



PROCESO DE IDENTIFICACIÓN DE COMPORTAMIENTO DE COMUNIDADES EDUCATIVAS BASADO EN RESULTADOS ACADÉMICOS

Tesista

Ing. Pablo CIGLIUTI

Directores

Dr. Ramón GARCÍA MARTÍNEZ (UNLa-UNLP) Ma. Florencia POLLO CATTANEO (UTN-FRBA)

Director

CoDirectora

TESIS PRESENTADA PARA OBTENER EL GRADO
DE
MAGISTER EN TECNOLOGÍA INFORMÁTICA APLICADA EN EDUCACIÓN

**FACULTAD DE INFORMÁTICA
UNIVERSIDAD NACIONAL DE LA PLATA**

JUNIO, 2015

RESUMEN

Los procesos de explotación de información se incorporan al ámbito educativo para ayudar a entender y mejorar tanto la enseñanza de los docentes como el aprendizaje de los alumnos. Entre estas cuestiones se destaca el análisis del comportamiento de comunidades educativas de forma tal de proveer al docente herramientas que ayuden a mejorar la enseñanza/aprendizaje. En este contexto, el presente trabajo de tesis tiene como objetivo proponer, estudiar y validar un proceso de explotación de información que permita identificar el comportamiento de comunidades educativas basado en resultados académicos.

ABSTRACT

The data mining processes are incorporated to the educational environment to help to understand and improve the way teachers teach, and the way students learn. These issues stands behavioral analysis of learning communities in such a way to provide teachers tools to help improve understanding of the processes of teaching and learning. In this context, this thesis aims to propose, study and validate a data mining process to identify the behavior of educational communities, based on academic results.

DEDICATORIA

A mis padres, esposa e hijos por alentarme y apoyarme...

A Paola y Ramiro por ser grandes compañeros en mi carrera académica...

A Edgardo Claverie quien fue mi mentor académico...

AGRADECIMIENTOS

A la Facultad de Informática de la Universidad Nacional de la Plata por acogerme con generosidad de “alma mater” para que pudiera llevar a cabo mis estudios de Magister en Tecnología Informática Aplicada a la Educación.

A la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional por brindar mi formación profesional de Grado y permitirme desarrollar como docente.

Al Grupo de Investigación en Sistemas de Información del Departamento de Desarrollo Productivo y Tecnológico de la Universidad Nacional de Lanús por recibirme para desarrollar las investigaciones de esta tesis, proveyendo un ambiente estimulante para mi formación.

Al Grupo de Estudio en Metodologías de Ingeniería de Software de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional por el soporte brindado a mi trabajo de tesis.

A mis directores de tesis: Dr. Ramón García-Martínez, y Ma. Florencia Pollo-Cattaneo por su gran guía, asistencia e insistencia en la elaboración de este trabajo de tesis.

A mis compañeros docentes de la tecnológica, en particular a la Lic. Susana Granado Peralta, Lic. María Alicia Piñeiro y Lic. Karina Cuzzani por su apoyo y contribución incondicional.

Al Ing. Andrés Burzryn y equipo de departamento de sistemas por cederme un espacio y marco de trabajo para esta tesis.

Al Lic. Sebastian Martins quien generosamente me ayudó a adaptar el proceso propuesto por él al dominio de la educación y colaboró como "dataminer" en la búsqueda de patrones en los datos disponibles.

A la ingeniera Paola Saavedra-Martinez por facilitarme los datos de sus propias investigaciones.

A todos los que me ayudaron a llegar aquí...

1. INTRODUCCIÓN	1
1.1. Contexto General de la Tesis	1
1.1.1. Educación a Distancia	1
1.1.2. Explotación de Información	3
1.1.3. Explotación de Información Educativa	5
1.2. Síntesis del Trabajo Realizado	8
1.3. Estructura de la Tesis	8
1.4. Producción Científica	9
2. ESTADO DE LA CUESTIÓN	11
2.1. Introducción General a la Explotación de Información	11
2.2. Procesos de Explotación de Información	12
2.3. Explotación de Información en el Ámbito Educativo	19
2.4. Casos de Estudio	30
2.4.1. Ámbito Internacional	30
2.4.2. Ámbito Nacional	39
3. PROBLEMA	47
3.1. Generalidades del Problema	47
3.2. Preguntas de Investigación	51
4. SOLUCIÓN	53
4.1. Marco de la Solución	53
4.2. Solución Propuesta	54
5. VALIDACIÓN	61
5.1. Identificación de Comportamiento de Estudiantes Recursantes Utilizando EVEAs	61
5.1.1. Descripción del Caso	61
5.1.2. Aplicación del Proceso	67
5.1.3. Conclusiones del Caso de Validación	76
5.2. Identificación de Conceptos Críticos de una Cursada	77
5.2.1. Descripción del Caso	77
5.2.2. Aplicación del Proceso	78
5.2.3. Conclusiones del Caso de Validación	86
5.3. Identificación de Contenidos Débilmente Apropriados por Estudiantes	86
5.3.1. Descripción del Caso	86
5.3.2. Aplicación del Proceso	87
5.3.3. Conclusiones del Caso de Validación	96

