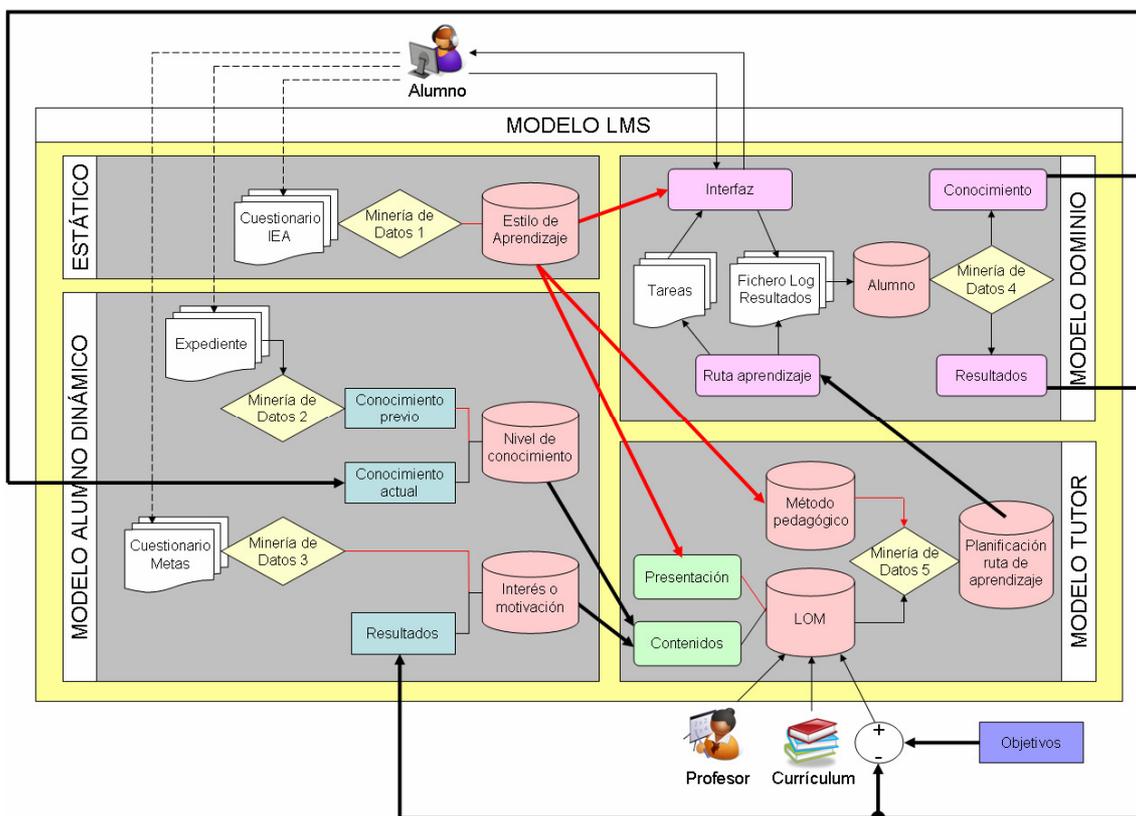


Figura 8.20: Integración de los tres modelos que conforman el LMS

Se ha modificado el modelo LMS añadiendo el bucle de control que incorpora una realimentación de los objetivos, con el fin de homogeneizarlo con el modelo expuesto en el capítulo 2.1 (ver figura 2.1). Las realimentaciones internas se producen a “tiempo real”, es decir, cuando se finaliza una unidad didáctica se toman los valores de los resultados de todos los alumnos para inferir la nueva ruta de aprendizaje de la siguiente unidad didáctica. En cambio, la velocidad de la realimentación exterior es más lenta, revisándose los objetivos y resultados con la finalización del curso (ver figura 8.21).

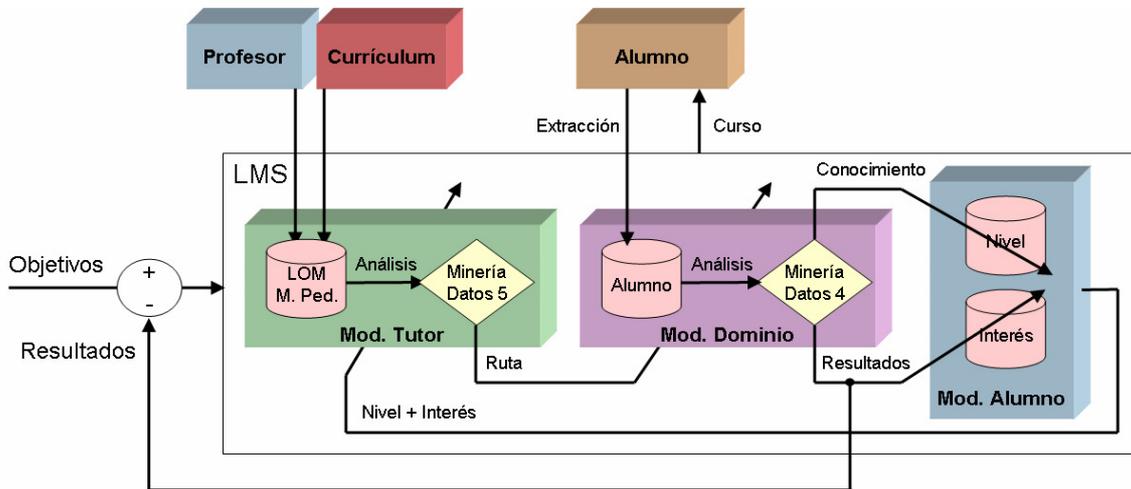


Figura 8.21: Esquema simplificado del bucle de control del LMS

Este modelo se ha presentado en un paper en las VI Jornadas CEA de Enseñanza a través de Internet-Web de la Ingeniería de Sistemas y Automática EIWISA'2010 con el ISBN-13: 978-84-693-1779-2 [90].

El trabajo “*A New Framework for the Control of LMS in ITS*” ha sido aceptado como un poster con el número 18, en la conferencia KMIS 2010 (International Conference on Knowledge Management and Information Sharing) que se celebrará en Valencia del 25 al 28 de octubre.

## 8.2 Bases de Datos

En las siguientes secciones se van a establecer las relaciones existentes entre las tablas que se han definido en el capítulo anterior, indicando cuál es la equivalencia entre los datos de las distintas tablas de la base de datos.

También se definirán los tipos de datos de todos los campos que constituyen las tablas, sobre los que se aplicará posteriormente minería de datos para analizar la información almacenada (ver el capítulo 8.3).

### Tablas: Definición de las interrelaciones

La primera relación con la que nos encontramos en el Modelo del Alumno es el Estilo de Aprendizaje que determina el Método Pedagógico del Modelo Tutor, como se muestra en la figura 8.22.

A partir de las definiciones de los distintos tipos de Métodos Pedagógicos estudiados [89], se han establecido las equivalencias con las diferentes dimensiones de los Estilos

de Aprendizaje, marcando el estilo que predomina en cada método, como se observa en la tabla 8.26.

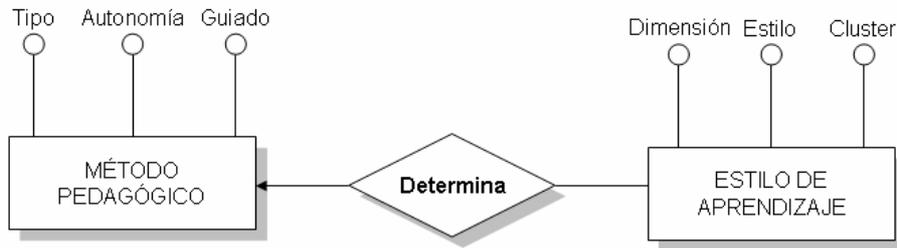


Figura 8.22: Interrelación Determina

			Estilos de Aprendizaje							
			Procesamiento		Percepción		Entrada		Comprensión	
			Activo	Reflexivo	Sensorial	Intuitiva	Visual	Verbal	Secuencial	Global
Autonomía	Guiado	Tipo								
Baja	Profesor	Instruccional		X		X		X	X	
Media		Orientador	X		X		X	X		
		Socrático	X			X		X	X	
Alta	Iguales	Constructivista	X		X			X		X
		Cooperativo	X		X		X	X	X	
		Contextual	X		X		X			X

Tabla 8.26: Relación entre los atributos de la entidad Método Pedagógico y los Estilos de Aprendizaje

Los objetos LOM, que son la base para definir las Tareas que realizará el alumno y que se almacenan en el Modelo Dominio, están constituidos por los atributos Formato, Semántica, Dificultad, Tipo de Recurso e Interactividad.

Estos atributos vienen condicionados por el Estilo de Aprendizaje, están marcados por el Nivel de Conocimiento y se ven favorecidos por la Motivación del alumno, según la figura 8.23.

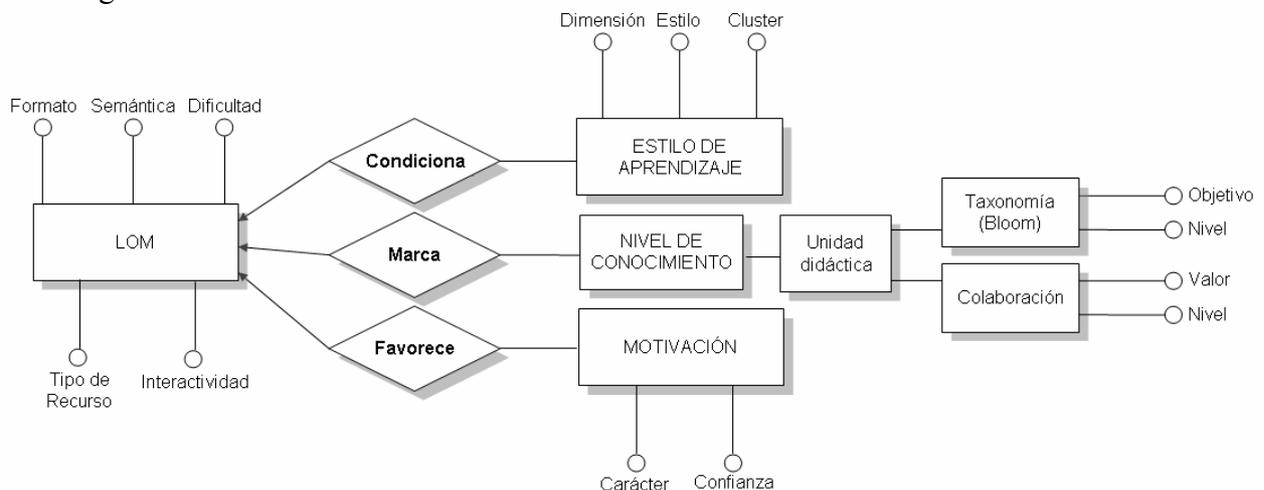


Figura 8.23: Interrelaciones Condiciona, Marca y Favorece

En el estudio realizado por [46] se establece la relación entre los diferentes tipos de atributos del objeto LOM y el Estilo de Aprendizaje, como se observan en las tablas 8.27 donde se indican los Tipos de Recursos y en la tabla 8.28 donde aparecen el resto de atributos (Formato, Densidad de semántica, Tipo de Interactividad).

Tipo de Recurso	Estilos de Aprendizaje							
	Procesamiento		Percepción		Entrada		Comprensión	
	Activo	Reflexivo	Sensorial	Intuitiva	Visual	Verbal	Secuencial	Global
Ejercicio	X			X		X	X	
Simulación	X		X		X			
Cuestionario	X					X	X	
Diagrama				X	X			X
Estadísticas			X		X			X
Gráfico			X		X			X
Índice						X		X
Diapositiva						X	X	
Tabla			X					X
Texto narrativo		X		X		X		
Examen	X							
Experimento	X		X					
Problema	X		X			X		
Auto evaluación	X						X	
Conferencia		X		X		X		

**Tabla 8.27:** Relación entre el Tipo de Recurso (LOM) y los Estilos de Aprendizaje

Formato	Estilos de Aprendizaje							
	Procesamiento		Percepción		Entrada		Comprensión	
	Activo	Reflexivo	Sensorial	Intuitiva	Visual	Verbal	Secuencial	Global
Texto		X		X		X	X	
Multimedia			X		X			
Gráfico			X		X			X
Vídeo			X		X			
Audio			X			X	X	
Densidad de Semántica								
Muy baja				X				
Baja				X				
Media			X					
Alta			X					
Muy alta			X					
Tipo de Interactividad								
Activo	X							
Expuesto		X						
Mixto	X	X						

**Tabla 8.28:** Relación entre el resto de atributos LOM y los Estilos de Aprendizaje

Por otro lado el Nivel del Objetivo de la taxonomía de Bloom alcanzado por el alumno en la materia marca la Dificultad de las tareas que se le encomendarán, según la tabla 8.29.

Nivel Objetivo Bloom						
Dificultad	Principiante		Medio		Experto	
	Conocimiento	Comprensión	Aplicación	Análisis	Síntesis	Evaluación
Muy fácil	X					
Fácil		X				
Media			X	X		
Difícil					X	
Muy difícil						X

**Tabla 8.29:** Relación entre la Dificultad (LOM) y los Objetivos Bloom (Nivel de Conocimiento)

Nivel Objetivo Bloom						
Tipo de Recurso	Principiante		Medio		Experto	
	Conocimiento	Comprensión	Aplicación	Análisis	Síntesis	Evaluación
Ejercicio			X	X	X	
Simulación			X	X	X	X
Cuestionario		X	X	X	X	X
Diagrama				X	X	X
Estadísticas				X	X	X
Gráfico				X	X	X
Índice	X	X				
Diapositiva	X	X				
Tabla		X	X	X	X	X
Texto narrativo	X	X				
Examen						X
Experimento			X	X	X	X
Problema			X	X	X	
Auto evaluación						X
Conferencia	X	X				

**Tabla 8.30:** Relación entre el Recurso (LOM) y los Objetivos Bloom (Nivel de Conocimiento)

Del mismo modo, el Tipo de Recurso también viene marcado por el Nivel del Objetivo Bloom, dado el carácter intrínseco de cada tipo de tarea y el propósito con el que se realiza o se propone al alumno. Por ejemplo, una práctica de simulación sirve para usar el material aprendido en situaciones nuevas y específicas (un laboratorio

virtual o remoto), es decir, se trata del Objetivo de Bloom conocido como Aplicación, aunque también se puede emplear en el Análisis, si se trata de una práctica más compleja que requiera una división del problema en diferentes partes, o bien, como Síntesis, si la misión del alumno es la creación de una práctica nueva a partir de la unión de conocimientos previos, o también como Evaluación, para calificar al alumno.

Al igual que en el ejemplo de la práctica de simulación, en la tabla 8.30 se han estudiado el resto de Tipos de Recursos estableciendo qué Objetivo de Bloom se puede alcanzar al realizarlos.

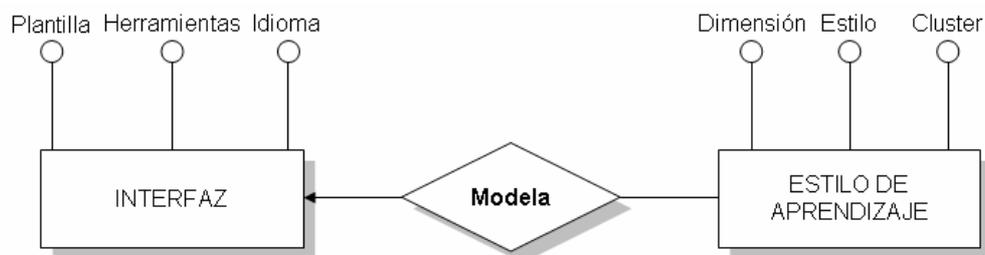
La Motivación del alumno favorece o inhibe su proceso de aprendizaje, por lo que se debe adaptar la Dificultad de los contenidos que se ofrecen al alumno en función de su orientación motivacional. Siguiendo las pautas indicadas en [84] se ha confeccionado la tabla 8.31, en dónde si el alumno presenta una tendencia motivadora dominante (según los valores establecidos en la tabla 8.15) y una confianza concreta, se debe ajustar la dificultad acorde a estos parámetros.

Dificultad	Motivación			
	Intrínseca		Extrínseca	
	Baja	Alta	Baja	Alta
Muy fácil			X	
Fácil			X	
Media	X		X	
Difícil	X	X		X
Muy difícil		X		X

**Tabla 8.31:** Relación entre la Dificultad (LOM) y la Motivación

Por último, la interfaz viene modelada por el Estilo de Aprendizaje (ver figura 8.24), aunque esta relación no es muy vinculante, ya que el alumno tiene libertad para modificar y adaptar el entorno de trabajo o interfaz a su gusto, guardándose los cambios que realice y manteniéndolos para las siguientes sesiones de trabajo que establezca.

A modo de inicialización de la aplicación LMS se pueden tomar en consideración las pautas establecidas en la tabla 8.32 dónde se han analizado las características que presentan los diferentes tipos de herramientas y su relación con las dimensiones del Estilo de Aprendizaje.



**Figura 8.24:** Interrelación Modela

Herramientas		Estilos de Aprendizaje			
		Procesamiento		Entrada	
		Activo	Reflexivo	Visual	Verbal
Interacción	Foro	X			X
	Chat	X			X
	Diario (blog)	X			X
	Wiki		X		X
	Votación	X			X
Utilidades	Agenda (calendario)		X		X
	Procesador de texto	X			X
	Presentación	X		X	
	Multimedia		X	X	
	Video		X	X	
	Audio		X		X
	Noticias (feeds)		X		X
	Aplicación Web (simulación)	X		X	

**Tabla 8.32:** Relación entre las Herramientas (Interfaz) y los Estilos de Aprendizaje

### Tablas: Definición de campos

En la figura 8.8 se nombran los campos de las distintas tablas que conforman el modelo del Alumno, a continuación, en la tabla 8.33, se definen los tipos de datos de estos campos. La tabla Registro, almacena toda la información del proceso de enseñanza/aprendizaje del alumno y está, por tanto, relacionada con el resto de tablas y aparece en todos los modelos que constituyen el sistema LMS, sin embargo, para no repetir la información, se ha definido únicamente en la tabla 8.33.

IEA		Registro	
Id	Autonumérico	Id	Autonumérico
Activo	Número	Alumno	Texto
Reflexivo	Número	IEA	Número
Sensorial	Número	Estilo_aprendizaje	Número
Intuitiva	Número	Expediente	Número
Visual	Número	Tipo_alumno	Número
Verbal	Número	Nivel_conocimiento	Número
Secuencial	Número	CMA	Número
Global	Número	Motivación	Número
		Sesión	Número
		Resultados	Número
		Interfaz	Número
		Método_pedagógico	Número
		Ruta_aprendizaje	Número
Estilo_aprendizaje			
Id	Autonumérico		
Dimensión	Texto		
Estilo	Texto		
Cluster	Texto		

Objetivo Bloom	
Id	Autonumérico
Objetivo	Texto
Nivel	Texto

Expediente	
Id	Autonumérico
Asignatura	Número
Nota	Texto
Ponderación	Número

Colaboración	
Id	Autonumérico
Valor	Texto
Nivel	Texto

Motivación	
Id	Autonumérico
Carácter	Texto
Nivel	Texto
Confianza	Texto

Nivel Conocimiento	
Id	Autonumérico
Unidad	Número
Sección	Número
Tiempo	Fecha/Hora
Nivel_Bloom	Número
Nivel_Colaboración	Número

CMA	
Id	Autonumérico
Ed	Número
Rs	Número
In	Número

Tipo alumno	
Id	Autonumérico
Tipo_alumno	Texto

**Tabla 8.33:** Definición de los tipos de datos de los campos de las tablas del Modelo del Alumno

En la figura 8.14 se nombran los campos de las tablas que conforman el modelo del Dominio, cuya definición de los tipos de datos se pueden consultar en la tabla 8.34. Dentro de la tabla Tarea, el tipo de datos del campo Ranking es Numérico, en el cual, el alumno expresa su grado de acuerdo o desacuerdo con la actividad propuesta, indicando si le ha gustado, con una escala del 1 a 5<sup>14</sup>, la escala de Likert, dado que está demostrado que esta distribución numérica satisface leyes estadísticas que se pueden aplicar posteriormente para su análisis con minería de datos.