6. CONCLUSIONES	99
6.1. Aportaciones de la Tesis	99
6.2. Futuras Líneas de Investigación	100
7. REFERENCIAS	103

ÍNDICE DE FIGURAS

Figura 2.1.	Proceso de Descubrimiento de Reglas de Comportamiento.	13
Figura 2.2.	Proceso de Descubrimiento de Grupos.	14
Figura 2.3.	Proceso de Ponderación de Interdependencia de Atributos.	15
Figura 2.4.	Proceso de Descubrimiento de Reglas Pertencientes a Grupos.	16
Figura 2.5.	Proceso de Ponderación de Reglas de Comportamiento o de Pertenencia a Grupos.	18
Figura 2.6.	El Ciclo de la Aplicación de la Minería de Datos en los Sistemas Educativos.	21
Figura 2.7.	Publicaciones por Tipo de Técnica de Minería de Datos Utilizadas.	29
Figura 2.8.	Cantidad de papers publicados hasta el 2009 agrupados por año.	30
Figura 2.9.	Minería de Datos sobre Moodle.	33
Figura 5.1.	Red Semántica del Modelo de Negocio aplicada al caso de estudio.	69
Figura 5.2.	Red Semántica del Problema de Explotación de Información aplicada al caso de estudio.	72
Figura 5.3.	Representación Conceptual del Dominio.	80
Figura 5.4.	Red Semántica del Problema de Explotación de Información aplicada al caso de estudio.	83
Figura 5.5.	Red Semántica del Modelo de Negocio aplicada al caso de estudio.	91
Figura 5.6.	Red Semántica del Problema de Explotación de Información aplicada al caso de estudio.	95

ÍNDICE DE TABLAS

Tabla 2.1.	Grupos de Aplicación de Procesos de Explotación de Información.	19
Tabla 2.2.	Mapeo de Procesos de Explotación de Información con Técnicas de Minería de Datos.	29
Tabla 4.1.	Procedimiento de derivación del proceso de explotación de información a partir del modelado de domino del negocio.	56
Tabla 4.2.	Detalle de los subpasos de derivación del proceso de explotación de información.	57
Tabla 4.3	Proceso Propuesto para el Estudio del Comportamiento de Comunidades Educativas.	59
Tabla 5.1.	Listado de Campos Registrados.	66
Tabla 5.2.	Término-Categoría-Definición del dominio aplicada al caso de estudio	68
Tabla 5.3.	Concepto-Atributo-Relación-Valor del domino aplicada al caso de estudio.	69
Tabla 5.4.	Concepto-Relación del dominio aplicado al caso de estudio.	69
Tabla 5.5.	Término-Categoría-Definición del Problema de Explotación de información aplicado al caso de estudio.	71
Tabla 5.6.	Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información aplicada al caso de estudio.	73
Tabla 5.7.	Concepto-Relación del Problema de Explotación de Información aplicada al caso de estudio.	73
Tabla 5.8.	Dependencias aplicadas al caso de estudio.	73
Tabla 5.9.	Término-Categoría-Definición del Problema aplicado al caso de estudio.	79
Tabla 5.10.	Concepto-Atributo-Relación-Valor Extendida del dominio aplicada al caso de estudio.	80
Tabla 5.11.	Término-Categoría-Definición del Problema de Explotación de Información aplicado al caso de estudio.	82
Tabla 5.12.	Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información aplicada al caso de estudio	83
Tabla 5.13.	Concepto-Relación del Problema de Explotación de Información aplicada al caso de estudio.	83
Tabla 5.14.	Término-Categoría-Definición del Problema de Explotación de Información aplicado al caso de estudio.	88
Tabla 5.15.	Término-Categoría-Definición del Problema aplicado al caso de estudio.	89
Tabla 5.16.	Concepto-Atributo-Relación-Valor del domino aplicada al caso de estudio.	90

Tabla 5.17. Concepto-Relación del dominio aplicado al caso de estudio.	90
Tabla 5.18. Término-Categoría-Definición del Problema de Explotación de Información aplicado al caso de estudio.	93
Tabla 5.19. Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información aplicada al caso de estudio.	94
Tabla 5.20. Concepto-Relación del Problema de Explotación de Información aplicada al caso de estudio.	94
Tabla 5.21. Dependencias aplicadas al caso de estudio.	94