Plantilla	
Id	Autonumérico
Fondo	Objeto OLE
Estilo	Objeto OLE
Forma	Texto
Menús	Objeto OLE

Interfaz	
Id	Autonumérico
Plantilla	Número
Herramientas	Número
Idioma	Texto

Actividad	
Id	Autonumérico
Tarea	Número
Acción	Texto
Opciones	Texto
Interacción_alumno	Texto
Tiempo_inicio	Fecha/Hora
Tiempo_fin	Fecha/Hora
Votación	Número
Nota	Texto
Aciertos	Número
Petición_ayuda	Hipervínculo

Herramientas	
Id	Autonumérico
Interacción	Objeto OLE
Utilidades	Objeto OLE

Sesión	
Id	Autonumérico
Tiempo_inicio	Fecha/Hora
Tiempo_fin	Fecha/Hora
Actividad	Texto

Resultados	
Id	Autonumérico
Unidad	Texto
Sección	Texto
Rendimiento	Texto
Nota	Texto
Confianza	Texto

Tarea	
Id	Autonumérico
Dirección	Hipervínculo
Tipo	Texto
Formato	Texto
Interacción	Texto
Dificultad	Texto
Semántica	Texto
Objetivo_bloom	Texto
Sección	Texto
Subsección	Texto
Ranking	Número
Ayuda	Hipervínculo

Competencia actual	
Id	Autonumérico
Competencia	Número
Nivel	Número

**Tabla 8.34:** Definición de los tipos de datos de los campos de las tablas del Modelo del Dominio

En la figura 8.19 se nombran los campos de las distintas tablas que conforman el modelo del Tutor, a continuación, en la tabla 8.35, se definen los tipos de datos de estos campos.

Método pedagógico	
Id	Autonumérico
Tipo	Texto
Autonomía	Texto
Guiado	Texto

Ruta aprendizaje	
Id	Autonumérico
Unidad	Texto
Sección	Texto
Secuenciación	Número
Temporalización	Fecha/Hora

Secuenciación	
Id	Autonumérico
Tarea	Número

**Tabla 8.35:** Definición de los tipos de datos de los campos de las tablas del Modelo del Tutor

### 8.3 Minería de datos

En este capítulo se analiza qué paradigma de computación (o la combinación de varios de ellos) es el más adecuado para los cinco casos de minería de datos que se han planteado en el modelo del LMS en los que se debe emplear (ver figura 8.20).

Existe una diferencia entre los tres primeros casos (ver figura 8.8), en los cuales se debe extraer información de cuestionarios o de expedientes académicos y sólo es necesario realizar esta operación una vez, cuando el usuario se registre en el sistema, como inicialización de la aplicación. Hay que destacar que el número de datos que se manejan en estos tres casos es muy inferior al del resto, por lo que se puede establecer directamente qué paradigma es el que mejor se ajusta al problema.

En cambio, para los dos casos últimos de minería de datos: modelo del Dominio (ver figura 8.14) y modelo del Tutor (ver figura 8.19) se ha procedido a aplicar la teoría de decisión multicriterio discreta (DMD) [81] con el fin de seleccionar el paradigma que mejor cumpla los criterios previstos. Hay que subrayar que en cada uno de estos casos se trata de realizar dos objetivos o tareas de minería de datos distintas: descriptiva y predictiva, respectivamente.

A modo de guía para la selección del método más adecuado, en [91] se establece una correspondencia entre los objetivos o tareas de la minería de datos más destacados y algunas de las técnicas o algoritmos que pueden abordarlas. La clasificación formulada en [91] se ha adecuado a este proyecto teniendo en cuenta la información extraída de los artículos consultados en el estado del arte del capítulo 3.3, resultando la tabla 8.36 que indica de forma orientativa las correspondencias entre tareas y algoritmos observadas en el campo de aplicación de la minería de datos para los ITS.

Bibliografía	Nombre	Predictivo		Descriptivo		
		Clasificación	Regresión	Agrupamiento	Secuencia de patrones y reglas de asociación	Valores anómalos
[9], [10], [11], [12], [29]	Redes de Bayes	✓	✓	✓	✓	✓
[13], [14], [15], [16]	Árboles de decisión	✓	✓			
[17], [18], [19], [20], [32]	Razonamiento basado en casos	✓			✓	
[21], [22]	Redes neuronales	✓	✓			
[19], [23], [24], [25]	Algoritmos genéticos	✓	✓	✓	✓	
[13], [20], [22], [26], [27], [28], [29], [39], [43]	Lógica borrosa	✓	✓	✓	✓	
[20], [22], [30], [31], [32], [33], [41]	Sistema multiagentes	✓	✓	✓	✓	✓
[33], [34], [35], [42]	Patrones secuenciales				✓	
[33], [36], [37], [38], [39], [40]	Reglas de asociación		✓		✓	
[41], [42], [43]	Agrupamiento	✓	✓	✓	✓	✓

**Tabla 8.36:** Correspondencia entre las tareas de la minería de datos y los paradigmas de computación

Como se puede observar de la tabla 8.36, la correspondencia entre tareas y técnicas es muy variada. Esta variedad es una de las razones por la que es necesario conocer las capacidades de cada técnica, los ámbitos donde suele funcionar mejor, la eficiencia, la robustez, etc., en definitiva, las características funcionales de cada técnica respecto a las demás.

Es por ello que, para cada una de las tareas de la minería de datos (descriptiva y predictiva) y sus correspondientes objetivos de la minería de datos (agrupamiento, clasificación, asociación y análisis de secuencias), se ha creado una matriz de decisión con los criterios o información que se desea obtener del proceso de interacción y aprendizaje del alumno en el sistema LMS.

## Minería de datos 1 (Modelo Alumno Estático)

En el primer caso de estudio (ver figura 8.25) se emplea la minería de datos para extraer la información del cuestionario IEA (Anexo 1) clasificándola, según la tabla 8.27, en los parámetros que establecen el Estilo de Aprendizaje del Alumno (definidos en la tabla 8.9).

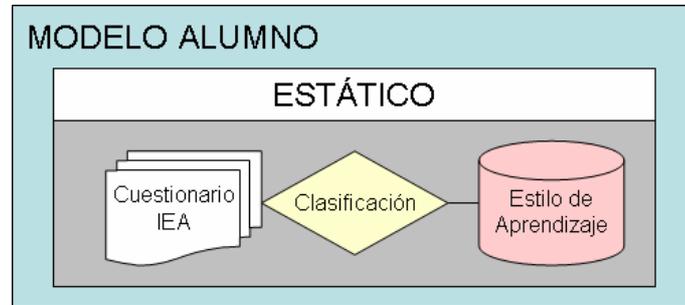


Figura 8.25: Diagrama del proceso de Minería de datos 1

La tabla 8.37 se ha confeccionado comparando las cuestiones del cuestionario con las definiciones de las dimensiones del Estilo de Aprendizaje [45], resultando 11 preguntas para cada dimensión (procesamiento, percepción, entrada, comprensión). Las dos posibles respuestas (a y b) de cada pregunta, determinan el valor de cada dimensión (activo/reflexivo, sensorial/intuitiva, visual/verbal, secuencial/global).

Dado que los datos de entrada se pueden transformar en valores numéricos (las respuestas *a* son 1 y las *b* son 0) y tienen un valor de salida conocido, determinado por la tabla 8.31, el método de minería de datos más simple para resolver el problema es un **clasificador basado en reglas** del tipo “SI (antecedente) ENTONCES (consecuente)”.

Únicamente se debe sumar y almacenar el valor de las respuestas correspondientes a cada dimensión, siendo el resultado el que defina la pertenencia a una clase (fuerte, moderado, equilibrado) como se observa en la figura 8.26. Este diagrama se puede traducir fácilmente en reglas, por ej. “SI activo/reflexivo = -1 o activo/reflexivo = -3 ENTONCES procesamiento = EquilibradoReflexivo”

	FUERTE	MOD.	EQUILIBRADO	MOD.	FUERTE	
ACT	11 9	7 5	3 1   -1 -3	-5 -7	-9 -11	REF
SEN	11 9	7 5	3 1   -1 -3	-5 -7	-9 -11	INT
VIS	11 9	7 5	3 1   -1 -3	-5 -7	-9 -11	VRB
SEQ	11 9	7 5	3 1   -1 -3	-5 -7	-9 -11	GLO

Figura 8.26: Clases para los Estilos de Aprendizaje de Felder

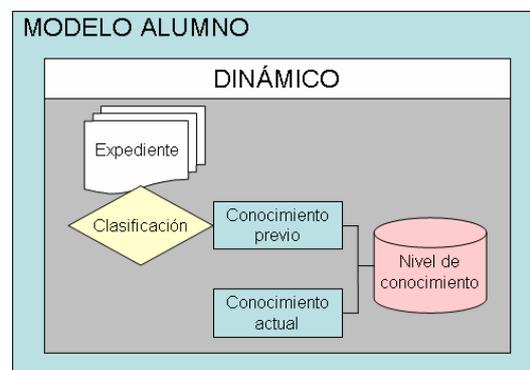
Estilos de Aprendizaje							
Procesamiento		Percepción		Entrada		Comprensión	
Activo	Reflexivo	Sensorial	Intuitiva	Visual	Verbal	Secuencial	Global
1a	1b	2a	2b	3a	3b	4a	4b
5a	5b	6a	6b	7a	7b	8a	8b
9a	9b	10a	10b	11a	11b	12a	12b
13a	13b	14a	14b	15a	15b	16a	16b
17a	17b	18a	18b	19a	19b	20a	20b
21a	21b	22a	22b	23a	23b	24a	24b
25a	25b	26a	26b	27a	27b	28a	28b
29a	29b	30a	30b	31a	31b	32a	32b
33a	33b	34a	34b	35a	35b	36a	36b
37a	37b	38a	38b	39a	39b	40a	40b
41a	41b	42a	42b	43a	43b	44a	44b

**Tabla 8.37:** Relación entre las Herramientas (Interfaz) y los Estilos de Aprendizaje

## Minería de datos 2 (Modelo Alumno Dinámico)

En este segundo caso (ver figura 8.27) se emplea la minería de datos para extraer la información del expediente del alumno y poder inferir qué tipo de alumno es: regular, normal, bueno o excelente, si la calificación es inferior a 5, está comprendida entre 5 y 7, entre 7 y 9 o es superior a 9, respectivamente.

Este análisis tiene como intención la de plantear la primera ruta de aprendizaje de la primera unidad didáctica del curso, con una dificultad adaptada al tipo de alumno. Para ello, se tiene en cuenta las calificaciones obtenidas en las asignaturas previas relacionadas con el curso que se impartirá en el sistema LMS, que pueden ser cuantitativas (de 0 a 10) o bien cualitativas (suspense, aprobado, notable, sobresaliente, matrícula de honor), en cuyo caso se definirán con los siguientes rangos: <5, 5-7, 7-9, 9-10 y 10, respectivamente. Estas calificaciones vienen ponderadas con un factor que determinará el profesor del curso (de 0 a 1), en función de la importancia que le dé a cada asignatura. Si existe más de una asignatura relacionada con el curso, se sacará el promedio de las calificaciones por su ponderación.



**Figura 8.27:** Diagrama del proceso de Minería de datos 2

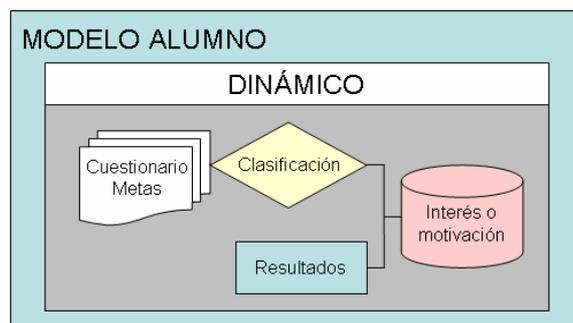
Mediante la aplicación de la **lógica borrosa** a la calificación ponderada, se puede clasificar al alumno (regular, normal, bueno) y con ello inferir un grado de dificultad para la unidad didáctica que cursará (muy fácil, fácil, media, difícil, muy difícil).

En las siguientes iteraciones (cada vez que el alumno realice una unidad didáctica o la repita si los resultados no son los deseados), se tomarán los datos nuevos de interacción del alumno con el sistema, los resultados obtenidos y las dudas planteadas o las dificultades observadas, para actualizar el nivel de dificultad de la ruta de aprendizaje.

Se comparará el nivel de dificultad asignado en la segunda iteración con el inicial, es decir, el éxito de la predicción, con el fin de actualizar los parámetros que definen el sistema borroso (la función de pertenencia) y ajustarlo para minimizar el error producido.

### Minería de datos 3 (Modelo Alumno Dinámico)

En el tercer caso de minería de datos (ver figura 8.28) se clasifica el interés o motivación del alumno inicial a partir de la realización de un cuestionario denominado CMA (ver el anexo 2). Éste aporta un valor para los factores Educativo y Reconocimiento Social que determinan el carácter motivacional del alumno, siempre que el resultado obtenido tenga el nivel dominante, es decir, que supere unos umbrales definidos en la tabla 8.15.



**Figura 8.28:** Diagrama del proceso de Minería de datos 3

Los umbrales definidos ya nos indican la idoneidad de las funciones de pertenencia para determinar el Nivel del Carácter motivacional del alumno, por tanto se aplicará la **lógica borrosa** en los dos factores: Educativo y Reconocimiento Social para ver cuál predomina o es más relevante.

En caso de salir como carácter el Extrínseco, haría falta determinar la Confianza del alumno en sus estudios con el fin de poder inferir un Nivel de Dificultad en la planificación de la unidad y de las tareas (según lo que indica la tabla 8.13).

Se han propuesto dos vías para fijar el nivel de dificultad de la unidad didáctica, por un lado a partir del expediente del alumno (minería de datos 2) y por otro a partir de la motivación (minería de datos 3). Los resultados deben coincidir pero, podría darse el caso de que fueran distintos. En esta situación se optará por seleccionar el peor de los casos, informando al alumno sobre la propuesta de ruta sugerida en función de sus condiciones pero dándole la opción de cambiarla si no se siente satisfecho. Por ejemplo, la combinación de un alumno con una trayectoria académica floja pero con alta autoestima y una meta del rendimiento daría un nivel de dificultad bajo para la unidad didáctica primera, lo que implica la obtención de una calificación proporcional a la

dificultad planteada (es decir, inferior al 100% de la nota que se obtendría con una ruta de aprendizaje más complicada). La proporcionalidad la implementará el profesor responsable de la asignatura teniendo en cuenta en que niveles mínimos de objetivos y contenidos se está trabajando en cada unidad.

Por otro lado, el factor Interpersonal se extrae también del cuestionario CMA (ver anexo 3 con los ítems que están directamente relacionados), definiéndose mediante unos umbrales que indican el Valor de colaboración del alumno, como indica la tabla 8.16. Aplicando el mismo criterio de relatividad que ofrecen los umbrales, el método más adecuado para analizar los datos es la **lógica borrosa**.

## Minería de datos 4 (Modelo Dominio)

De los objetivos o tareas de la minería de datos descritos en el capítulo 3.2, interesa para esta aplicación de análisis de los datos de interacción del alumno y de su proceso de enseñanza y aprendizaje, las tareas que sean descriptivas, cuyo objetivo no es predecir nuevos datos sino describir los existentes. Lo que se busca es describir “cómo es y que ha hecho el alumno”. En [91] se citan qué técnicas de la minería de datos forman parte de los objetivos o tareas descriptivas:

- **Agrupamiento:** el objetivo de esta tarea es obtener grupos o conjuntos entre los elementos de  $\delta$ , de tal manera que los elementos asignados al mismo grupo sean *similares*. La diferencia respecto a la clasificación es que son precisamente los grupos y la pertenencia a los grupos lo que se quiere determinar y, a priori, no se sabe ni cómo son los grupos ni cuantos hay.
- **Reglas de asociación:** el objetivo es similar a los estudios correlacionales y factoriales pero para atributos nominales. Dado los ejemplos del conjunto  $E = A_1x_1A_2x_2...x_nA_n$ , una regla de asociación se define generalmente de la siguiente forma: “SI  $A_i=a$  Y  $A_j=b$  Y ... Y  $A_k=h$  ENTONCES  $A_r=u$  Y  $A_s=v$  Y ... Y  $A_z=w$ ”
- **Detección de valores e instancias anómalas:** la detección de valores anómalos o atípicos (*outlier detection*) tienen el objetivo de realizar limpieza de datos aunque también puede ser muy útil para detectar comportamientos anómalos, que pueden sugerir fraudes, gallos, intrusos o comportamientos diferenciados. La instancia anómala es más general en el sentido que no sólo considera un único atributo, sino que los considera todos. La manera de abordar el problema es generalmente la de agrupar los ejemplos y ver aquellas instancias que se quedan “desplazadas” de los grupos mayoritarios.

De las técnicas que se acaban de enumerar, se ha encontrado información relativa a las mismas en la bibliografía consultada en el capítulo 3.3 y que se refleja a modo resumen en la tabla 8.30. Estos datos se han ampliado teniendo en cuenta una serie de criterios para poder aplicar la teoría de decisión multicriterio discreta (DMD) y se han clasificado en tres tablas (ver tablas 8.38, 8.39, y 8.40) una por cada tarea descriptiva. En las alternativas de las tablas se ha empleado una nomenclatura para designar los diferentes métodos de minería de datos:

a	Redes de Bayes	f	Lógica borrosa
b	Árboles de decisión	g	Sistema multiagentes
c	Razonamiento basado en casos	h	Patrones secuenciales
d	Redes neuronales	i	Reglas de asociación
e	Algoritmos genéticos	j	Agrupamiento

**Agrupamiento**

	$w_j$	Alternativas									
		a	b	c	d	e	f	g	h	i	j
Agrupar alumnos por características de aprendizaje (actividades realizadas)	6	X			X	X	X	X		X	X
Agrupar alumnos por sus habilidades (notas obtenidas)	5	X			X		X	X			X
Agrupar alumnos por el tiempo de navegación	2	X			X		X	X	X	X	X
Caracterizar grupos con comportamientos similares	4	X			X		X	X	X		X
Promover aprendizaje colaborativo basado en grupos (muy activos, activos o inactivos)	3	X			X		X		X		X
Agrupar tests y cuestiones en función de la puntuación	1	X									
$r_j$		21	0	0	20	6	20	17	9	8	20

**Tabla 8.38:** Matriz de decisión para valorar el Agrupamiento**Reglas de asociación y patrones de secuencias**

	$w_j$	Alternativas									
		a	b	c	d	e	f	g	h	i	j
Guiar las actividades automáticamente	10			X		X	X	X	X	X	
Recomendar materiales de aprendizaje	9			X		X	X	X	X	X	
Determinar qué material es el más adecuado	11			X		X	X	X	X	X	X
Ofrecer consejo a los alumnos	8	X		X		X	X	X	X		X
Descubrir patrones de comportamiento de los alumnos (estilos de aprendizaje)	4	X			X	X	X	X	X	X	X
Optimizar el contenido del LMS determinado por los intereses de los alumnos	7		X			X			X	X	X
Evaluar las actividades de los alumnos	2			X			X		X		X
Personalizar itinerarios basados en Objetos de Aprendizaje (LO)	6	X								X	
Generar actividades personalizadas	5					X		X			
Descubrir y comparar patrones de comportamiento con un itinerario de aprendizaje predefinido	3									X	
Definir patrones que son marcadores de éxito	1									X	
$r_j$		18	7	40	4	54	51	47	51	51	32

**Tabla 8.39:** Matriz de decisión para valorar las Reglas de asociación y los patrones de secuencias**Detección de valores anómalos**

	$w_j$	Alternativas									
		a	b	c	d	e	f	g	h	i	j
Diagnosticar problemas de aprendizaje	3	X					X		X		X
Determinar actitudes positivas y negativas	1	X									X
Detectar los conceptos más difíciles	2	X					X		X		X
$r_j$		6	0	0	0	0	5	0	5	0	6

**Tabla 8.40:** Matriz de decisión para valorar la Detección de valores anómalos

A continuación se ha seleccionado la ponderación lineal [81] como método de DMD. Su esquema básico consiste en construir una función de valor  $U(A_i)$  para cada alternativa propuesta en las tablas 8.38, 8.39 y 8.40:

$$U(A_i) = \sum_j w_j r_{ij}$$

Siendo  $w_j$  el peso del criterio  $j$  y  $r_{ij}$  la evaluación (rating) de la alternativa  $i$  respecto al criterio  $j$ .

El resultado de la aplicación del método de DMD se muestra en la tabla 8.41, donde se observan los tres paradigmas de computación que obtienen mayor puntuación, y por tanto, se adecuan mejor al problema planteado de minería de datos son la Lógica borrosa, los Sistemas multiagentes y los Patrones secuenciales.

	$U(A_j)$	
Alternativas	Redes de Bayes	45
	Árboles de decisión	7
	Razonamiento basado en casos	40
	Redes neuronales	24
	Algoritmos genéticos	60
	Lógica borrosa	<b>76</b>
	Sistema multiagentes	<b>64</b>
	Patrones secuenciales	<b>65</b>
	Reglas de asociación	59
	Agrupamiento	58

**Tabla 8.41:** Resultados de la Ponderación Lineal para las distintas Alternativas

Este resultado se obtiene a partir del estado del arte de la minería de datos aplicada a los ITS, por tanto es un indicativo de la tendencia actual y de los algoritmos y las aplicaciones que más se están utilizando, pero que no implican que sean las que mejor resuelven el problema en su conjunto ni que no existan otros algoritmos igual de válidos u óptimos para plantear el problema de los ITS.

## Minería de datos 5 (Modelo Tutor)

El objetivo o la tarea de la minería de datos que interesa para este caso es la predicción, ya que la meta que se pretende alcanzar con el sistema ITS es proponer una ruta de aprendizaje adaptada al proceso de aprendizaje del alumno que prediga o garantice buenos resultados académicos y la máxima adquisición de conocimientos en función de cómo es el alumno, sus necesidades, intereses, conocimientos actuales... es decir, datos que se han analizado en el Modelo del Dominio mediante la tarea descriptiva de minería de datos y que realimentan al Modelo Tutor.

En [91] se describen que técnicas de la minería de datos forman parte de los objetivos o tareas predictivas:

- **Clasificación:** el objetivo es aprender una función  $\lambda: E \rightarrow S$ , denominada clasificador, que represente la correspondencia existente en los ejemplos, conjuntos de pares de elementos de dos conjuntos,  $\mathcal{D} = \{ \langle e, s \rangle : e \in E, s \in S \}$  donde  $S$  es el conjunto de valores de salida. Es decir, para cada valor de  $E$  tenemos un único valor para  $S$ . Además  $S$  es nominal, puede tomar un conjunto de valores  $c_1, c_2, \dots, c_m$ , denominados clases.
- **Regresión:** el conjunto de evidencias son correspondencias entre dos conjuntos  $\mathcal{D}: E \rightarrow S$ , donde  $S$  es el conjunto de valores de salida. El objetivo es aprender una función  $\lambda: E \rightarrow S$  que represente la correspondencia existente en los ejemplos. La diferencia respecto a la clasificación es que  $S$  es numérico. También se conoce como interpolación (cuando el valor predicho está en medio de otros) o estimación (cuando se trata de algo futuro).

Al igual que en el apartado anterior, se han extraído qué información de la bibliografía consultada en el capítulo 3.3 empleaba estas técnicas de minería de datos. Estos datos se han clasificado en una tabla que considera las técnicas de Clasificación (ver tabla 8.42). La regresión se considera incluida dentro de la clasificación, por lo que no se ha añadido una tabla con estos datos.

### Clasificación

	$w_j$	Alternativas									
		a	b	c	d	e	f	g	h	i	j
Descubrir grupos de estudiantes con características similares y reacciones a una estrategia pedagógica	5	X	X	X	X	X	X	X	X		X
Predecir la actuación de los estudiantes	7	X	X		X	X	X		X		
Predecir la nota final de los estudiantes	6		X			X	X				
Detectar un mal uso o distracción de los estudiantes	1	X	X			X			X		X
Identificar alumnos con poca motivación	2	X	X				X	X	X		X
Promover actividades motivadoras o correctoras	4		X			X	X		X	X	
Encontrar las malas concepciones comunes a los estudiantes	3	X	X								X
	$r_j$	18	28	5	12	23	24	7	19	4	11

**Tabla 8.42:** Matriz de decisión para valorar el Clasificación

A continuación se ha seleccionado la ponderación lineal [81] como método de DMD. Su esquema básico consiste en construir una función de valor  $U(A_i)$  para cada alternativa:

$$U(A_i) = \sum_j w_j r_{ij}$$

Siendo  $w_j$  el peso del criterio  $j$  y  $r_{ij}$  la evaluación (*rating*) de la alternativa  $i$  respecto al criterio  $j$ .

El resultado de la aplicación del método de DMD se muestra en la tabla 8.43, donde se observan los dos paradigmas de computación que obtienen mayor puntuación, y por tanto, se adecuan mejor al problema planteado de minería de datos son los Árboles de decisión y la Lógica borrosa.

		$U(A_j)$
Alternativas	Redes de Bayes	18
	<b>Árboles de decisión</b>	<b>28</b>
	Razonamiento basado en casos	5
	Redes neuronales	12
	Algoritmos genéticos	23
	<b>Lógica borrosa</b>	<b>24</b>
	Sistema multiagentes	7
	Patrones secuenciales	19
	Reglas de asociación	4
	Agrupamiento	11

**Tabla 8.43:** Resultados de la Ponderación Lineal para las distintas Alternativas

Al igual que se ha comentado en el apartado anterior, los algoritmos que resultan del estudio de adecuación para el Modelo del Tutor son consecuencia del estado del arte de la aplicación de la minería de datos a los ITS, y por tanto únicamente reflejan qué algoritmos son los que mejor se adaptan dentro de los más empleados actualmente.

## 8.4 Sistema de control: objetivos/resultados

La forma de evaluar el buen funcionamiento del sistema de control propuesto para el LMS y la satisfacción con el mismo consiste en establecer una comparación con los **resultados** (calificación media de los estudiantes) obtenidos al aplicar el curso LMS y los valores almacenados del curso anterior, en el cual no existía el programa LMS.

Por otro lado, también es indicativo del buen funcionamiento una reducción en la **tasa de abandono** de la asignatura, entendiéndose como la relación porcentual entre el número total de estudiantes de una cohorte de nuevo ingreso que, estando matriculados, no se presentaron (o suspendieron) la asignatura y no se han vuelto a matricular en los dos cursos siguientes. Para el título de Ingeniería Industrial el valor de la tasa de abandono es del **31.4%** para el año **2002**, estimándose un valor para el año de implantación de los nuevos grados del **35%**.

Por último, la Universidad establece una **tasa de eficiencia** como la relación porcentual entre el número total de créditos teóricos del plan de estudios del título y el número total de créditos en los que realmente se ha matriculado el alumno. En el año **2000** este valor era del **70%** (lo que implica que los alumnos deben matricularse en un 30% más de créditos que lo estipulado), estimándose una tasa de eficiencia con los nuevos grados del **74%**. Esta tasa se puede trasladar al curso ITS, adaptando el número de créditos al número de veces que se matricula el alumno hasta que consigue aprobar la asignatura.

## 8.5 Conclusiones

En este capítulo se ha establecido la metodología para la implementación del ITS adaptativo, seleccionado los componentes necesarios que lo integran, como es el caso de la plataforma o LMS y se han definido, mediante bases de datos, los parámetros y las interrelaciones entre los modelos alumno, tutor y dominio del ITS.

Se ha analizado en profundidad las técnicas de minería de datos empleadas en la literatura, clasificándolas en función de los objetivos o tareas de minería de datos que se quieren obtener en el ITS: descriptiva y predictiva. Los resultados obtenidos reflejan las tendencias actuales utilizadas en la minería de datos y sirven como guía para la determinación de los algoritmos más adecuados para el sistema ITS adaptativo. Comentar que de las cinco fases en las que se aplica la minería de datos (figura 8.20) para la extracción de conocimiento, hay que diferenciar una primera etapa de inicialización del sistema, en la que se aplican las tres primeras (realmente se combinan las tres para determinar el agrupamiento de los alumnos), y una segunda etapa, durante el transcurso del curso, en la que se aplican las dos últimas, consistentes en la determinación de la ruta de aprendizaje (minería de datos 5) y un reagrupamiento de los alumnos con la incorporación de los datos obtenidos de su interacción con el LMS (minería de datos 4). En el siguiente capítulo se indican los algoritmos que se emplearán en el sistema ITS adaptativo.

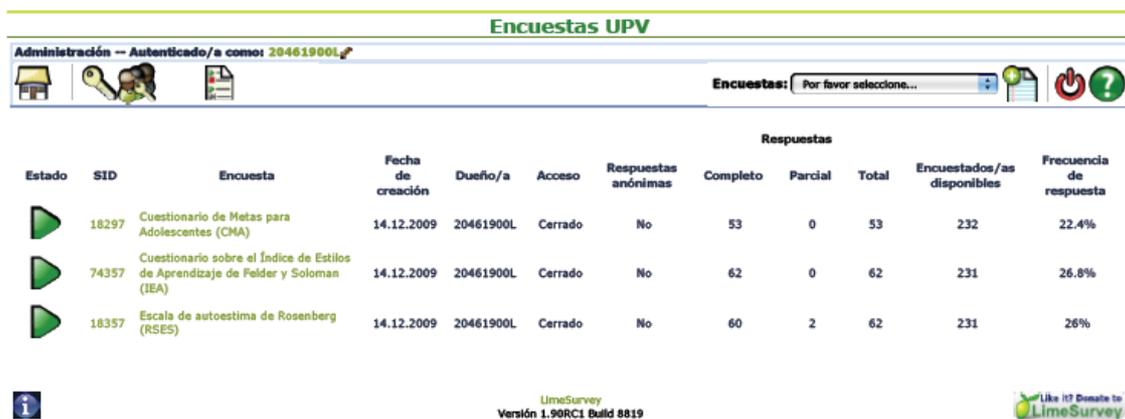


## Capítulo 9

# Aplicación del Sistema de Control

En este capítulo se ha implementado el sistema ITS adaptativo para un caso real: los alumnos de la asignatura de Automatización Industrial de la carrera de Ingeniería Industrial y los alumnos de la asignatura de Control Avanzado del Master de Automática e Informática Industrial del curso académico 2009/2010.

Para la inicialización del sistema se han creado tres cuestionarios online (Anexos 1, 2 y 4) que deben rellenar los alumnos para conocer sus intereses, estilos de aprendizaje y motivación y poder configurar así el Modelo del Alumno estático. Los cuestionarios se han diseñado con la herramienta de software libre *LimeSurvey* que pone a disposición de los docentes la Universidad Politécnica de Valencia (UPV). El porcentaje de participación medio de todos los alumnos en las tres encuestas ha sido bajo, del orden del 20,8%, recogiendo un total de 48 cuestionarios completados totalmente (Fig. 9.1).



Estado	SID	Encuesta	Fecha de creación	Dueño/a	Acceso	Respuestas anónimas	Respuestas			Encuestados/as disponibles	Frecuencia de respuesta
							Completo	Parcial	Total		
	18297	Cuestionario de Metas para Adolescentes (CMA)	14.12.2009	20461900L	Cerrado	No	53	0	53	232	22.4%
	74357	Cuestionario sobre el Índice de Estilos de Aprendizaje de Felder y Soloman (IEA)	14.12.2009	20461900L	Cerrado	No	62	0	62	231	26.8%
	18357	Escala de autoestima de Rosenberg (RSES)	14.12.2009	20461900L	Cerrado	No	60	2	62	231	26%

Figura 9.1: Encuestas para inicializar el Modelo del Alumno

Los datos obtenidos son mixtos cualitativos y cuantitativos. Debido a la naturaleza de los datos, se ha realizado un estado del arte para determinar qué algoritmo es capaz de agrupar eficientemente dichos datos y determinar las clases de alumnos iniciales.

El algoritmo elegido: *fuzzy k-prototypes* se ha mejorado modificando el cálculo de la función de la distancia, y se ha realizado una comparativa entre los dos algoritmos sobre dos bases de datos ampliamente estudiadas del repositorio *UCI Machine Learning Repository*: Heart Disease Data Set y Statlog (Australian Credit Approval) Data Set con el fin de verificar la mejora.

Por último, una vez demostrada la validez del algoritmo, se ha aplicado la técnica de agrupamiento no supervisado a los datos de los alumnos de la UPV. Con el fin de determinar el número de clases óptimo, se ha definido un índice de validez de clusters apto para datos cualitativos y cuantitativos. Con los grupos iniciales de alumnos, el sistema ITS adaptativo ya está listo para asignar una ruta de aprendizaje adaptada a las necesidades de dichos grupos. En trabajos futuros se implementará el sistema borroso experto para inferir la metodología más adecuada, el tipo de contenidos y la dificultad de los mismos y se integrará en la plataforma LMS.

## 9.1 Estado del arte de las técnicas de agrupamiento para el análisis de datos cuantitativos y cualitativos

Existe una gran variedad de algoritmos de agrupamiento que funcionan correctamente para datos cualitativos (i.e. k-means, PAM, CLARA, BIRCH, CURE, FCM, FMLE, GDBSCAN, etc.) y otros para los datos cuantitativos (k-modes, COOLCAT, ROCK, LIMBO, STIRR, CACTUS, CLICKS, KEROUAC) [3].

El primer enfoque para resolver el problema del agrupamiento de datos mixtos ha sido la aplicación de los algoritmos anteriores y por ello se ha optado por las siguientes estrategias:

- Convertir los atributos cualitativos a cuantitativos. Resulta difícil la correcta asignación de valor numérico para atributos en los que no existe una corrección, i.e. los colores, etc.
- Discretizar los atributos cuantitativos y aplicar algoritmos de agrupamiento cualitativo. Tiene como inconveniente la pérdida de información derivada de la discretización.

En vista de las desventajas que se aprecian en estas opciones, en [92] se presenta el algoritmo de agrupamiento aglomerativo basado en la medida de similitud de Goodall (SBAC) que pese a trabajar bien con datos mixtos cualitativos y cuantitativos supone un coste computacional elevado. El algoritmo k-prototypes creado en [93] propone una función de coste que considera los atributos cualitativos y cuantitativos por separado, con dos definiciones distintas de distancias. Como se trata de un algoritmo particional, tiene un eficiente coste computacional. En [94] se demuestra que una versión borrosa del algoritmo k-prototypes produce una mejora en el ajuste puesto que el algoritmo es robusto frente al ruido y a los registros incompletos a los que les faltan parte de los datos. En [95] se aplica una versión anterior del algoritmo k-prototypes, llamada k-mean, a datos mixtos y se propone una nueva función de coste y medida de la distancia basada en la co-ocurrencia de atributos. La relevancia de los atributos se basa en la co-ocurrencia y determina el peso asociado a dicho atributo a la hora de calcular la función de coste. Los datos numéricos deben ser discretizados para el cálculo de la co-ocurrencia con lo que se pierde precisión en el proceso. Otro enfoque propuesto por [96] agrupa de forma independiente el subconjunto de atributos cualitativos y por otro lado el subconjunto de datos cuantitativos y aplica una técnica de agrupamiento conjunto basado en la acumulación de indicios, con el fin de combinar los resultados y conseguir las clases finales. Los algoritmos que particionan los datos tienden a ser inestables, aleatorios y poco precisos según [97].

De la revisión anterior se extrae que el método más efectivo y que supone un bajo coste computacional es el algoritmo fuzzy k-prototypes, que se enuncia en el siguiente apartado.

## 9.2 Fuzzy k-prototypes

El algoritmo *fuzzy k-prototypes* [94] integra los algoritmos *k-means* y *k-modes* para agrupar conjuntos de datos con valores mixtos cuantitativos y cualitativos. En este algoritmo se define una medida de disimilitud combinada para tratar con ambos atributos cualitativos y cuantitativos. El proceso de agrupamiento es similar al algoritmo *k-means* con la excepción de que éste usa el enfoque *k-modes* para actualizar

los valores de los atributos cualitativos de los prototipos de agrupamiento, así se conserva la eficiencia del algoritmo *k-means*.

Dado el conjunto  $[X^{(n)}|X^{(c)}]^T = [x_1, x_2, \dots, x_n]$  de  $n$  objetos con  $m$  atributos mixtos cuantitativos y cualitativos, en donde  $X^{(n)}$  representa los atributos cuantitativos y  $X^{(c)}$  representa los atributos cualitativos y dada la matriz  $Z = [Z^{(n)}|Z^{(c)}]^T = [z_1, z_2, \dots, z_k]$  de dimensiones  $m \times k$  que contiene los  $k$  centros de las clases, en donde  $Z^{(n)}$  representa los atributos cuantitativos y  $Z^{(c)}$  representa los atributos cualitativos.

La medida de disimilitud combinada entre dos objetos,  $x_l = [x_l^{(n)}|x_l^{(c)}]^T$  y  $x_i = [x_i^{(n)}|x_i^{(c)}]^T$  con atributos mixtos cuantitativos y cualitativos, se define como:

$$d_n(x_l^{(n)}, x_i^{(n)}) + \beta d_c(x_l^{(c)}, x_i^{(c)}) \quad (1)$$

donde  $d_n(\cdot, \cdot)$  es la distancia Euclídea, normalmente empleada en el algoritmo *k-means*,  $d_c(\cdot, \cdot)$  es la medida de disimilitud utilizada en el algoritmo *k-modes*, definida por:

$$d_c(z_l, x_i) \equiv \sum_{j=1}^m \delta(z_{lj}, x_{ij}) \quad (2)$$

donde  $z_l = [z_{l1}, \dots, z_{lm}]^T$ ,  $x_i = [x_{i1}, \dots, x_{im}]^T$  y

$$\delta(z_{lj}, x_{ij}) = \begin{cases} 0 & \text{si } z_{lj} = x_{ij} \\ 1 & \text{si } z_{lj} \neq x_{ij} \end{cases} \quad (3)$$

y  $\beta$  es un peso para ponderar la parte cuantitativa y cualitativa y evitar así que se favorezca algún tipo de atributo.

El problema de optimización para el agrupamiento borroso de un conjunto de  $n$  objetos descritos por  $m$  atributos mixtos cualitativos y cuantitativos en  $k$  clases es:

$$\min_{W, Z} F(W, Z) = \sum_{l=1}^k \sum_{i=1}^n w_{li}^\alpha [d_n(z_l^{(n)}, x_i^{(n)}) + \beta d_c(z_l^{(c)}, x_i^{(c)})] \quad (4)$$

sujeto a las restricciones:

$$0 \leq w_{li} \leq 1, \quad 1 \leq l \leq k, \quad 1 \leq i \leq n \quad (5)$$

$$\sum_{l=1}^k w_{li} = 1, \quad 1 \leq i \leq n \quad (6)$$

$$0 < \sum_{i=1}^n w_{li} < n, \quad 1 \leq l \leq k \quad (7)$$

El método más común para la optimización de  $F$  en la Ec. (4) consiste en emplear una optimización parcial para  $Z$  y para  $W$ . En este método, primero se fija  $Z$  y se encuentran las condiciones necesarias para  $W$  de modo que se minimice  $F$ . A continuación se fija  $W$  y se minimiza  $F$  con respecto a  $Z$ .

Este proceso se formaliza en el algoritmo siguiente:

- Paso 1: Elegir un punto inicial  $|Z^{(1)} \in \mathfrak{R}^{km}$ .
  - Determinar  $W^{(1)}$  de forma que se minimice  $F(W, Z^{(1)})$ . Configurar  $\tau=1$ .
- Paso 2: Determinar  $|Z^{(\tau+1)}$  de forma que se minimice  $|F(W^{(\tau)}, Z^{(\tau+1)})$ .
  - Si  $|F(W^{(\tau)}, Z^{(\tau+1)}) = |F(W^{(\tau)}, Z^{(\tau)})$  se finaliza el algoritmo, en caso contrario, ir al Paso 3.
- Paso 3: Determinar  $|W^{(\tau+1)}$  de forma que se minimice  $|F(W^{(\tau+1)}, Z^{(\tau+1)})$ .
  - Si  $|F(W^{(\tau+1)}, Z^{(\tau+1)}) = |F(W^{(\tau)}, Z^{(\tau+1)})$  se finaliza el algoritmo, en caso contrario, se establece  $\tau=\tau+1$  y se va al Paso 2.

Las matrices  $W$  y  $Z$  pueden calcularse del mismo modo que en el algoritmo fuzzy  $k$ -modes. Dada una matriz  $Z$  fija, por ejemplo  $z_l$  para unas clases  $l=1, 2, \dots, k$  dadas,  $W$  se puede encontrar utilizando la fórmula:

$$w_{li} = \begin{cases} 1 & \text{si } x_i = z_l, \\ 0 & \text{si } x_i = z_h \text{ con } h \neq l, \\ 1 / \sum_{h=1}^k \left[ \frac{d_n(z_l^{(n)}, x_i^{(n)}) + \beta d_c(z_l^{(c)}, x_i^{(c)})}{d_n(z_h^{(n)}, x_i^{(n)}) + \beta d_c(z_h^{(c)}, x_i^{(c)})} \right] & \text{si } x_i \neq z_l \text{ y } x_i \neq z_h, 1 \leq h \leq k \end{cases} \quad (8)$$

para  $1 \leq l \leq k, 1 \leq i \leq n$ .