NOMENCLATURA

A	Aprobado.
AP	Tema de la asignatura Inteligencia Artificial: Análisis de Protocolos.
ARQ	Tema de la asignatura Inteligencia Artificial: Modelos de Arquitectura de Sistemas Inteligentes.
BUSQ	Tema de la asignatura Inteligencia Artificial: Métodos de Búsqueda
CAVD	Formalismo para el modelado del domino: Tabla Concepto-Atributo-Valor del Domino.
CAVI	Formalismo para el modelado del domino: Tabla Concepto-Atributo-Valor Integrada.
CAVPEI	Formalismo para el modelado del domino: Tabla Concepto-Atributo-Valor del Problema de Explotación de Información.
CRD	Formalismo para el modelado del domino: Tabla Concepto-Relación del Domino.
CRI	Formalismo para el modelado del domino: Tabla Concepto-Relación Integrada.
CRISP-DM	Acrónimo en inglés del Proceso Estándar para la Minería de Datos (Cross Industry Standard Process for Data Mining).
CRPEI	Formalismo para el modelado del domino: Tabla Concepto-Relación del Problema de Explotación de Información.
D	Desaprobado.
DEDNAT	Tema de la asignatura Inteligencia Artificial: Deducción Natural
DM	Acrónimo en inglés de Minería de Datos (Data Mining)
DFD	Diagrama de Flujo de Datos.
E/S-PEI	Formalismo para el modelado del domino: Tabla de Entradas y Salidas para el Problema de Explotación de Información.
EDM	Acrónimo en inglés de la Minería de Datos aplicada a la Educación (Educational Data Mining).
EEUU	Estados Unidos.
EIE	Explotación de Información Educativa.
EMP	Tema de la asignatura Inteligencia Artificial: Emparrillado
EVEA	Entorno Virtual de Educación y Aprendizaje.
FRBA	Facultad Regional Buenos Aires.
GR	Atributo Clase Identificado perteneciente a los modelos de explotación de información.
HTML	Acrónimo en inglés del lenguaje de programación de páginas web (Hyper Text Markup Language).
M	Mal.
MSU	Acrónimo en inglés de la Universidad del Estado de Michigan (Michigan State University).
PROMEI	Programa de Mejoramiento para la Enseñanza de la Ingeniería.
R	Regular.

RSI-MN-EI	Formalismo para el modelado del domino: Red Semántica Integrada Modelo de Negocio - Explotación de Información.
RSI-MN–EI-DE/S	Formalismo para el modelado del domino: Red Semántica Integrada Modelo de Negocio - Explotación de Información con Datos de Entrada y Salida identificados.
RSMN	Formalismo para el modelado del domino: Red Semántica del Negocio
RSPEI – DE/S	Formalismo para el modelado del domino: Red Semántica del Problema de Explotación de Información con Datos Identificados de Entrada y Salida.
SOM	Acrónimo en inglés de Mapas Auto Organizados (Self Organized Map).
TDIDT	Acrónimo en inglés para árboles de decisión (Top Down Induction Decision Trees).
TIC	Tecnologías de la Información y Comunicación.
TPA	Acta de Trabajos Prácticos.
TRAD	Tema de la asignatura Inteligencia Artificial: Traducción Lógica
UTN	Universidad Tecnológica Nacional.
V/F	Verdadero / Falso

1. INTRODUCCION

En este capítulo se presenta el contexto general de la tesis (sección 1.1), luego se expone la síntesis del trabajo realizado (sección 1.2), se aborda la estructura de la tesis (sección 1.3), y se presenta la producción científica realizada (sección 1.4).

1.1. Contexto General de la Tesis

En esta sección se presentan las generalidades propias de la educación a distancia (sección 1.1.1), luego se aborda el contexto general de la explotación de información (sección 1.1.2), y finalmente se exponen las generalidades propias de la explotación de información educativa (sección 1.1.3).

1.1.1. Educación a distancia

[Tiffin y Ragasingham, 1997] sostienen que tradicionalmente, nunca se había concebido una fórmula de enseñanza que no contemplara lo presencial, es decir, la coincidencia física en espacio y tiempo del maestro y el alumno. Así, a lo largo de la historia, los procesos educativos se basaban en la idea de que el contacto educador-educado era la actividad básica que posibilitaba una formación esencialmente oral, ayudada por un soporte escrito que podía presentar ciertas limitaciones.

A partir de la invención de la imprenta, la relación educativa, tradicionalmente considerada un asunto exclusivo entre maestro y alumno, incorporó un tercer elemento: los libros. La creación de los grandes sistemas escolares públicos, significó la generalización y la expansión de unas formas de enseñanza y aprendizaje basadas, lógicamente, en la acción de un maestro o maestros sobre un grupo de alumnos más o menos numeroso, en un espacio y en un tiempo concreto (la clase) y con el soporte de libros de texto. En tanto, la enseñanza universitaria no fue ajena a esta evolución [Tiffin y Ragasingham, 1997].