Dada una matriz de pertenencias fija  $W$ , se puede obtener  $Z^{(c)}$  por el método de actualización de  $k$ -modes. Cada objeto cualitativo se describe por  $m$  atributos cualitativos, de modo que el atributo número  $j$  dispone de  $n_j$  categorías:  $a_j^{(1)}, a_j^{(2)}, \dots, a_j^{(n_j)}$  para  $1 \leq j \leq m$ . Siendo el  $l$ ésimo centro de la clasificación  $z_l = [z_{l1}, z_{l2}, \dots, z_{lm}]^T$ . Entonces se minimiza  $F(W, Z)$  si y sólo si:

$$z_{lj} = a_j^{(r)} \text{ donde } \sum_{i, x_{ij}=a_j^{(r)}} w_{li}^\alpha \geq \sum_{i, x_{ij}=a_j^{(t)}} w_{li}^\alpha, \quad 1 \leq t \leq n_j \quad (9)$$

y  $Z^{(n)}$  se puede calcular según:

$$z_{lj} = \frac{\sum_{i=1}^n w_{li}^\alpha x_{ij}}{\sum_{i=1}^n w_{li}^\alpha}, \quad 1 \leq l \leq k, \quad 1 \leq j \leq m \quad (10)$$

### 9.3 Fuzzy k-prototypes modificado

El algoritmo *fuzzy k-prototypes* se ha modificado, añadiéndole dos mejoras. Por un lado, se ha ampliado el peso  $\beta$  que en un principio era genérico y afectaba a todos los registros cualitativos por igual, pasando a ser ahora  $\beta_i$  que puede adoptar valores diferentes para cada registro cualitativo. Con ello se gana el poder ponderar aquellos atributos donde se verifique, mediante un análisis de correlación, que tienen una mayor importancia o peso sobre la salida.

$$d_n(x_l^{(n)}, x_i^{(n)}) + \beta_i d_c(x_l^{(c)}, x_i^{(c)}) \quad (1b)$$

Por otro lado en [98] se plantea que, para el aprendizaje supervisado, es cierto que la distancia observada es  $\delta(p, q) = 0$  para  $p = q$ , pero no es necesariamente cierto que  $\delta(p, q) = 1$  para  $p \neq q$ . De acuerdo con esto,  $\delta(p, q)$  adopta valores diferentes para las distintas distancias entre pares de atributos cualitativos y depende de las frecuencias relativas de los pares de datos dentro de una clase.

Con la idea de reflejar estas distancias de disimilitud borrosas se ha añadido una matriz de pertenencias intermedia  $\mu$ , en el paso 2 del algoritmo, el correspondiente al cálculo de los nuevos centroides  $Z$ . La matriz de pertenencia intermedia  $\mu$  se obtiene de la combinación de  $w_z$  para los datos cualitativos (Ec. (9)) y de  $w$  para los datos cuantitativos (Ec. (8)) normalizados según:

$$\mu_{li} = \begin{cases} w_{li} / \sum_{i=1}^n w_{li} & \text{para } x_i^{(n)} \\ w_{li}^\alpha / \sum_{j=1}^m w_{li}^\alpha & \text{para } x_i^{(c)} \end{cases} \quad (11)$$

Esta matriz modifica el cálculo de la distancia de la Ec.(1) al ponderarla según la ecuación:

$$\mu_{li} (d_n(x_l^{(n)}, x_i^{(n)}) + \beta_i d_c(x_l^{(c)}, x_i^{(c)})) \quad (1c)$$

para  $1 \leq l \leq k$ ,  $1 \leq i \leq n$ .

Este proceso se formaliza en el algoritmo siguiente:

- Paso 1: Elegir un punto inicial aleatorio  $Z^{(1)} \in \mathfrak{R}^{km}$  y asignar  $\mu_{li} = 1$ .
  - Determinar  $W^{(1)}$  de forma que se minimice  $F(W, Z^{(1)})$ . Configurar  $\tau = 1$ .
- Paso 2: Determinar  $Z^{(\tau+1)}$  y  $\mu_{li}^{(\tau+1)}$  de forma que se minimice  $F(W^{(\tau)}, Z^{(\tau+1)})$ .
  - Si  $|F(W^{(\tau)}, Z^{(\tau+1)}) - F(W^{(\tau)}, Z^{(\tau)})| < \xi$  se finaliza el algoritmo, en caso contrario, ir al Paso 3.
- Paso 3: Determinar  $W^{(\tau+1)}$  con  $\mu_{li}^{(\tau+1)}$  de forma que se minimice  $F(W^{(\tau+1)}, Z^{(\tau+1)})$

- Si  $|F(W^{(\tau+1)}, Z^{(\tau+1)}) - F(W^{(\tau)}, Z^{(\tau+1)})| < \xi$  se finaliza el algoritmo, en caso contrario, se establece  $\tau = \tau + 1$  y se va al Paso 2.

Siendo  $\xi$  el error admisible, que se define con un valor de 0,0001.

Con la nueva definición de la distancia (Ec. 1c), se calcula la nueva matriz de pertenencias según la ecuación:

$$w_{li} = \begin{cases} 1 & \text{si } x_i = z_l, \\ 0 & \text{si } x_i = z_h \text{ con } h \neq l, \\ 1 / \sum_{h=1}^k \left[ \frac{\mu_{li}(d_n(z_l^{(n)}, x_i^{(n)}) + \beta_i d_c(z_l^{(c)}, x_i^{(c)}))}{\mu_{hi}(d_n(z_h^{(n)}, x_i^{(n)}) + \beta_i d_c(z_h^{(c)}, x_i^{(c)}))} \right] & \text{si } x_i \neq z_l \text{ y } x_i \neq z_h, 1 \leq h \leq k \end{cases} \quad (8b)$$

para  $1 \leq l \leq k, 1 \leq i \leq n$ .

Del mismo modo, la nueva función de coste a minimizar se define según la ecuación:

$$\min_{W, Z} F(W, Z) = \sum_{l=1}^k \sum_{i=1}^n w_{li}^\alpha \left[ \mu_{li}(d_n(z_l^{(n)}, x_i^{(n)}) + \beta_i d_c(z_l^{(c)}, x_i^{(c)})) \right] \quad (4b)$$

sujeta a las restricciones:

$$0 \leq w_{li} \leq 1, \quad 1 \leq l \leq k, \quad 1 \leq i \leq n \quad (5)$$

$$\sum_{l=1}^k w_{li} = 1, \quad 1 \leq i \leq n \quad (6)$$

$$0 < \sum_{i=1}^n w_{li} < n, \quad 1 \leq l \leq k \quad (7)$$

## 9.4 Resultados experimentales

Para evaluar el funcionamiento del algoritmo fuzzy k-prototypes modificado, se ha llevado a cabo el agrupamiento en clases de una serie de conjuntos de datos mixtos (cualitativos y cuantitativos) extraídas del repositorio *UCI machine learning repository*: Heart Disease y Statlog (Australian Credit Approval) cuya descripción se muestra en la tabla 9.1.

Como los datos cuantitativos de cada atributo tienen diferentes valores, se han normalizado, con media cero y varianza unitaria, para poder medir correctamente las distancias entre ellos, antes de aplicar el algoritmo. También se observa en la Tabla 9.1 que los datos de los objetos están incompletos, faltando información.

	Número de atributos		Número de objetos	Número de clases	Datos incompletos
	Cualitativos	Cuantitativos			
Heart Disease	6	8	303 (276)	2	Si
Credit Approval	9	6	690 (653)	2	Si

**Tabla 9.1:** Descripción de los conjuntos de datos

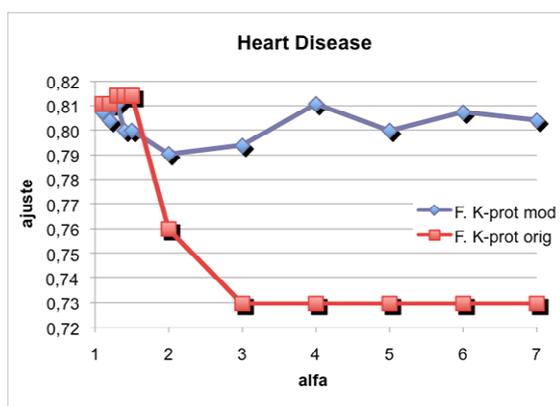
El modo para poder aplicar el algoritmo ha sido el eliminar estos registros incompletos, quedando el número de objetos que aparecen entre paréntesis.

Se ha realizado una comparativa del agrupamiento no supervisado, empleando los dos algoritmos: fuzzy k-prototypes original y modificado, para cada conjunto de datos. Con el fin de comparar su eficiencia, se define el ajuste como:

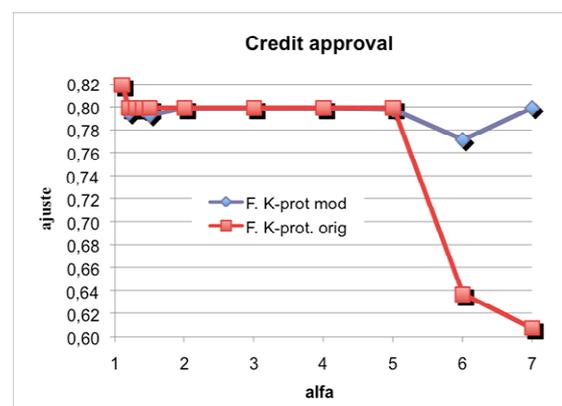
$$r = \frac{\sum_{l=1}^k a_l}{n} \quad (12)$$

Donde  $a_l$  es el número de objetos, de los  $n$  objetos que conforman el conjunto de datos, que ha sido correctamente agrupado en cada una de las  $k$  clases.

Las figuras 9.1 y 9.2 muestran las gráficas del conjunto de datos Heart Disease y Credit Approval, respectivamente. En ellas se aprecia el ajuste para diferentes valores del coeficiente de borrosidad  $\alpha$ . Se observa que con  $\alpha=1.1$ , se obtiene el valor máximo, siendo ligeramente superior el ajuste para fuzzy k-prototypes modificado, aunque la diferencia es mínima. Pese a que se trata de dos conjuntos de datos con diferente número y tipo de atributos, el ajuste es muy similar para ambos (0,81 frente a 0,82, respectivamente). Para el algoritmo fuzzy k-prototypes original existe una bajada del ajuste con respecto al coeficiente de borrosidad, más rápido y de forma exponencial en el caso de Heart Disease y un poco más lento en el otro, a partir del valor  $\alpha=1.5$  y  $\alpha=5$ , respectivamente. En cambio, el algoritmo modificado mantiene un buen ajuste más constante e independiente del factor de borrosidad.



**Figura 9.1:** Heart Disease



**Figura 9.2:** Credit Approval

Como se ha observado que el factor de borrosidad más efectivo es  $\alpha=1.1$ , se ha establecido este valor para comparar el efecto del siguiente parámetro:  $\beta$  que pondera los atributos cualitativos respecto a los cuantitativos, que no tienen ponderación.

Para cada conjunto de datos se ha calculado la matriz de coeficientes de correlación que es una medida normalizada de la intensidad de la relación lineal entre dos variables: cada uno de los atributos de entrada respecto a la clase de salida. Los datos no correlados resultan con un coeficiente de correlación 0 y los correlados de 1 (tablas 9.2 y 9.3).

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14
0,23	0,29	<b>0,41</b>	0,15	0,08	0,01	0,18	0,43	<b>0,43</b>	0,43	<b>0,34</b>	0,47	<b>0,53</b>	1

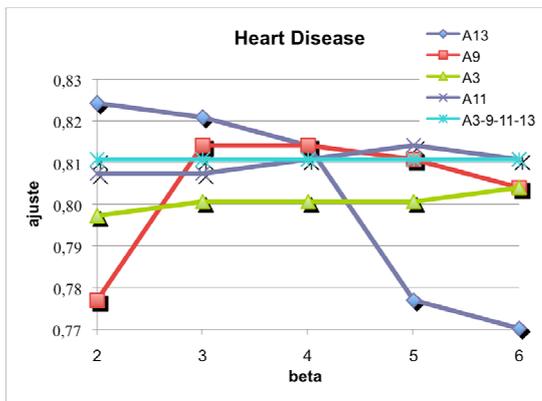
**Tabla 9.2:** Coeficientes de correlación para Heart Disease

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	A11	A12	A13	A14	A15
0,02	0,18	0,21	0,18	<b>0,37</b>	0,25	0,33	<b>0,74</b>	<b>0,45</b>	0,41	0,05	0,10	0,09	0,17	1

**Tabla 9.3:** Coeficientes de correlación para Credit Approval

En las figuras 9.3 y 9.4 se pueden observar los resultados con  $\alpha=1.1$ , respecto al ajuste, al incrementar el valor de  $\beta$  para los atributos cualitativos que tienen un coeficiente de correlación superior a 0,3 (valores resaltados en negrita en las tablas 9.2 y 9.3).

Se aprecia un incremento en el ajuste con esta mejora (2% y 5% respectivamente), siendo el valor máximo para  $\beta=2$  y  $\beta=3$ , respectivamente. Para el conjunto de datos Heart Disease el atributo de mayor correlación produce un ajuste mayor, en cambio, para Credit Approval, se obtiene el mayor ajuste con el segundo atributo de mayor correlación.



**Figura 9.3:** Heart Disease



**Figura 9.4:** Credit Approval

De los ejemplos anteriores se deduce que los parámetros que optimizan el ajuste del método de agrupamiento se deben fijar en un coeficiente de borrosidad de  $\alpha=1.1$  y un incremento de los atributos más correlados con la salida, en una proporción  $\beta=2$  ó 3.

### 9.5 Agrupamiento de los datos de los alumnos de la UPV

Los datos obtenidos de las encuestas de los alumnos han necesitado de un pre-procesado antes de aplicar el algoritmo de agrupamiento fuzzy k-prototypes modificado.

Como se observa en la figura 9.1, se trata de 3 cuestionarios independientes que han sido respondidos de forma desigual por los distintos alumnos y en momentos temporales diferentes, de modo que nos podemos encontrar con que un alumno haya completado 1, 2, 3 o ningún cuestionario. Como el registro no era anónimo, se ha utilizado el correo electrónico de los alumnos como identificador en los distintos cuestionarios para fusionar el conjunto de datos. Al igual que en los ejemplos de conjuntos de datos reales anteriores, se han eliminado aquellos registros en los que el alumno no hubiera completado las tres encuestas correctamente (ver el Anexo 5). A continuación se han convertido los valores cuantitativos de los índices de estilos de aprendizaje (IEA) a valores cualitativos con sus correspondientes clasificadores: Equilibrado, Moderado y Fuerte (ver Anexo 6). No se han convertido los valores cuantitativos del Cuestionario de Metas de Adolescentes (CMA) ni de la Escala de Autoestima de Rosenberg (RSES) para evitar la pérdida de información en la clasificación, puesto que el intervalo de las clases es muy amplio.

En el último paso del algoritmo, cuando se informa de los clusters obtenidos al usuario, sí que se convierten estos atributos (CMA y RSES) para que el usuario obtenga una información más intuitiva y útil.

Se ha añadido también la calificación cualitativa (NP, Suspenso, Aprobado y Notable) con la que han sido evaluados los alumnos en la correspondiente asignatura en este curso escolar 2009/10, que representaría la nota que obtienen los alumnos en cada unidad didáctica a lo largo del curso. Es importante mencionar que la nota de la asignatura Control Avanzado del master no se ha podido incluir por encontrarse pendiente de acta en el momento de la elaboración de la presente memoria. Por tanto, a los 5 alumnos del master se les ha estipulado como calificación NC (no conocida).

	Número de atributos		Número de objetos	Número de clases	Datos incompletos
	Cualitativos	Cuantitativos			
Alumnos UPV	11	4	70 (48)	¿?	Si

**Tabla 9.4:** Descripción del conjunto de datos

En la tabla 9.4 se encuentra un resumen de las características del conjunto de datos de los alumnos de la UPV. Se trata de un conjunto de datos sin una clasificación previa, por lo que a priori no se conoce el número de clases o clusters necesarios para el agrupamiento, de ahí los símbolos de interrogación de la tabla.

Se realizará, por tanto, un agrupamiento no supervisado del conjunto de datos y con el fin de validar el número de clases óptimo, se define a continuación el índice de validación de los clusters.

### Índice de validación de los clusters

El índice de validación de clusters se basa en un criterio de compactación y separación para determinar la bondad de las particiones obtenidas.

En cuanto a los índices para agrupamiento borroso, destaca el criterio de varianza [99], que se define como:

$$S(k) = \sum_{l=1}^k \sum_{i=1}^n w_{li}^{\alpha} \left( \|z_i - c_l\|^2 - \|c_l - \bar{z}\|^2 \right) \quad (12)$$

donde  $n$  es el número de datos,  $k$  es el número de clases,  $\alpha$  es el factor de borrosidad,  $w_{li}$  es la matriz de pertenencias,  $z_i$  es el  $i$ -ésimo vector de datos,  $\bar{z}$  es la media de los datos y  $c_l$  es el centro de la  $l$ -ésima clase. El número de clases se determina incrementando  $k$  hasta que  $S(k)$  alcanza un mínimo local.

Como se trata del agrupamiento de un conjunto de datos mixtos (cualitativos y cuantitativos)  $\bar{z}$  se determina como la media de los datos cuantitativos (que será nula al tratarse de datos normalizados, con media nula y varianza unitaria) y el valor con mayor frecuencia de aparición para los datos cualitativos.

Otra medida de validez para el caso de particiones borrosas puede ser el valor que va tomando la función objetivo a minimizar.

Aplicando estos dos criterios: el índice de validación  $S(k)$  y la función objetivo a minimizar  $F(k)$  para el algoritmo fuzzy k-prototypes modificado se obtienen los resultados que se muestran en la figura 9.5.

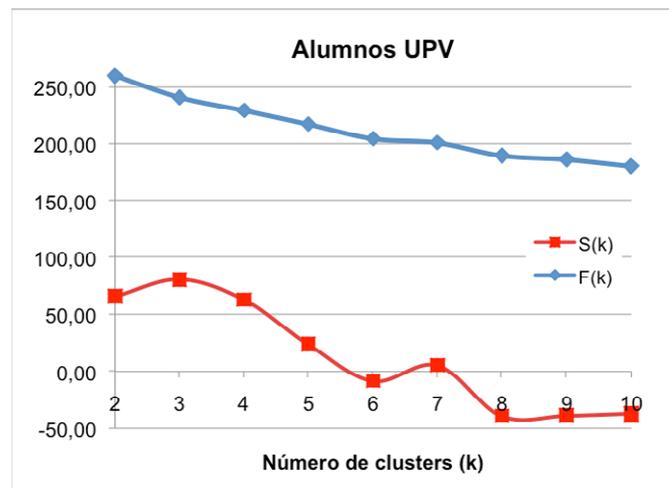


Figura 9.5: Índice de validación de clusters

Se aprecian dos mínimos locales para  $S(k)$ , uno para  $k=6$  y otro para  $k=8$ . También se observa que la función objetivo disminuye con mayor pendiente (21%) hasta alcanzar  $k=6$ , a partir de este punto continúa descendiendo pero con menor pendiente (12%). Con el fin de minimizar el número de clusters y corroborando que la función objetivo también se encuentra en un mínimo local, se elige  $k=6$  como el número de clusters óptimo para el conjunto de datos.

## Resultados

Con la idea de diseñar un sistema de control adaptativo rápido, con baja complejidad computacional, se ha optado por agrupar a los alumnos en función de sus características o atributos y así no tener que proponer tantos cursos distintos como alumnos. Con esta técnica se obtienen clusters cuyos centroides son una media de las características de un grupo o tipo de alumnos con un estilo de aprendizaje común. Es evidente que a mayor número de grupos más ajustados serán los cursos propuestos y menos alumnos lejanos a la media de esas clases tendremos. Con todo, se debe llegar a un compromiso entre el número de grupos y la eficacia del agrupamiento no supervisado.

Con los datos iniciales, aplicando el índice de validación de los clusters basado en el criterio de la varianza (Ec. 12), se obtiene que el número de grupos óptimo es 6, cuyos atributos se muestran en la tabla 9.5.

Ref	Mod	Sen	Mod	Vis	Eq	Sec	Mod	Rs_Pa	Ed_Pa	In_Mi	C_Ba	In	NP
Act	Eq	Sen	Mod	Vis	Fue	Sec	Eq	Rs_Pa	Ed_Do	In_Mi	C_Ba	In	Ap
Ref	Eq	Sen	Eq	Vis	Fue	Glo	Mod	Rs_Do	Ed_Do	In_Mi	C_Ba	In	Su
Act	Mod	Sen	Mod	Vis	Mod	Sec	Eq	Rs_Do	Ed_Pa	In_Mi	C_Ba	In	Ap
Act	Mod	Sen	Eq	Vis	Mod	Glo	Eq	Rs_Pa	Ed_Pa	In_Mi	C_Ba	M	NC
Ref	Eq	Sen	Mod	Vis	Fue	Sec	Eq	Rs_Pa	Ed_Pa	In_Mi	C_Ba	In	Su

*Tabla 9.5:* Características de los clusters

En siguientes iteraciones, a medida que el alumno realice las unidades del curso, se añadirán otros atributos al conjunto de datos (por ejemplo, autonomía, tiempo de dedicación, gustos o preferencias, rendimiento, etc.) y se modificarán algunos atributos iniciales (por ejemplo, la colaboración, la nota, la motivación, la confianza, etc.). Con estos cambios o redefiniciones de parámetros, se recalculará el número de clusters y se realizará de nuevo el agrupamiento no supervisado de los alumnos.

## Sistema borroso experto

En la figura 9.6 se muestra el diagrama de bloques del sistema ITS adaptativo. Como se aprecia, existe una primera fase de inicialización, que es la que se ha implementado en este trabajo de investigación. A partir de los grupos de los alumnos, mediante un sistema borroso experto, se asociarían los grupos con las metodologías didácticas más idóneas. La idea es implementar, mediante una asociación de reglas, las tablas definidas en el capítulo 8.2 que especifican la relación existente entre los estilos de aprendizaje y la autonomía del alumno y la metodología pedagógica (tabla 8.26), el tipo de recurso (tablas 8.27 y 8.30), la dificultad (tablas 8.28 y 8.31) y las herramientas de colaboración (tabla 8.32)). El hecho de que sea un sistema borroso va ligado a las matrices de pertenencia que se obtienen de los agrupamientos y que hacen más robusto al sistema de control.

La secuenciación (planificación de la unidad didáctica) tiene una estructura que viene fijada según el criterio del profesor del curso (las metodologías didácticas descritas en el capítulo 8.1) y las exigencias del currículum, y cuyo desarrollo puede estar sujeto, en ciertos casos, a variaciones en función de la metodología asociada al grupo de alumnos (más tradicional o más cooperativa).

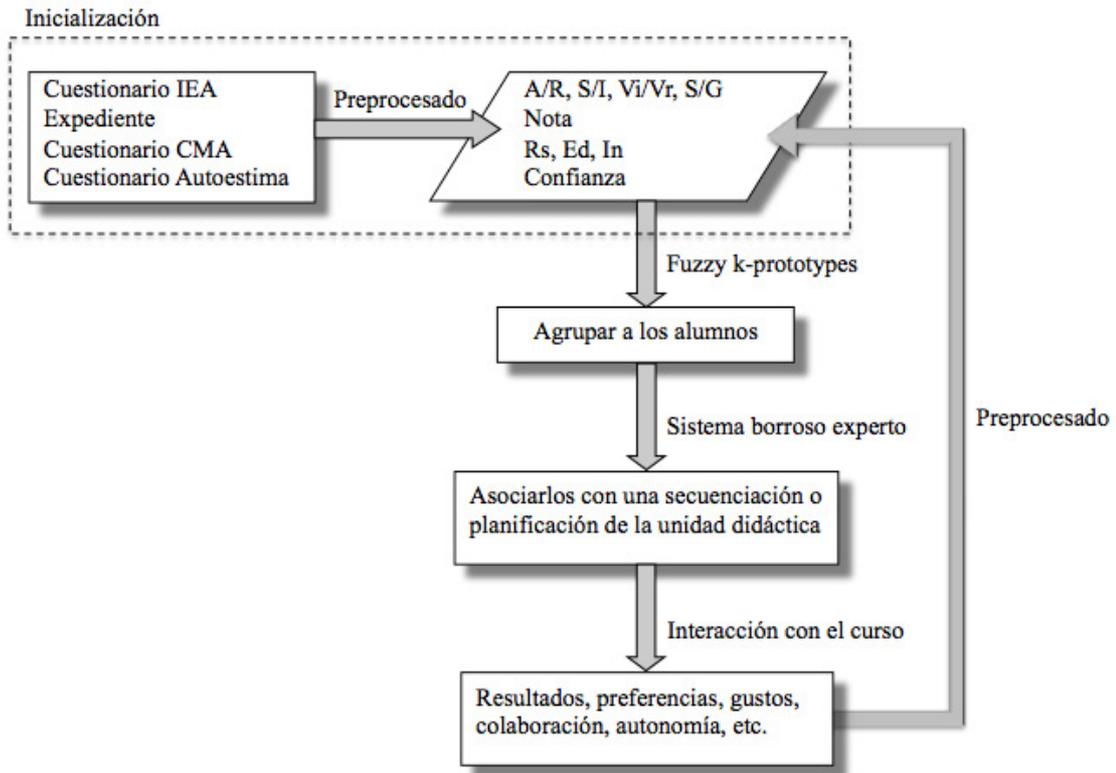


Figura 9.6: Diagrama de bloques del sistema ITS adaptativo

## 9.6 Conclusiones

El tamaño de la muestra utilizada no es muy grande, aunque inicialmente habían 212 alumnos de la asignatura Automatización Industrial y 19 alumnos de Control avanzado, la participación ha sido baja, del 20,8% debido en parte, al poco tiempo en el que han estado operativos los cuestionarios, coincidiendo además con el periodo vacacional de Pascua.

Para estimar si el tamaño de la muestra obtenido es representativo de la población de alumnos encuestados, se debe aplicar la siguiente fórmula [100]:

$$n = \frac{z^2 pq}{B^2} \quad (13)$$

donde  $n$  es el tamaño de la muestra,  $z$  (1,96 para el 95% de confianza, o 2,56 para el 99%),  $p$  es la frecuencia esperada del factor a estudiar,  $q$  corresponde a  $1-p$  y  $B$  es la precisión o error admitido.

El valor de  $n$  obtenido por la Ec.13 indica el tamaño de la muestra para una población infinita, a efectos prácticos, se considera población infinita cuando la muestra supone menos del 5% de la población total.

Cuando la población es pequeña, la muestra obtenida mediante la Ec.13 es demasiado grande, por lo que se debe aplicar la fórmula correctora:

$$\frac{1}{n'} = \frac{1}{n} + \frac{1}{N} \quad (14)$$

donde  $n'$  es el tamaño de la muestra necesario,  $n$  el tamaño de la muestra según la Ec.13 y  $N$  el tamaño de la población.

Si aplicamos la Ec.13, siendo el factor a estudiar la generalización del número de grupos realizados a los alumnos, nos encontramos con que 6 grupos corresponde a una frecuencia del 16,7%, por tanto, con una precisión del 5% y un nivel de confianza del 95% sería necesario un tamaño de muestra de 214 alumnos. Este tamaño de muestra es inferior al 5% del tamaño de la población, por lo que aplicando la fórmula correctora Ec.14 se obtiene un tamaño de muestra de 111 alumnos.

Se puede concluir que la muestra no es representativa y por tanto no se puede inferir que con la totalidad de la población de alumnos se alcanzara el mismo número de grupos que se ha obtenido en la sección 9.5. Sin embargo, la muestra es válida para demostrar la validez del algoritmo y de las técnicas propuestas.

En este capítulo se han revisado las técnicas de agrupamiento de datos cualitativos y cuantitativos. Se ha aportado una mejora al algoritmo fuzzy k-prototypes con la cual se consigue un buen ajuste, más constante e independiente del factor de borrosidad, dado que el algoritmo modificado incluye un mayor grado de borrosidad para los datos cualitativos, con la definición de una matriz de pertenencias adicional para la búsqueda de los nuevos centroides que minimicen la función objetivo.

El algoritmo propuesto es relativamente simple, con un bajo coste computacional del orden  $O(tkmn)$ , donde  $t$  es el número de iteraciones del algoritmo,  $k$  el número de clusters,  $m$  el número de atributos y  $n$  el número de datos o registros. En muchos casos (exceptuando el actual, que tiene un bajo número de registros) se cumple que  $n \gg t, k, m$  por lo que se puede aproximar el coste a  $O(n)$ .



# Capítulo 10

## Conclusiones y trabajo futuro

### 10.1 Conclusiones

El objetivo fundamental de este trabajo de investigación se ha centrado en el desarrollo de un nuevo modelo de Sistema Tutor Inteligente que integre todos los parámetros que definen el proceso de enseñanza y aprendizaje de los alumnos y sea capaz de adaptar automáticamente el curso a las necesidades de los alumnos, proponiendo a cada uno la mejor ruta de aprendizaje en función de su tipología. A lo largo de los primeros capítulos de esta memoria, se ha llevado a cabo una importante recopilación bibliográfica y revisión teórica sobre algunos aspectos básicos relacionados con el tema.

La primera aportación de este trabajo consiste en la aplicación de la teoría de control al sistema ITS, siendo el proceso a controlar el aprendizaje del alumno. El estudiante presenta un estilo de aprendizaje y unos conocimientos iniciales que se modifican dinámicamente a partir de la interacción con el curso. La sensorización del proceso consiste en la extracción y almacenamiento de información en una base de datos para su posterior análisis mediante técnicas de minería de datos. La minería de datos realimenta al controlador del ITS, adaptando los modelos que conforman el LMS. El profesor supervisa el proceso y define los contenidos o currículum del curso. Además de este primer bucle de control descrito, existe otro bucle externo más lento que verifica y controla el buen funcionamiento del ITS y su adecuación al alumno, mediante la comparación de los resultados obtenidos por el alumno respecto de los objetivos que se plantean en el curso. En la bibliografía consultada se plantea el problema de una definición del modelo del dominio del ITS limitada y estructurada, que obliga al profesor a una modificación manual en el proceso de mejora. Este inconveniente queda resuelto en el modelo propuesto donde sí se incorpora una adaptabilidad automatizada para los tres modelos que constituyen la arquitectura del ITS. Además se propone una medida de evaluación del funcionamiento del ITS intrínseca al modelo.

La segunda aportación constituye una definición exhaustiva del modelo del alumno, que lo conforman dos tipos de informaciones, estática y dinámica, en función de si ven o no modificadas por la interacción del alumno con el sistema. La información estática la determina el estilo de aprendizaje del alumno y la dinámica se obtiene de la interacción del usuario y valora el nivel de conocimiento que va adquiriendo, sus motivaciones y la confianza en sí mismo. En este sentido, para inicializar el ITS y poder agrupar a los alumnos en función de su tipología de aprendizaje, se han seleccionado los cuestionarios más idóneos para medir estos factores y se ha realizado un estudio en una muestra de alumnos de la Universidad Politécnica de Valencia. Pese a que el tamaño de la muestra no es apto para la generalización de las conclusiones a la totalidad de alumnos de la UPV, sí que cumple el objetivo de demostrar la validez del algoritmo de agrupamiento y de las técnicas propuestas.

Finalmente destacar la mejora que se ha aportado al algoritmo de agrupamiento fuzzy k-prototypes, con la cual se consigue un buen ajuste, más constante e indepen-

diente del factor de borrosidad. Esto se debe a que el algoritmo modificado incluye un mayor grado de borrosidad para los datos cualitativos, con la definición de una matriz de pertenencias adicional para la búsqueda de los nuevos centroides que minimicen la función objetivo. El número de clases óptimo se ha elegido mediante un índice de validación de clusters, basado en el criterio de varianza de Sugeno y Yasukawa para datos mixtos, cualitativos y cuantitativos y en el valor que va tomando la función objetivo a minimizar.

## 10.2 Trabajo futuro

Fundamentalmente el trabajo desarrollado en este proyecto establece las bases y el estudio previo para, en la Tesis Doctoral, implementar completamente el ITS adaptativo propuesto.

En cuanto a trabajos futuros, podemos sugerir varias líneas de investigación:

1. Demostrar formalmente los algoritmos de agrupamiento y validación del número de clusters enunciados, mediante la realización de una prueba de convergencia y la comprobación de que la propuesta de la matriz de pertenencia borrosa, minimiza la función de coste.
2. Desarrollar el sistema borroso experto que implemente, mediante una asociación de reglas, las tablas definidas en el capítulo 8.2 que especifican la relación entre los diferentes parámetros que definen el proceso de enseñanza y aprendizaje del alumno y permite obtener la ruta de aprendizaje adaptativa para cada estudiante.
3. Implementar el sistema ITS adaptativo en el entorno LMS, añadiendo las formalidades necesarias para incluir los algoritmos de minería de datos propuestos en la plataforma seleccionada.
4. Experimentar y analizar los resultados obtenidos por el sistema ITS adaptativo en una asignatura del área de Ingeniería de Sistemas y Automática para los nuevos cursos del grado, incorporando las prácticas de laboratorios virtuales y remotos al entorno LMS.

# Anexos

## 1. Cuestionario sobre el Índice de Estilos de Aprendizaje de Felder y Soloman (IEA) [47]

Para cada una de las 44 cuestiones de abajo, selecciona bien “a” o “b” para indicar tu respuesta. Por favor, elige sólo una respuesta para cada pregunta. Si ambas “a” y “b” son posibles, elige la que se ajuste más frecuentemente.

1. Entiendo algo mejor después de
  - (a) probarlo
  - (b) pensar en ello
2. Preferiría ser considerado
  - (a) realista
  - (b) innovador
3. Cuando pienso en lo que hice ayer, es más fácil que me venga
  - (a) una imagen
  - (b) palabras
4. Yo suelo
  - (a) entender los detalles de una asignatura pero puede ser que me pierda con la estructura general
  - (b) entender la estructura general pero puede ser que me pierda con los detalles
5. Cuando estoy aprendiendo algo nuevo, me ayuda
  - (a) hablar sobre ello
  - (b) pensar en ello
6. Si fuera un profesor, preferiría enseñar un curso
  - (a) que tratara de hechos y situaciones de la vida real
  - (b) que tratara de ideas y teorías
7. Yo prefiero conseguir nueva información en forma de
  - (a) imágenes, diagramas, gráficos o mapas
  - (b) instrucciones escritas o información verbal
8. Una vez que he entendido
  - (a) todas las partes, entiendo el conjunto
  - (b) el conjunto, veo cómo encajan las partes

9. En un grupo de trabajo con una tarea difícil para realizar, es más probable que yo
- (a) me lance y contribuya con ideas
  - (b) me siente al fondo y escuche
10. Encuentro más sencillo
- (a) aprender hechos
  - (b) aprender conceptos
11. En un libro con muchas imágenes y gráficos, es más probable que yo
- (a) mire las imágenes y gráficos detenidamente
  - (b) me centre en el texto escrito
12. Cuando resuelvo problemas matemáticos
- (a) normalmente trabajo para conseguir la solución paso a paso
  - (b) a menudo veo la solución directamente, pero luego tengo que cavilar para extraer los pasos que llegan hasta ella
13. En las clases a las que he asistido
- (a) solía conocer a la mayoría de los alumnos
  - (b) raramente conocía a muchos de los alumnos
14. A la hora de leer novelas que no sean de ficción, prefiero
- (a) algo que me enseñe nuevos hechos o me cuente cómo hacer algo
  - (b) algo que me de nuevas ideas sobre las que pensar
15. Me gustan los profesores que
- (a) ponen muchos diagramas en la pizarra
  - (b) pasan mucho tiempo explicando
16. Cuando estoy analizando una historia o una novela
- (a) pienso en los incidentes e intento juntarlos para extraer los hilos conductores o temas
  - (b) sólo sé de que van los temas cuando acabo de leerla y luego tengo que volver al inicio y encontrar los incidentes que los demostraban
17. Cuando empiezo con los deberes, es más probable que
- (a) empiece a trabajar en la solución directamente
  - (b) intente entender completamente el problema primero
18. Prefiero la idea de
- (a) certidumbre
  - (b) teoría

19. Recuerdo mejor
- (a) lo que veo
  - (b) lo que escucho
20. Para mí es más importante que un profesor
- (a) exponga el material en pasos secuenciales claros
  - (b) me dé una imagen general y relacione el material con otras asignaturas
21. Prefiero estudiar
- (a) en un grupo de trabajo
  - (b) solo
22. Es más probable que me consideren
- (a) cuidadoso con los detalles de mi trabajo
  - (b) creativo en cómo hacer mi trabajo
23. Cuando me dan indicaciones para llegar a un sitio, prefiero
- (a) un mapa
  - (b) instrucciones escritas
24. Aprendo
- (a) con un paso regular y estable. Si estudio duro, lo conseguiré.
  - (b) a tropezones. Puede que esté totalmente confundido, pero de repente, todo encaja.
25. Seguramente primero
- (a) intentaría sacar las cosas
  - (b) pensaría en cómo voy a hacerlo
26. Cuando estoy leyendo por hobby, prefiero a los escritores que
- (a) dicen claramente lo que están pensando
  - (b) dicen cosas de forma creativa e interesante
27. Cuando veo un diagrama o un croquis en clase, es más probable que recuerde
- (a) la imagen
  - (b) lo que el profesor dijo sobre él
28. Cuando considero un cuerpo de información, es más probable que
- (a) me centre en los detalles y me pierda la imagen general
  - (b) intente entender la imagen general antes de ponerme con los detalles
29. Recuerdo más fácilmente
- (a) algo que he hecho
  - (b) algo sobre lo que he estado pensando mucho

30. Cuando tengo que realizar una tarea, prefiero
- (a) controlar una forma concreta de hacerla
  - (b) descubrir nuevas formas de hacerla
31. Cuando alguien me está enseñando datos, prefiero
- (a) gráficos o diagramas
  - (b) texto que sintetice los resultados
32. Cuando estoy escribiendo un artículo, es más probable que
- (a) trabaje (pensando y escribiendo) en el principio del artículo y progrese hacia adelante
  - (b) trabaje (pensando y escribiendo) en diferentes partes del artículo y luego las ordene
33. Cuando tengo que trabajar en un proyecto en grupo, lo que primero quiero es
- (a) tener una “lluvia de ideas” en la que todos contribuyan con ideas
  - (b) pensar individualmente y luego juntarnos como grupo para comparar las ideas
34. Tengo en mayor estima llamar a alguien
- (a) sensato
  - (b) imaginativo
35. Cuando conozco a gente en una fiesta, es más probable que recuerde
- (a) cómo eran
  - (b) lo que dijeron sobre ellos
36. Cuando estoy aprendiendo una asignatura nueva, prefiero
- (a) estar en centrado en esa asignatura, aprendiendo lo máximo que pueda
  - (b) intentar hacer conexiones entre esa asignatura y las asignaturas relacionadas.
37. Es más probable que me consideren
- (a) sociable
  - (b) reservado
38. Prefiero los cursos que enfatizan
- (a) los materiales concretos (hechos, datos)
  - (b) los materiales abstractos (conceptos, teorías)
39. Para distraerme, preferiría
- (a) mirar la televisión
  - (b) leer un libro

- 
40. Algunos profesores empiezan sus clases con un esquema de lo que tratarán. Dichos esquemas me resultan
- (a) de alguna manera útiles
  - (b) muy útiles
41. La idea de hacer trabajos en grupos, que se valoren con puntos adicionales para el grupo entero
- (a) me llama la atención
  - (b) no me llama la atención
42. Cuando estoy haciendo cálculos largos,
- (a) suelo repetir todos los pasos y comprobar mi trabajo cuidadosamente
  - (b) comprobar mi trabajo me resulta tedioso y tengo que forzarme para hacerlo
43. Suelo visualizar sitios en los que he estado
- (a) fácilmente y con bastante precisión
  - (b) con dificultad y sin mucho detalle
44. Cuando estoy resolviendo problemas en un grupo, es más probable que
- (a) piense en los pasos del proceso de resolución
  - (b) piense en posibles consecuencias o aplicaciones de la solución en un amplio rango de áreas

## 2. Cuestionario de Metas para Adolescentes (CMA) [86] [87]

ES IMPORTANTE PARA MÍ:						
	1. Si es nada importante	2. Si es poco importante	3. Si es importante	4. Si es bastante importante	5. si es muy importante	6. si es importantísimo
1. Ser buen estudiante.						
2. Ser considerado un héroe.						
3. Divertirme en todo momento.						
4. Pertenecer a un equipo deportivo.						
5. Aprobar los exámenes.						
6. Ayudar a los demás.						
7. Hacer las cosas a mi manera.						
8. Sobresalir en deporte.						
9. Aprender cosas nuevas.						
10. Ser sincero con los demás.						
11. Respetar a la otra persona en una relación de amistad.						
12. Tener mucho poder.						
13. Poder vivir por mi cuenta.						
14. Jugar en equipos destacados.						
15. Cambiar el sistema político actual.						
16. Finalizar las tareas en el tiempo asignado.						
17. Escuchar a los demás.						
18. Ser líder de un grupo.						
19. Tener bastante dinero.						
20. Ser mejor que otros en deporte.						
21. Terminar las etapas educativas sin repetir ningún curso.						
22. Defender mis ideales políticos.						
23. Ser justo con los demás.						
24. Ser perdonado si ofendo a otras personas.						
25. Alcanzar prestigio social.						
26. Llevar los últimos diseños de moda.						
27. Estar en buena forma física.						
28. Conseguir notas altas para poder acceder a la universidad.						
29. Ser leal con los demás.						
30. Ser centro de atención de los demás.						
31. Asistir a muchas fiestas.						
32. Realizar actividades de riesgo.						
33. Vestir de manera llamativa.						

<b>ES IMPORTANTE PARA MÍ:</b>	1. Si es nada importante	2. Si es poco importante	3. Si es importante	4. Si es bastante importante	5. si es muy importante	6. si es importantísimo
34. Obtener mejores notas que mis amigos.						
35. Ser una persona de confianza.						
36. Pertenecer a un partido político.						
37. Realizar actividades distintas de las habituales.						
38. Garantizar el derecho de autodeterminación de un pueblo.						
39. Vivir alejado de la familia.						
40. Relacionarme con la gente.						
41. Mejorar mis marcas deportivas.						
42. Esforzarme en tareas difíciles.						
43. Comparar los contenidos nuevos con los ya adquiridos.						
44. Ser considerado superior a los demás.						
45. Comparar siempre lo que quiero.						
46. Participar en manifestaciones políticas.						
47. Estudiar utilizando resúmenes, cuadros, diagramas, etc.						
48. Tener libertad para tomar mis propias decisiones.						
49. Satisfacer mis impulsos lo antes posible.						
50. Llegar a ser gerente de una empresa.						
51. Estar entre los mejores en todas las actividades.						
52. Ser mencionado en situaciones especiales.						
53. Apoyar a quienes utilizan la violencia en la política.						
54. Planificar mis tareas y controlar su realización.						
55. Hacer cosas prohibidas.						
56. Practicar el ejercicio físico.						
57. Obtener la aprobación de los profesores.						
58. Ser reconocido como un líder político.						
59. Aceptar a los demás tal como son.						
60. Programar las actividades físicas.						
61. Tener libertad para hacer lo que yo quiera.						
62. Rechazar lo que me imponen.						
63. Disfrutar con el deporte.						

<b>ES IMPORTANTE PARA MÍ:</b>	1. Si es nada importante	2. Si es poco importante	3. Si es importante	4. Si es bastante importante	5. si es muy importante	6. si es importantísimo
64. Disfrutar del sexo.						
65. Relajarme con el deporte.						
66. Expresar mis opiniones sin ofender a nadie.						
67. Identificarme con quienes me comunican sus emociones.						
68. Reunirme con amigos.						
69. Ser delicado en las conversaciones con los demás.						
70. Conseguir las metas que me propongo.						
71. Esforzarme poco en el logro de las metas.						
72. Esforzarme mucho en el logro de las metas.						
73. Tomar las metas con cierta tranquilidad.						
74. Cambiar de metas dependiendo de cómo vayan las cosas.						
75. Evaluar las metas que consigo.						
76. Renunciar a las metas que requieren mucho esfuerzo.						
77. Estudiar la posibilidad de alcanzar determinadas metas.						
78. Analizar el tiempo que tardaré en lograr una meta.						
79. Trabajar con ilusión para conseguir las metas que deseo.						

### 3. Corrección de los factores [86]

NI	Factor	Definición del ítem: <i>Es importante para mí</i>
1	Ed	Ser buen estudiante
2	Rs	Ser considerado un héroe
5	Ed	Aprobar los exámenes
6	In	Ayudar a los demás
9	Ed	Aprender cosas nuevas
10	In	Ser sincero con los demás
11	In	Respetar a la otra persona en una relación de amistad
12	Rs	Tener mucho poder
16	Ed	Finalizar las tareas en el tiempo asignado
17	In	Escuchar a los demás
18	Rs	Ser líder de un grupo
19	Rs	Tener bastante dinero
21	Ed	Terminar las etapas educativas sin repetir ningún curso
23	In	Ser justo con los demás
24	In	Ser perdonado si ofendo a otras personas
25	Rs	Alcanzar prestigio social
26	Rs	Llevar los últimos diseños de moda
28	Ed	Conseguir notas altas para poder acceder a la Universidad
29	In	Ser leal con los demás
30	Rs	Ser centro de atención de los demás
33	Rs	Vestir de manera llamativa
34	Rs	Obtener mejores notas que mis amigos
35	In	Ser una persona de confianza
40	In	Relacionarme con la gente
42	Ed	Esforzarme en tareas difíciles
43	Ed	Comparar los contenidos nuevos con los ya adquiridos
44	Rs	Ser considerado superior a los demás
45	Rs	Comprar siempre lo que quiero
47	Ed	Estudiar utilizando resúmenes, cuadros, diagramas, etc.
50	Rs	Llegar a ser gerente de una empresa
51	Rs	Estar entre los mejores en todas las actividades
52	Rs	Ser mencionado en situaciones especiales
54	Ed	Planificar mis tareas y controlar su realización
57	Ed	Obtener la aprobación de los profesores
59	In	Aceptar a los demás tal como son
66	In	Expresar mis opiniones sin ofender a nadie
67	In	Identificarme con quienes me comunican sus emociones
68	In	Reunirme con amigos
69	In	Ser delicado en las conversaciones con los demás

#### 4. Escala de autoestima de Rosenberg (RSES) [87]

		Totalmente de acuerdo	De acuerdo	En desacuerdo	Totalmente en desacuerdo
1	En general, estoy satisfecho conmigo mismo				
2	A veces pienso que no soy bueno en nada				
3	Tengo la sensación de que poseo algunas buenas cualidades				
4	Soy capaz de hacer las cosas tan bien como la mayoría de las personas				
5	Siento que no tengo demasiadas cosas de las que sentirme orgulloso				
6	A veces me siento realmente inútil				
7	Tengo la sensación de que soy una persona de valía al menos igual que la mayoría de la gente				
8	Ojalá me respetara más a mi mismo				
9	En definitiva, tiendo a pensar que soy un fracasado				
10	Tengo una actitud positiva hacia mi mismo				

## 5. Conjunto de datos de los alumnos (pre-procesados 1)

Identificador	IEA				CMA			RSES	Origen
	Pr	Pe	E	C	Rs	Ed	In		
'alumno1@etsid.upv.es'	7	1	7	-3	25	47	68	27	'Master'
'alumno2@upvnet.upv.es'	-7	5	-9	3	30	43	67	26	'Master'
'alumno3@upvnet.upv.es'	-7	-7	9	-5	51	49	66	25	'Master'
'alumno4@etsid.upv.es'	7	3	5	1	48	52	74	29	'Master'
'alumno5@upvnet.upv.es'	7	-3	5	-1	27	44	60	21	'Master'
'alumno6@etsii.upv.es'	1	5	5	3	26	53	79	23	'Industriales'
'alumno7@etsid.upv.es'	5	7	7	1	60	46	63	26	'Industriales'
'alumno8@etsii.upv.es'	-3	7	3	7	35	52	67	22	'Industriales'
'alumno9@etsii.upv.es'	3	3	7	5	24	39	67	27	'Industriales'
'alumno10@euitiv.upv.es'	-5	7	3	5	36	39	65	26	'Industriales'
'alumno11@etsid.upv.es'	-5	5	9	1	52	55	71	26	'Industriales'
'alumno12@etsid.upv.es'	1	5	1	-5	31	54	81	27	'Industriales'
'alumno13@etsii.upv.es'	-1	3	11	-1	39	48	69	21	'Industriales'
'alumno14@teleco.upv.es'	3	7	11	5	60	54	69	24	'Industriales'
'alumno15@etsid.upv.es'	9	5	3	1	50	50	69	26	'Industriales'
'alumno16@etsid.upv.es'	-9	-3	-5	3	34	59	81	26	'Industriales'
'alumno17@etsid.upv.es'	3	5	9	1	47	47	74	26	'Industriales'
'alumno18@etsid.upv.es'	5	5	7	-1	54	37	58	22	'Industriales'
'alumno19@etsid.upv.es'	-1	-1	5	5	52	60	69	25	'Industriales'
'alumno20@etsid.upv.es'	-1	3	9	-5	38	57	72	26	'Industriales'
'alumno21@etsid.upv.es'	3	1	9	-7	48	49	67	27	'Industriales'
'alumno22@euitiv.upv.es'	-3	1	11	-7	57	54	66	24	'Industriales'
'alumno23@etsid.upv.es'	9	1	7	1	33	55	76	26	'Industriales'
'alumno24@etsii.upv.es'	5	7	11	-3	54	48	70	26	'Industriales'
'alumno25@etsid.upv.es'	-3	5	5	1	47	50	69	24	'Industriales'
'alumno26@etsid.upv.es'	5	5	5	11	19	44	67	27	'Industriales'
'alumno27@etsid.upv.es'	7	-1	11	3	43	49	66	22	'Industriales'
'alumno28@etsid.upv.es'	-5	3	9	3	49	51	62	23	'Industriales'
'alumno29@etsid.upv.es'	9	-1	7	-1	71	58	69	25	'Industriales'
'alumno30@etsid.upv.es'	5	5	7	1	35	53	71	23	'Industriales'
'alumno31@etsid.upv.es'	1	1	1	7	34	45	69	25	'Industriales'
'alumno32@etsid.upv.es'	-3	1	11	5	55	52	69	25	'Industriales'
'alumno33@etsid.upv.es'	-7	11	9	5	47	58	73	21	'Industriales'
'alumno34@etsid.upv.es'	1	7	9	3	27	56	71	26	'Industriales'
'alumno35@etsii.upv.es'	3	3	11	-3	34	36	68	24	'Industriales'
'alumno36@etsid.upv.es'	7	3	1	5	34	46	64	20	'Industriales'
'alumno37@etsid.upv.es'	5	1	9	5	44	48	64	25	'Industriales'
'alumno38@etsid.upv.es'	3	7	11	-5	52	46	56	29	'Industriales'
'alumno39@etsid.upv.es'	3	7	5	5	37	46	62	26	'Industriales'
'alumno40@posgrado.upv.es'	-1	7	9	-1	55	57	65	30	'Master'
'alumno41@etsid.upv.es'	5	5	9	1	57	47	60	25	'Industriales'
'alumno42@etsid.upv.es'	-1	-3	7	5	35	50	61	28	'Industriales'
'alumno43@etsid.upv.es'	3	7	11	1	42	41	60	25	'Industriales'
'alumno44@etsid.upv.es'	3	3	11	-1	49	59	64	27	'Industriales'
'alumno45@etsid.upv.es'	-1	-7	5	-1	71	57	68	24	'Industriales'
'alumno46@etsid.upv.es'	5	7	-1	1	24	45	74	28	'Industriales'
'alumno47@etsid.upv.es'	-1	7	9	1	40	42	62	26	'Industriales'
'alumno48@etsii.upv.es'	-1	9	9	-1	24	26	58	22	'Industriales'

## 6. Conjunto de datos de los alumnos (pre-procesados 2)

	IEA							CMA				Origen	Nota
	Pr	Pe	E	C	Rs	Ed	In	RSES					
'Activo'	'Moderado'	'Sensorial'	'Equilibrado'	'Visual'	'Moderado'	'Global'	'Equilibrado'	25	47	68	27	'Master'	NC'
'Reflexivo'	'Moderado'	'Sensorial'	'Moderado'	'Verbal'	'Fuerte'	'Secuencial'	'Equilibrado'	30	43	67	26	'Master'	NC'
'Reflexivo'	'Moderado'	'Intuitiva'	'Moderado'	'Visual'	'Fuerte'	'Global'	'Moderado'	51	49	66	25	'Master'	NC'
'Activo'	'Moderado'	'Sensorial'	'Equilibrado'	'Visual'	'Moderado'	'Secuencial'	'Equilibrado'	48	52	74	29	'Master'	NC'
'Activo'	'Moderado'	'Intuitiva'	'Moderado'	'Visual'	'Moderado'	'Global'	'Equilibrado'	27	44	60	21	'Master'	NC'
'Activo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Moderado'	'Secuencial'	'Equilibrado'	26	53	79	23	'Industriales'	Aprobado'
'Activo'	'Moderado'	'Sensorial'	'Moderado'	'Visual'	'Moderado'	'Secuencial'	'Equilibrado'	60	46	63	26	'Industriales'	Aprobado'
'Reflexivo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Equilibrado'	'Secuencial'	'Moderado'	35	52	67	22	'Industriales'	Aprobado'
'Activo'	'Equilibrado'	'Sensorial'	'Equilibrado'	'Visual'	'Moderado'	'Secuencial'	'Moderado'	24	39	67	27	'Industriales'	Aprobado'
'Reflexivo'	'Moderado'	'Sensorial'	'Moderado'	'Visual'	'Equilibrado'	'Secuencial'	'Moderado'	36	39	65	26	'Industriales'	NP'
'Reflexivo'	'Moderado'	'Sensorial'	'Moderado'	'Visual'	'Fuerte'	'Secuencial'	'Equilibrado'	52	55	71	26	'Industriales'	Notable'
'Activo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Equilibrado'	'Global'	'Moderado'	31	54	81	27	'Industriales'	Aprobado'
'Reflexivo'	'Equilibrado'	'Sensorial'	'Equilibrado'	'Visual'	'Fuerte'	'Global'	'Equilibrado'	39	48	69	21	'Industriales'	Aprobado'
'Activo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Fuerte'	'Secuencial'	'Moderado'	60	54	69	24	'Industriales'	Aprobado'
'Activo'	'Fuerte'	'Sensorial'	'Moderado'	'Visual'	'Equilibrado'	'Secuencial'	'Equilibrado'	50	50	69	26	'Industriales'	Aprobado'
'Reflexivo'	'Fuerte'	'Intuitiva'	'Fuerte'	'Verbal'	'Moderado'	'Secuencial'	'Equilibrado'	34	59	81	26	'Industriales'	Aprobado'
'Activo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Fuerte'	'Secuencial'	'Equilibrado'	47	47	74	26	'Industriales'	Aprobado'
'Activo'	'Moderado'	'Sensorial'	'Moderado'	'Visual'	'Moderado'	'Global'	'Equilibrado'	54	37	58	22	'Industriales'	Suspense'
'Reflexivo'	'Equilibrado'	'Intuitiva'	'Equilibrado'	'Visual'	'Moderado'	'Secuencial'	'Moderado'	52	60	69	25	'Industriales'	Suspense'
'Reflexivo'	'Equilibrado'	'Sensorial'	'Equilibrado'	'Visual'	'Fuerte'	'Global'	'Moderado'	38	57	72	26	'Industriales'	Aprobado'
'Activo'	'Equilibrado'	'Sensorial'	'Equilibrado'	'Visual'	'Fuerte'	'Global'	'Moderado'	48	49	67	27	'Industriales'	Notable'
'Reflexivo'	'Equilibrado'	'Sensorial'	'Equilibrado'	'Visual'	'Fuerte'	'Global'	'Moderado'	57	54	66	24	'Industriales'	Suspense'
'Activo'	'Fuerte'	'Sensorial'	'Equilibrado'	'Visual'	'Moderado'	'Secuencial'	'Equilibrado'	33	55	76	26	'Industriales'	Aprobado'
'Activo'	'Moderado'	'Sensorial'	'Moderado'	'Visual'	'Fuerte'	'Global'	'Equilibrado'	54	48	70	26	'Industriales'	Aprobado'
'Reflexivo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Moderado'	'Secuencial'	'Equilibrado'	47	50	69	24	'Industriales'	Suspense'
'Activo'	'Moderado'	'Sensorial'	'Moderado'	'Visual'	'Moderado'	'Secuencial'	'Fuerte'	19	44	67	27	'Industriales'	Aprobado'
'Activo'	'Moderado'	'Intuitiva'	'Moderado'	'Visual'	'Fuerte'	'Secuencial'	'Equilibrado'	43	49	66	22	'Industriales'	Aprobado'
'Reflexivo'	'Moderado'	'Sensorial'	'Equilibrado'	'Visual'	'Fuerte'	'Secuencial'	'Equilibrado'	49	51	62	23	'Industriales'	Aprobado'
'Activo'	'Fuerte'	'Intuitiva'	'Fuerte'	'Visual'	'Moderado'	'Global'	'Equilibrado'	71	58	69	25	'Industriales'	Aprobado'
'Activo'	'Moderado'	'Sensorial'	'Moderado'	'Visual'	'Moderado'	'Secuencial'	'Equilibrado'	35	53	71	23	'Industriales'	Suspense'
'Activo'	'Equilibrado'	'Sensorial'	'Equilibrado'	'Visual'	'Equilibrado'	'Secuencial'	'Moderado'	34	45	69	25	'Industriales'	NP'
'Reflexivo'	'Equilibrado'	'Sensorial'	'Equilibrado'	'Visual'	'Fuerte'	'Secuencial'	'Moderado'	55	52	69	25	'Industriales'	Aprobado'
'Reflexivo'	'Moderado'	'Sensorial'	'Fuerte'	'Visual'	'Fuerte'	'Secuencial'	'Moderado'	47	58	73	21	'Industriales'	Aprobado'
'Activo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Fuerte'	'Secuencial'	'Equilibrado'	27	56	71	26	'Industriales'	Aprobado'
'Activo'	'Equilibrado'	'Sensorial'	'Equilibrado'	'Visual'	'Fuerte'	'Global'	'Equilibrado'	34	36	68	24	'Industriales'	Notable'
'Activo'	'Moderado'	'Sensorial'	'Equilibrado'	'Visual'	'Equilibrado'	'Secuencial'	'Moderado'	34	46	64	20	'Industriales'	Aprobado'
'Activo'	'Moderado'	'Sensorial'	'Equilibrado'	'Visual'	'Fuerte'	'Secuencial'	'Moderado'	44	48	64	25	'Industriales'	Notable'
'Activo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Fuerte'	'Global'	'Moderado'	52	46	56	29	'Industriales'	Aprobado'
'Activo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Moderado'	'Secuencial'	'Moderado'	37	46	62	26	'Industriales'	NP'
'Reflexivo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Fuerte'	'Global'	'Equilibrado'	55	57	65	30	'Master'	NC'
'Activo'	'Moderado'	'Sensorial'	'Moderado'	'Visual'	'Fuerte'	'Secuencial'	'Equilibrado'	57	47	60	25	'Industriales'	Aprobado'
'Reflexivo'	'Equilibrado'	'Intuitiva'	'Equilibrado'	'Visual'	'Moderado'	'Secuencial'	'Moderado'	35	50	61	28	'Industriales'	Suspense'
'Activo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Fuerte'	'Secuencial'	'Equilibrado'	42	41	60	25	'Industriales'	Suspense'
'Activo'	'Equilibrado'	'Sensorial'	'Equilibrado'	'Visual'	'Fuerte'	'Global'	'Equilibrado'	49	59	64	27	'Industriales'	Aprobado'
'Reflexivo'	'Equilibrado'	'Intuitiva'	'Equilibrado'	'Visual'	'Moderado'	'Global'	'Equilibrado'	71	57	68	24	'Industriales'	Aprobado'
'Activo'	'Moderado'	'Sensorial'	'Moderado'	'Verbal'	'Equilibrado'	'Secuencial'	'Equilibrado'	24	45	74	28	'Industriales'	NP'
'Reflexivo'	'Equilibrado'	'Sensorial'	'Moderado'	'Visual'	'Fuerte'	'Secuencial'	'Equilibrado'	40	42	62	26	'Industriales'	Suspense'
'Reflexivo'	'Equilibrado'	'Sensorial'	'Fuerte'	'Visual'	'Fuerte'	'Global'	'Equilibrado'	24	26	58	22	'Industriales'	Suspense'

## 7. Código Matlab del algoritmo fuzzy k-prototypes modificado

```
function [P,F,C,W] = fuzzykprototypes(data, textdata, k, w, alfa)

%Entradas:
% - data: matriz de datos cuantitativos
% - textdata: matriz de datos cualitativos
% - k: número de clusters
% - w: peso de los datos cualitativos
% - alfa: factor de borrosidad
%Salidas:
% - P: matriz de pertenencia
% - F: función de coste
% - C: matriz de centroides de los clusters
% - W: matriz de pertenencia borrosa

%Inicialización
iter=0;
converged=false;
lanz=0; %Se lanza el programa varias veces con el fin de verificar
        %si otra matriz aleatoria diera una distancia menor
[a,b]=size(textdata);
Ftot=0; %Se inicializa la distancia total de comparación
[m,s,data]=normalitzacio(data); %Se normalizan los datos numericos

%Matriz de pertenencia borrosa variable, inicialmente unitaria
for i=1:a
    for j=1:b
        for t=1:k
            mu{i,j,t}=1;
        end
    end
end

%Bucle
while true
    if (iter==0),
        %Crea la matriz de los centroides con k filas aleatorias
        C=createcluster(textdata,data,k);

        %Genera la función distancia entre los datos y los clusters
        D=distcluster(textdata,data,k,C,w,mu);

        %Crea la matriz W de pertenencia borrosa
        W=createfuzzy(textdata,data,k,C,D,alfa);

        %Comienza la la iteración
        iter=1;
        %Calcula la función coste
        F=costfunction(W,D,k,textdata,alfa);

    elseif ((iter==100)|(converged==true)), %Converge o llega a 100it
        lanz=lanz+1; %Se lanza el programa de nuevo
        iter=0; %Se inicializa el contador de las iteraciones
        converged=false; %Se inicializa la convergencia
        if Ftot==0 %Es la 1a vez que se lanza el programa
```

```

        Cant=Cnew; %Se guarda el cluster anterior para compararlo
        Ftot=Fnew;
    else
        if (Fnew<Ftot),
            Cant=Cnew; %El nuevo cluster es mínimo y se guarda
            Ftot=Fnew;
        end
    end
    if (lanz==40), %Finaliza el programa al lanzarlo 40 veces
        %Reescribe los datos numéricos originales (sin normalizar)
        C=originaldata(textdata,Cant,m,s);
        %Genera la función distancia entre los datos y las clases
        D=distcluster(textdata,data,k,Cant,w,mu);
        %Crea la matriz W de pertenencia borrosa
        W=createfuzzy(textdata,data,k,Cant,D,alfa);
        %Asigna los datos a los clusters más cercanos (mayor W)
        P=asigcluster(W);
        %Calcula la función coste
        F=costfunction(W,D,k,textdata,alfa);
        break;
    end
else
    %Step 2: Se recalcula el centroide Cnew siguiendo el método de
    %actualización de k-modes
    [Cnew,Z,Wz]=updatecluster(textdata,data,k,W,alfa);
    mu=matrizpertenencias(textdata,Z,W,Wz,k);
    %Se determina el valor de la función coste con el nuevo
    %centroide y para ello es necesario recalcular distancias
    Dnew=distcluster(textdata,data,k,Cnew,w,mu);
    Fnew=costfunction(W,Dnew,k,textdata,alfa);
    %Se busca minimizar F, si Fnew es igual a F, se para el
    %algoritmo
    if (abs(Fnew-F)<0.0001)
        converged=true;
    else
        %Step 3: Se determina la nueva matriz W a partir de Cnew
        Wnew=createfuzzy(textdata,data,k,Cnew,Dnew,alfa);
        %Se calcula la 2a nueva función de coste Fnew2
        Fnew2=costfunction(Wnew,Dnew,k,textdata,alfa);
        %Se busca minimizar F, si Fnew2 es igual a Fnew, se
        %para el algoritmo
        if (abs(Fnew-Fnew2)<0.0001)
            converged=true;
        else
            %Continua con la siguiente iteración
            W=Wnew;
            iter=iter+1;
            F=Fnew2; %Se modifica F con el de la iter anterior
        end
    end
end
end
end

```

---

```
function [m,s,P_norm] = normalitzacio(P)
```

```
%P: Matriz de datos, donde cada columna es una variable diferente. La
%normalización se realiza de cada variable por separado.
```

---

```

m=mean(P);
s=diag(cov(P))';
for i=1:size(P,1)
    for j=1:size(P,2)
        P_norm(i,j)=(P(i,j)-m(j))./sqrt(s(j));
    end;
end;

```

---

```

function [C]=createcluster(textdata, data, k)

%Determina el tamaño de la matriz de texto
[a,b]=size(textdata);

for i=1:k %se definen las k filas
    n=1;
    f=random('discrete uniform',a); %se elige una fila aleatoria
    for j=1:b
        if isequal(textdata{f,j}, ''), %se trata de atributos numéricos
            C{i,j}=data(f,n);
            n=n+1; %nueva numeración para la matriz numérica
        else %se trata de atributos cualitativos
            C{i,j}=textdata{f,j};
        end
    end
end
end

```

---

```

function [dist]=distcluster(textdata, data, k, C, w, mu)

[a,b]=size(textdata);

for j=1:k %se mide la distancia con los k clusters
    for i=1:a
        num(i)=0;
        cat(i)=0;
    end
    n=1;
    for t=1:b
        if isequal(textdata{1,t}, ''), % columna numérica
            for i=1:a %calcula parte de la distancia euclidia
                D{i,t,j}=abs(data(i,n)-C{j,t});
                valor=mu{i,t,j}*D{i,t,j};
                num(i)=num(i)+valor^2;
            end
            n=n+1;
        else %se trata de datos cualitativos
            for i=1:a %medida de disimilitud (con ponderación)
                if isequal(textdata{i,t},C{j,t}),
                    D{i,t,j}=0;
                else
                    D{i,t,j}=1;
                end
                cat(i)=cat(i)+mu{i,t,j}*w(t)*D{i,t,j}; %suma ponderada
            end
        end
    end
end
end

```

---

```
    for i=1:a
        dist{i,j}=cat(i)+sqrt(num(i));
    end
end
```

---

```
function W=createfuzzy(textdata,data,k,C,dist,alfa)
```

```
%Determina el tamaño de la matriz de texto
```

```
[a,b]=size(textdata);
coef=1/(alfa-1);
```

```
for j=1:a
    valor(j)=false;
    for i=1:k
        if isequal (dist{j,i},0), %coincide con algun cluster
            W{i,j}=1;
            valor(j)=true;
        else
            W{i,j}=0;
        end
    end
    if (valor(j)==false)
        for i=1:k
            frac=0;
            suma=0;
            for s=1:k
                frac=dist{j,i}/dist{j,s};
                frac=frac^coef;
                suma=suma+frac;
            end
            div=1/suma;
            W{i,j}=div;
        end
    end
end
```

---

```
function F=costfunction(W,D,k,textdata,alfa)
```

```
[a,b]=size(textdata);
```

```
%Multiplica las distancias de cada cluster por el coeficiente W^2
```

```
F=0;
for i=1:a
    for j=1:k
        coef=W{j,i};
        sum=coef^alfa*D{i,j};
        F=sum+F;
    end
end
```

---

```

function [Cnew,Z,Wz]=updatecluster(textdata,data,k,W,alfa)

%Determina el tamaño de la matriz de texto
[a,b]=size(textdata);

%Crea la matriz Z compuesta de los l diferentes valores que pueden
%tomar las t variables en las a observaciones realizadas (k-modes
%update method)
for t=1:b
    for i=1:a
        if i==1
            l=1;
            Z{l,t}=textdata{i,t};
        else
            flag=false;
            for n=1:l
                if isequal(textdata{i,t},Z{n,t}),
                    flag=true;
                end
            end
            if flag==false
                l=n+1;
                Z{l,t}=textdata{i,t};
            end
        end
    end
end

%Se define el sumatorio de factores de pertenencia correspondiente a
%Zlj
for j=1:k
    for t=1:b
        [x,y]=size(Z);
        l=0;
        for i=1:x
            if isequal(Z{i,t},''),
                break
            else
                l=l+1;
            end
        end
        for n=1:l
            Wz{n,t,j}=0;
            for i=1:a
                if isequal(textdata{i,t},Z{n,t}),
                    num=W{j,i};
                    Wz{n,t,j}=Wz{n,t,j}+num^alfa;
                end
            end
        end
    end
end

%Se consideran los datos cuantitativos
for j=1:k
    s=1;
    for t=1:b
        if isequal (textdata{1,t},''),
            z=0;

```

```

        Wtot=0;
        for i=1:a
            z=z+W{j,i}^alfa*data(i,s);
            Wtot=Wtot+W{j,i}^alfa;
        end
        Zc(j,s)=z/Wtot;
        s=s+1;
    end
end
end

%Se elige el Zlj cuyo sumatorio de los factores de pertenencia da un
%valor mayor y realiza el proceso inverso: obtener Cnew a partir de
%Zlj

[a,b2,c]=size(Wz);

for j=1:k
    s=1;
    for t=1:b2
        n=0;
        for i=1:a
            if isequal(Wz{i,t,j}, ''),
                if (n==0)
                    Cnew{j,t}=Zc(j,s);
                    s=s+1;
                end
                break
            else
                v(i)=Wz{i,t,j};
                n=n+1;
            end
        end
        if (n~=0)
            maximo=max(v(:));
            for i=1:n
                if maximo==v(i)
                    Cnew{j,t}=Z{i,t};
                end
            end
        end
    end
end
if (b>b2)
    for t=(b2+1):b
        Cnew{j,t}=Zc(j,s);
        s=s+1;
    end
end
end
end

```

---

```
function mu=matrizpertenencias(textdata,Z,W,Wz,k)
```

```
[a,b]=size(textdata);
[c,d]=size(Z);
```

```
%Se normaliza Wz
```

```

for t=1:k
    for j=1:d
        tot(j,t)=0;
        for l=1:c
            if isequal (Z{l,j}, ''),
                break;
            else
                tot(j,t)=tot(j,t)+Wz{l,j,t};
            end
        end
    end
end

for t=1:k
    for j=1:d
        for l=1:c
            if isequal (Z{l,j}, ''),
                break;
            else
                valor=Wz{l,j,t};
                Wz{l,j,t}=valor/tot(j,t);
            end
        end
    end
end

%Se normaliza W
for t=1:k
    sum(t)=0;
    for i=1:a
        sum(t)=sum(t)+W{t,i};
    end
end

for t=1:k
    for i=1:a
        W{t,i}=W{t,i}/sum(t);
    end
end

%Se implementa la matriz de pertenencias
n=1;
for j=1:b
    if isequal (textdata{1,j}, ''),
        for i=1:a
            for t=1:k
                mu{i,j,t}=W{t,i};
            end
        end
    else
        for i=1:a
            for l=1:c
                if isequal (textdata{i,j},Z{l,j}),
                    for t=1:k
                        mu{i,j,t}=Wz{l,j,t};
                    end
                    break;
                end
            end
        end
    end
end

```

```
        end
        n=n+1;
    end
end
```

---

```
function [C] = originaldata(textdata,Cant,m,s)

%A partir de la media y la desviación estándar, desnormaliza los datos

[a,b]=size(Cant);
l=1;

for j=1:b
    if isequal(textdata{1,j}, ''), %Se trata de datos numéricos
        for i=1:a
            C{i,j}=Cant{i,j}*sqrt(s(l))+m(l);
        end
        l=l+1;
    else
        for i=1:a
            C{i,j}=Cant{i,j};
        end
    end
end
```

## 8. Código Matlab del algoritmo de validación del número de clusters

```
function [S]=varianza_sugeno(textdata,data,D,C,Z,k,W,alfa,m,s)

[a,b]=size(textdata);

%Calculo de [C(i)-Z]^2=dif(i)
for i=1:k
    dif(i)=0;
    num=0;
    l=1;
    for j=1:b
        if isequal(textdata{1,j},''), %dato cuantitativo
            %Si se normalizan los datos, la media de Z es cero
            %Por tanto la diferencia es el dato_norm al cuadrado
            dato_norm=(C{i,j}-m(l))/sqrt(s(l));
            num=num+dato_norm^2;
            l=l+1;
        else
            if isequal(C{i,j},Z{1,j}),
                %si son iguales, la diferencia es nula
            else %se considera la unidad
                dif(i)=dif(i)+1;
            end
        end
    end
    dif(i)=dif(i)+sqrt(num);
end

S=0;
for i=1:k
    for j=1:a
        coef=W{i,j};
        resta=D{j,i}-dif(i);
        sum=coef^alfa*resta;
        S=S+sum;
    end
end
```

---

```
function [Z]=media_datos(textdata,data)

[a,b]=size(textdata)

m=1;
for j=1:b
    if isequal(textdata{1,j},''), %Se trata de datos cuantitativos
        v=0;
        for i=1:a
            v=v+data(i,m);
        end
        Z{1,j}=v/a; %Calcula la media de los datos cuantitativos
        m=m+1;
    else %busca el valor cualitativo que más se repite
```

```
for i=1:a %determina los distintos valores y sus repeticiones
    if i==1
        l=1;
        c{1}=textdata(i,j);
        rep(l)=1;
    else
        flag=false;
        for n=1:l
            if isequal(textdata(i,j),c{n}),
                rep(n)=rep(n)+1;
                flag=true;
            end
        end
        if flag==false
            l=n+1;
            c{1}=textdata(i,j);
            rep(l)=1;
        end
    end
end
valor=max(rep()); %Elige el máximo número de repeticiones
for i=1:l
    if(rep(i)==valor)
        Z{1,j}=c{i}; %Asigna el valor cualitativo
    end
end
end
end
```

## 9. Código Matlab del pre-procesado de los datos de los alumnos

```
function [S,F]=algoritmo()

IEA=cuestionarioIEA();
[CMA,textdataCMA]=cuestionarioCMA();
[RSES, textdataRSES]=cuestionarioRSES();

%Se configura el conjunto de datos a partir de los extraídos de los 3
%cuestionarios (CMA,IEA,RSES), ordenados y convertidos los datos del
%cuestionario IEA a datos cualitativos
[textdata,data]=conversion(IEA,CMA,textdataCMA,RSES,textdataRSES);

%Se añaden las calificaciones cualitativas obtenidas
clases_cat=clases_alumnos_cat()

[a,b]=size(textdata);
for i=1:a
    textdata{i,b+1}=clases_cat{i,1};
end

alfa=1.1;
k=2;
for i=1:15
    w(i)=1;
end

%Se aplica el algoritmo de agrupación fuzzykprototypes modificado
[P,F,C,W,D,m,s] = fuzzykprototypes(data, textdata, k, w, alfa);

%Se calcula la media para el índice de validación de clusters
Z=media_datos(textdata,data);
S=varianza_sugeno(textdata,data,D,C,Z,k,W,alfa,m,s);

%Se convierten los datos de los centroides C de numéricos a
%cualitativos
[Cfinal]=conversionfinal(data,textdata,C);
```

---

```
function [textdata,
num]=conversion(IEA,CMA,textdataCMA,RSES,textdataRSES)
```

```
%textdata:
% 1: email
% 2: Pr: procesamiento (activo/reflexivo)
% 3: Cluster: Equilibrado (1-3), Moderado (5-7), Fuerte (9-11)
% 4: Pe: percepcion (sensorial/intuitiva)
% 5: Cluster: Equilibrado (1-3), Moderado (5-7), Fuerte (9-11)
% 6: E: entrada (visual/verbal)
% 7: Cluster: Equilibrado (1-3), Moderado (5-7), Fuerte (9-11)
% 8: C: comprension (secuencial/global)
% 9: Cluster: Equilibrado (1-3), Moderado (5-7), Fuerte (9-11)
% 10: Rs: motivacion extrinseca (Dominante/Pasivo)
% 11: Ed: motivacion intrinseca (Dominante/Pasivo)
```

```

% 12: In: valor colaborativo (Activo/Mixto/Pasivo)
% 13: A: autoestima (Alta/Normal/Baja)
% 14: Procedencia (Master/Industriales)

[datos]=cuestionarios(IEA,CMA,txtdataCMA,RSES,txtdataRSES);

[a,b]=size(datos);

for i=1:a
    txtdata{i,1}=datos{i,1}; %1: email
    %Rs, Ed, In y la Autoestima se realiza el cluster cuantitativo y
    %con las clasificaciones hechas, se convierten a datos
    %cualitativos.
    txtdata{i,10}='';
    num(i,1)=datos{i,6}; %10: Rs
    txtdata{i,11}='';
    num(i,2)=datos{i,7}; %11: Ed
    txtdata{i,12}='';
    num(i,3)=datos{i,8}; %12: In
    txtdata{i,13}='';
    num(i,4)=datos{i,9}; %13: Autoestima
    txtdata{i,14}=datos{i,10}; %14: Procedencia
end

for i=1:a
    if isequal (datos{i,2}, '');
        v(i)=0;
    else
        if (datos{i,2}>0) ; %2: procesamiento
            v(i)=datos{i,2};
            txtdata{i,2}='Activo';
        else
            v(i)=datos{i,2};
            v(i)=abs(v(i));
            txtdata{i,2}='Reflexivo';
        end
    end
end

[vc]=IEAclustering(v); %3: Se determina el cluster del procesamiento

for i=1:a
    txtdata{i,3}=vc{i};
end

for i=1:a
    if isequal (datos{i,3}, '');
        v(i)=0;
    else
        if (datos{i,3}>0) ; %4: percepcion
            v(i)=datos{i,3};
            txtdata{i,4}='Sensorial';
        else
            v(i)=datos{i,2};
            v(i)=abs(v(i));
            txtdata{i,4}='Intuitiva';
        end
    end
end
end

```

```

[vc]=IEAclustering(v); %5: Se determina el cluster de la percepcion

for i=1:a
    textdata{i,5}=vc{i};
end

for i=1:a
    if isequal (datos{i,4}, '');
        v(i)=0;
    else
        if (datos{i,4}>0) ; %6: entrada
            v(i)=datos{i,4};
            textdata{i,6}='Visual';
        else
            v(i)=datos{i,4};
            v(i)=abs(v(i));
            textdata{i,6}='Verbal';
        end
    end
end

[vc]=IEAclustering(v); %7: Se determina el cluster de la entrada

for i=1:a
    textdata{i,7}=vc{i};
end

for i=1:a
    if isequal (datos{i,5}, '');
        v(i)=0;
    else
        if (datos{i,5}>0) ; %8: comprensión
            v(i)=datos{i,5};
            textdata{i,8}='Secuencial';
        else
            v(i)=datos{i,5};
            v(i)=abs(v(i));
            textdata{i,8}='Global';
        end
    end
end

[vc]=IEAclustering(v); %9: Se determina el cluster de la comprensión

for i=1:a
    textdata{i,9}=vc{i};
end

```

---

```

function [Cfinal]=conversionfinal(data,textdata,C)

```

```

[a,b]=size(textdata);
[e,f]=size(C);
m=1;

```

```

%Se cogen los valores iniciales (sin la normalizacion de media y
varianza)
for i=1:e

```

```
for j=1:a
    if isequal (C{i,1},textdata{j,1}),
        for k=1:f
            if isequal (textdata{j,k},''), %datos cuantitativos
                Cfinal{i,k}=data(j,m);
                m=m+1;
            else
                Cfinal{i,k}=textdata{j,k}; %datos cualitativos
            end
        end
        end
        m=1;
    end
end
end

%Se convierten los factores Rs, Ed e In cuantitativos a cualitativos:
%Rs: Dominante (45<Rs<84), Pasivo
%Ed: Dominante (48<Ed<66), Pasivo
%In: Activo (77<In<84), Mixto (56<In<77), Pasivo
%RSES: Alto (36<RSES<40), Normal (28<RSES<36), Bajo

for i=1:e
    if (Cfinal{i,10}>44), %Motivación Extrínseca
        Cfinal{i,10}='Rs_Dominante';
    else
        Cfinal{i,10}='Rs_Pasivo';
    end
end

for i=1:e
    if (Cfinal{i,11}>47), %Motivación Intrínseca
        Cfinal{i,11}='Ed_Dominante';
    else
        Cfinal{i,11}='Ed_Pasivo';
    end
end

for i=1:e
    if (Cfinal{i,12}>76), %Colaboración
        Cfinal{i,12}='In_Activo';
    else
        if (Cfinal{i,12}>55),
            Cfinal{i,12}='In_Mixto';
        else
            Cfinal{i,12}='In_Pasivo';
        end
    end
end

for i=1:e
    if (Cfinal{i,13}>35), %Autoconfianza o Autoestima
        Cfinal{i,13}='Conf_Alta';
    else
        if (Cfinal{i,13}>27),
            Cfinal{i,13}='Conf_Normal';
        else
            Cfinal{i,13}='Conf_Baja';
        end
    end
end
```

# Bibliografía

- [1] Ong, J. y Ramachandran, S. Intelligent Tutoring Systems: Using AI to Improve Training Performance and ROI. *Slottler Henke Associates, Inc.* 2003.
- [2] Romero, C., Ventura, S. y De Castro, C., Hall, W. y Hong, M. Using Genetic Algorithms for Data Mining in Web-based Educational Hypermedia Systems. *Adaptative Hypermedia 2002, Workshop on Adaptive Systems for Web-based Education*, pp. 137-142, 2002.
- [3] Benítez, I. J. Trabajo de investigación. Técnicas de agrupamiento para el análisis de datos cuantitativos y cualitativos. *Universidad Politécnica de Valencia*, 2005.
- [4] Romero, C. y Ventura, S. Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33, pp. 135-146, 2007.
- [5] Klossgen, W. y Zytkow, J. Handbook of data mining and knowledge discovery. *Oxford University Press*. 2002.
- [6] Shoham, Y. y Leyton-Brown, K. Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations, *Cambridge University Press*, 2008.
- [7] Hernández, J., Ramírez, M. J. y Ferri, C. Introducción a la minería de datos, *Pearson Educación, S. A.*, 2004.
- [8] Pal, S. K., Talwar, V. y Mitra, P. Web Mining in Soft Computing Framework: Relevance, State of the Art and Future Directions, *IEEE Transactions on Neural Networks*, vol. 13, nº 5, pp. 1163-1177, 2002.
- [9] Talavera, L, Gaudioso, E. Mining student data to characterize similar behaviour groups in unstructured collaboration spaces, *Workshop on Artificial Intelligence in CSCL*, pp. 17-23, 2004.
- [10] Hamalainen, W., suhonen, J. Sutinen, E. y toivonen, H. Data mining in personalizing distance education courses. *World conference on open learning and distance education*, 2004.
- [11] Ueno, M. Online Outlier Detection System for Learning Time Data in E-Learning and ITS Evaluation, *Proceedings of Computers and Advanced Technology in Education*, pp. 248-253, 2004.
- [12] Arrollo, I., Murray, T., Wooldf, B. P., Beal, C. R. Inferring Unobservable Learning Variables from Students' Help Seeking Behaviour, *Conference on Intelligent Tutoring Systems*, pp. 782-784, 2004.
- [13] Jing, Y., Zhong, S., Li, X. y Cheng, X. Using Instruction Strategy for a Web-Based Intelligent Tutoring System, *Technologies for E-Learning and Digital Entertainment*, vol. 3942, pp. 132-139, 2006.
- [14] Yoo, J., Yoo, S., Lance, C. y Hankins, J. Student Progress Monitoring Tool Using Treeview, *SIGCSE'06*, pp. 373-377, 2006.
- [15] Cocea, M. y Wibelzahl, S. Can Log Files Analysis Estimate Learners' Level of Motivation? *Proceedings of the workshop week Lernen*, pp. 32-35, 2006.
- [16] Sterbini, A. y Temperini, M. Adaptive Construction and Delivery of Web-Based Learning Paths, *39th ASEE/IEEE Frontiers in Education Conference*, sesión W1C-1, 2009.
- [17] Shen, R., Han, P. y Yang, F. Data Mining and Case-based Reasoning for Distance Learning. *International Journal of Distance Education Technologies*, vol. 1 (3), pp. 46-58, 2003.
- [18] Heraud, J. M. A Web Adaptive Learning Environment where the link structure is generated by experience reuse. *Conference on Intelligent Tutoring Systems*, 2004.

- [19] Huang, M. J., Huang, H. S. y Chen, M. Y. Constructing a personalized e-learning system based on genetic algorithm and case-based reasoning approach, *Expert Systems with Applications*, vol. 33, pp. 551-564, 2007.
- [20] Alves, P., Amaral, L. y Pires, J. Case-based Reasoning Approach to Adaptive Web-Based Educational Systems, *8th IEEE International Conference on Advanced Learning Technologies*, pp. 260-261, 2008.
- [21] Yeh, S. W. y Lo, J. J. Assessing metacognitive knowledge in Web-based CALL: a neural network approach, *Computers & Education*, vol. 44, pp. 97-113, 2005.
- [22] Dorça, F., Lopes, C., Fernandes, M. y Lopes, R. Adaptativity Supported by Neural Networks in Web-based Educational Systems, *Journal of Education, Informatics and Cybernetics*, 2008.
- [23] Minaei-Bidgoli y B., Punch, W. F. Predicting student performance: an application of data mining methods with the educational web-based system LON-CAPA. *IEEE Frontiers in Education*, pp. 1-9, 2003.
- [24] Minaei-Bidgoli y B., Punch, W. F. Using Genetic Algorithms for Data Mining Optimization in an Educational Web-Based System. *GECCO*, pp. 2252-2263, 2003.
- [25] Romero, C., Ventura, S. y de Bra, P. Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Author. *User Modeling and User Adapted Interaction*, vol. 14, nº 5, 2005.
- [26] Chen, C. M. y Duh, L. J. Personalized web-based tutoring system based on fuzzy item response theory, *Expert Systems with Applications*, vol. 34, pp. 2298-2315, 2008.
- [27] Huang, C. J., Liu, M.C., Chu, S. S. y Cheng, C.L. An intelligent learning diagnosis system for Web-based thematic learning platform, *Computers & Education*, vol. 48, pp. 658-679, 2007.
- [28] Hwang, K. A. y Yang, C. H. Assessment of effective state in distance learning based on image detection by using fuzzy fusion, *Knowledge-Based Systems*, 2009.
- [29] Huang, C.J., Liu, M.C., Chu, S.S. y Cheng, C.L. An intelligent learning diagnosis system for Web-based thematic learning platform, *Computer & Education*, 48, pp. 658-679, 2007.
- [30] Piramuthu, S. Knowledge-based web-enabled agents and intelligent tutoring systems, *IEEE Transactions on Educations*, vol. 48, pp. 750-756, 2005.
- [31] Mohammed, P. A cooperative multi-agent approach in support of learning object recommendations, *Proceedings of the 9th WSEAS International Conference on Data Networks, Communications, Computers*, pp. 424-430, 2007.
- [32] Ovalle, D. A., Jiménez, J. y Viccari, R. M. Sistemas de Enseñanza/Aprendizaje basados en Agentes Inteligentes Pedagógicos, *Avances en Sistemas e Informática 2*, vol. 2, pp. 17-26, 2005.
- [33] Nkambou, R., Fournier-Viger, P. y Mephu, E. Improving the Behavior of Intelligent Tutoring Agents with Data Mining, *IEEE Intelligent Systems*, pp. 46-53, 2009.
- [34] Kay, J., Maisonneuve, N., Yacef, K. y Zaiane, O. R. Mining patterns of events in student's teamwork data, *Proceedings of educational data mining workshop*, pp. 1-8, 2006.
- [35] Romero, C., Porras, A. R., Ventura, S., Hervás, C. y Zafra, A. Using sequential pattern mining for links recommendation in adaptive hipermedia educational systems, *Current Developments in Technology-Assisted Education*, pp. 1016-1020, 2006.

- [36] Zaïne, O. R. Building a Recommender Agent for e-learning Systems, International Conference on computers in Education, pp. 55-59, 2002.
- [37] Wang, F. On Analysis and Modeling of Student Browsing Behavior in Web-Based Asynchronous Learning Environments, Conference on Web-based Learning, pp. 69-80, 2002.
- [38] Freyberger, J., Heffernan, N. T., Ruiz, C. Using Association Rules to Guide a Search for Best Fitting Transfer Models of Student Learning, Conference on Intelligent Tutoring Systems, 2004.
- [39] Yu, P., Own, C., Lin, L. On the Learning Behaviour Analysis of Web Based Interactive Environment, ICCE, 2001.
- [40] Popescu, E. Dynamic adaptive hypermedia systems for e-learning. Tesis doctoral de la Universidad de Craiova (Rumania), 2008.
- [41] Huang, Y. M. y Liu, C. H. Applying adaptive swarm intelligence technology with structuration in web-based collaborative learning, Computers & Education, n° 52, pp. 789-799, 2009.
- [42] Perera, D., Kay, J., Koprinska, I., Yacef, K. y Zaïane, O. R. Clustering and Sequential Pattern Mining of Online Collaborative Learning Data, IEEE Transactions on Knowledge Data Engineering, vol. 21, n° 6, pp. 759-772, 2009.
- [43] Dogan, B. y Camurcu, Y. Visual Clustering of Multidimensional Educational Data From an Intelligent Tutoring System, Wiley Periodicals Inc., 2009.
- [44] Popescu, E. An Artificial Intelligence Course Used to Investigate Students' Learning Style, Lecture Notes in Computer Science, vol. 5145, pp. 122-131, 2008.
- [45] Felder, R. y Silverman, L. Learning and Teaching Styles in Engineering Education. Engineering Education, 78 (7), pp. 674-681, 1988.
- [46] Baldiris, S., Santos, O. C., Barrera, C., Boticario, J. G., Velez, J. y Fabregat, R. Integration of Educational Specifications and Standards to Support Adaptative Learning Scenarios in ADAPTAPlan, International Journal of Computer & Applications, vol. 5, n° 1, pp. 88-107, 2008.
- [47] Felder, R. M. y Soloman, B. A. Index of Learning Styles, <http://www.engr.ncsu.edu/learningstyles/ilsweb.html>
- [48] Cha, H. J., Kim, Y. S., Park, S. H., Yoon, T. B., Jung, Y. M. y Lee, J. H. Learning Styles Diagnosis based on User Interface Behaviors for the Customization of Learning Interfaces in an Intelligent Tutoring System, Lecture Notes in Computer Science, 2006.
- [49] Rishi, O., Govil, R. y Sinha, M. Distributed Case Based Reasoning for Intelligent Tutoring System: An Agent Based Student Modeling Paradigm, Proceedings of World Academy of Science, Engineering and Technology, vol. 23, pp. 273-276, 2007.
- [50] García, P., Amandi, A., Schiaffino, S. y Campo, M. Evaluating Bayesian networks' precision for detecting students' learning styles, Computers & Education, vol. 49, pp. 794-808, 2007.
- [51] Bloom, B. S. Taxonomy of Educational Objectives. New York: David Mckay, 1956.
- [52] Paquette, G., Léonard, M., Lundgren-Cayrol, K., Mihaila, S. y Gareau, D. Learning Design based on Graphical Knowledge-Modelling, Educational Technology & Society, 9 (1), pp. 97-112, 2006
- [53] Benmohamed, H., Leleve, A. y Prevot, P. Remote Laboratories: New Technology and Standard Based Architecture. International Conference on Information and Communication Technologies from Theory to Applications, pp. 101-102, 2004.

- [54] Sánchez, J., Morilla, F., Dormino, S., Aranda J. y Ruipérez, P. Virtual control lab using Java and Matlab: A qualitative approach. *IEEE Control Systems Magazine*, pp. 8-20, 2002.
- [55] Casini, M., Prattichizzo, D. y Vicino, A. E-Learning by Remote Laboratories: A New Tool for Control Education. *6th IFAC Symposium on Advances in Control Education*, 2003.
- [56] Liou, S., Soelaeman, H. y Leung, P. Distance Learning Power Engineering Laboratory. *IEEE Computer Applications in Power*, 1999.
- [57] Imbrie, P., K. y Raghavan, S. Work in Progress – A Remote e-Laboratory for Student Investigation, Manipulation and Learning. *35th ASEE/IEEE Frontiers in Education Conference*, 2005.
- [58] Viedma, G., Dancy, I. Y Lundberg, K. Web-based linear-systems iLab. *American Control Conference*, 2005.
- [59] Ko, C. C., Chen, B. M., Chen, J., Zhuang, Y. y Tan, K. C. Development of a Web-Based Laboratory for Control Experiments on a Coupled Tank Apparatus. *IEEE Transactions on Education*, vol. 44, pp. 76-86, 2001.
- [60] Tzafestas, C. S., Paliologou, N. Y Alifragis, M. Virtual and Remote Robotic Laboratory: Comparative Experimental Evaluation. *IEEE Transactions on Education*, vol. 49, nº 3, pp. 360-369, 2006.
- [61] Leleve, A., Benmohamed, H., Prevot, P. y Meyer, C. Remote Laboratory – Towards an integrated training system. *Proceedings of 4th International Conference on Information Technology Based Higher Education and Training*, 2003.
- [62] Callaghan, M., Harkin, J., McGinnity, T. M. y Maguire, L. P. Intelligent User Support in Autonomous Remote Experimentation Environments. *IEEE Transactions on Industrial Electronics*, vol. 55, nº 6, 2008.
- [63] Clause, A., Spanias, A., Xavier, A. y Tampi, M. A Java signal analysis tool for signal processing experiments. *Acoustics Speech and Signal Processing, ICASSP*, vol. 3, pp. 1849-1852, 1998.
- [64] Qingcang, Y. U., Chen, B. y Cheng, H. H. Web based control system design and analysis. *Control System Magazine, IEEE*, vol. 24, pp. 45-57, 2004.
- [65] Exel, M., Gentil, S. Michau, F. y Rey, D. Simulation workshop and remote laboratory: two Web-based training approaches for control. *Proceedings American Control Conference*, vol. 5, pp. 3468-3472, 2000.
- [66] Sucar, E., Noguez, J., Huesca, G. y Rodríguez, E. A Semi-Open Learning Environment for Mobile Robotics. *International Journal of Online Engineering*, 2007.
- [67] García, L. ¿Web 2.0 vs Web 1.0?, Editorial del BENED, 2007.
- [68] Safran, C., Helic, D. y Gütl, C. E-Learning practices and Web 2.0, *Conference ICL2007*, 2007.
- [69] Jarabo, J. Trabajo de Investigación ¿Han llegado los nativos digitales a la Universitat Jaume I?, *Universitat Jaume I de Castellón*, 2010.
- [70] Ferreti, S., Mirri, S., Muratori, L. A., Rocchetti, M. y Salomoni, P. E-learning 2.0: you are We-LCoME!, *Proceedings of the 2008 international cross-disciplinary conference on Web accessibility (W4A)*, vol. 317, pp. 116-125, 2008.
- [71] Pea, R. y Lindgren, R. Video Collaboratories for Research and Education: An Analysis of Collaboration Design Patterns, *IEEE Transactions on Learning Technologies*, vol. 1, nº 4, 2008.

- [72] Vargas, H., Sánchez, J. y Dormido, S. Experimentación Remota en Automática: Nuevos Entornos Basados en la Web 2.0, XXX Jornadas de Automática, pp. 37-43, 2009.
- [73] Pérez, R., Caeiro, M. y Anido, L. Adding Process-Driven Collaboration Support in Moodle, 39th ASEE/IEEE Frontiers in Education Conference, sesión W2A, 2009.
- [74] Bittencourt, I. I., Costa, E., Silva, M. y Soares, E. A Computational Model for Developing Semantic Web-based Educational Systems, Knowledge-Based Systems, 2009.
- [75] Lin, F., Holt, P., Leung, S., Hogeboom, M. y Cao, Y. A Multi-Agent and Service-Oriented Architecture for Developing Integrated and Intelligent Web-based Education Systems, Proceedings of International Workshop on Semantic Web Technologies for E-learning, pp. 11-20, 2005.
- [76] Millard, D. E., Koody, K., Davis, H. C., Gilbert, L., Howard, Y., Tao, F. y Wills, G. (Semantic Web) Services for e-Learning, International Journal of Knowledge and Learning, vol. 4, nº 2/3, pp. 298-315, 2008.
- [77] Dietze, S., Gugliotta, A. y Domingue, J. A Semantic Web Service oriented Framework for adaptive Learning Environments, The Semantic Web: Research and Applications, vol. 4519, pp. 701-715, 2007.
- [78] Mikic, F. A., Burguillo, J. C., Llamas M. y Fernández, D. Using Semantics in INES, an Intelligent Educational System. 39th ASEE/IEEE Frontiers in Education Conference, sesión M1G, pp. 1-6, 2009.
- [79] Monk, D. Using Data Mining for e-Learning Decision Making, Electronic Journal of e-Learning, vol. 3(1), pp. 41-54, 2005.
- [80] Wolpers, M., Najjar, J., Verbert, K. y Duval, E. Tracking Actual Usage: the Attention Metadata Approach, Educational Technology & Society, vol. 10 (3), pp. 106-121, 2007.
- [81] Barba-Romero, S. Panorámica Actual de la Decisión Multicriterio Discreta, Investigaciones Económicas (segunda época), vol. 11, nº 2, pp. 279-308, 1987.
- [82] Boneu, J. M. Plataformas abiertas de e-learning para el soporte de contenidos educativos abiertos, Revista de Universidad y Sociedad del Conocimiento, vol. 4, nº 1, 2007.
- [83] Harrigan, M., Kravcik, M., Steiner, C. y Wade, V. GRAPPLE – D9.1: Requirements Specification by Academic Users on Integrated Adaptive Learning Services, LMedia: Deliverables and Milestones, 2009.
- [84] Burgos, D., Tattersall, C. y Dougiamas, M. Mapping IMS Learning Design and Moodle. A first understanding, Proceedings of Simposio Internacional de Informática Educativa (SIIE06), 2006.
- [85] González, R., Valle, A., Núñez, J. C. y González-Pienda, J. A. Una aproximación teórica al concepto de metas académicas y su relación con la motivación escolar, Psicothema, vol. 8, nº 1, pp. 45-61, 1996.
- [86] Sanz de Acedo, M., Ugarte, M. D. y Lumbreras, M. V. Desarrollo y Validación de un Cuestionario de Metas para Adolescentes, Psicothema, vol.15, nº 3, pp. 493-499, 2003.
- [87] Martín-Albo, J., Núñez, J. L. y Grijalvo, F. The Rosenberg Self-Esteem Scale: Translation and Validation in University Students. *The Spanish Journal of Psychology*, vol. 10, nº 2, pp. 458-467, 2007.
- [88] García, M. S., González-Pienda, J. A., Núñez, J.C., González-Pumariiega, S., Álvarez, L., Roces, C., González, R. y Valle, A. El Cuestionario de Metas Académicas (C.M.A.) Un instrumento para la evaluación de la orientación

- motivacional de los alumnos de educación secundaria, *Aula abierta*, nº 71, pp. 175-199, 1998.
- [89] Perkins, D. *La escuela inteligente*, Ed. Gedisa, 1995.
- [90] Amela, V., Díez J.-L. y Vallés, M. Un Nuevo Enfoque para el Diseño de Sistemas Tutores Inteligentes con Aprendizaje, VI Jornadas CEA de Enseñanza a través de Internet-Web de la Ingeniería de Sistemas y Automática, 2010.
- [91] Hernández, J., Ramírez, M.J. y Ferri, C. *Introducción a la Minería de Datos*, Pearson Educación S. A., 2004.
- [92] Li, C. y Biswas, G. Unsupervised learning with mixed numeric and nominal data, *IEEE Transactions on Knowledge and Data Engineering* nº 14 (4), pp. 673-690, 2002.
- [93] Huang, J. Z., Ng, M. K., Rong, H. y Li, Z. Automated variable weighting in k-mean type clustering, *IEEE Transactions on PAMI* nº 27 (5), 2005.
- [94] Chen, N., Chen, A. y Zhou, L-X. Fuzzy *k*-Prototypes Algorithm for Clustering Mixed Numeric and Categorical Valued Data, *Journal of Software*, Vol. 12, nº 8, pp. 1107-1113, 2001.
- [95] Ahmad, A. y Dey, L. A *k*-mean clustering algorithm for mixed numeric and categorical data, *Data & Knowledge Engineering*, nº 63, pp. 503-527, 2007.
- [96] Luo, H., Kong, F. y Li, Y. Clustering mixed data based on evidence accumulation, *Lecture Notes on Artificial Intelligence* 4093, 2006.
- [97] Zhao, Y., Li, B., Li, X., Liu, W. H. y Ren, S. J. Cluster ensemble method for databases with mixed numeric and categorical values. *Journal of Tsinghua University (Science and Technology)*, vol. 46, nº 10, pp. 1673-1676, 2006.
- [98] Stanfill, C. y Waltz, D. Toward memory based reasoning, *Communication of the ACM* 29 (12), pp. 1213-1228, 1986.
- [99] Sugeno, M. y Yasukawa, T. A fuzzy-logic based approach to qualitative modeling, *IEEE Transactions on Fuzzy Systems*, nº 1, 1993.
- [100] Mateu E., Casal J. Tamaño de la muestra. *Rev. Epidem. Med. Prev.* nº 1, pp. 8-14, 2003.