A finales del siglo XIX, a causa de la disminución progresiva del precio del papel y la creación de los grandes servicios públicos de correos, aparecen las primeras fórmulas

alternativas de enseñanza no presencial, concebidas entonces como enseñanza por correspondencia [Moore y Kearsley, 1996]. Todas ellas se basan en la concepción de que aquello que el maestro enseña oralmente (la lección magistral) puede ser sustituido por aquello que puede ser explicado sobre un soporte de papel impreso, pues al fin y al cabo, según esta visión prototípica y tradicional de la enseñanza a distancia, se trata de transmitir contenidos desde la mente del profesor hasta la del estudiante.

A medida que la tecnología fue evolucionando, fueron cambiando las modalidades de la educación a distancia utilizando diferentes soportes para lograr el objetivo: radiodifusión, televisión, videos, informática y telemática fueron algunas de ellas, hasta llegar a la actualidad con internet. Pues ahora la educación a distancia utiliza mayormente plataformas virtuales especialmente preparadas para la ocasión [Tiffin y Ragasingham, 1997].

La educación a distancia y presencial tienen diferencias tangibles. En la educación a distancia el alumno y el profesor se encuentran separados físicamente y carecen de una interacción cara a cara. La educación a distancia es un sistema de enseñanza mediado por tecnología, donde el profesor y el alumno no comparten un espacio físico en común para llevar a cabo este proceso. Como principales características podemos mencionar [Aretio, 1987]: (a) la separación alumno-profesor, (b) la utilización sistemáticas de medios y recursos tecnológicos, (c) el aprendizaje individual, (d) el apoyo de una organización de carácter tutorial y (e) la comunicación bidireccional.

En la educación a distancia la interacción alumno-profesor, en general, se hace por intermedio de recursos tecnológicos [Quesada, 2006], donde la posibilidad de comunicarse con lenguaje gestual, corporal y oral, está dificultada. Estos recursos tecnológicos son diversos y han ido cambiando a lo largo del tiempo: correo postal, cassette y teléfono son alguno de los ejemplos. Con la aparición de la computadora personal, y mucho más adelante de internet, la tecnología utilizada como soporte para la educación a distancia ha ido evolucionando rápidamente. Actualmente los cursos

de educación dictados a distancia se apoyan, en su mayoría, en Entornos Virtuales de Enseñanza y Aprendizaje (EVEAs). Estas plataformas, además de contener información de los participantes, cuentan también con herramientas que dan soporte a la comunicación entre alumnos (foros, mensajes, chats, repositorios de materiales de lectura, entre otros). Estos instrumentos generan gran cantidad de información [Charum, 2007].

En [Saavedra-Martinez *et al.*, 2012] se señala que para Spotts y Bowman [1995] “...el aprendizaje es definido como un cambio estable en la conducta. Antes de aprender, un alumno no es capaz de lograr un objetivo y, luego del aprendizaje, será capaz de alcanzarlo. En este contexto, la evaluación es la herramienta que utilizan los docentes que les permite observar los avances en el desarrollo de las habilidades de conducta, los objetivos del programa y detectar los errores frecuentes evaluando la eficacia del programa de la materia propuesto. Sin embargo, las evaluaciones por sí mismas no permiten al docente diagnosticar las dificultades de aprendizaje de los alumnos y sus causas...”

1.1.2. Explotación de Información

En [Larose, 2005] se define el término explotación de información como el proceso de descubrir nuevas correlaciones, patrones y tendencias utilizando grandes cantidades de datos almacenados en repositorios, utilizando tecnologías de reconocimiento de patrones así como herramientas matemáticas y de estadística.

La Explotación de Información es una subdisciplina de los Sistemas de Información [Negash y Gray, 2008] que provee herramientas de análisis y síntesis para extraer conocimiento no trivial el cual se encuentra (implícito) en los datos disponibles en diferentes fuentes de información [Schiefer *et al.*, 2004].

Un Proceso de Explotación de Información, puede ser definido como un conjunto de tareas relacionadas lógicamente [Curtis *et al.*, 1992] que son ejecutadas para lograr, desde un conjunto de información con un grado de valor para la organización, otro

conjunto de información con un grado mayor de valor que el primero [Ferreira *et al.*, 2005; Kanungo, 2005].

En [Britos, 2008] se proponen los siguientes procesos de explotación de información: descubrimiento de reglas de comportamiento, descubrimiento de grupos, descubrimiento de atributos significativos, descubrimiento de reglas de pertenencia a grupos y ponderación de reglas de comportamiento o de pertenencia. El proceso de descubrimiento de reglas de comportamiento aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema. El proceso de descubrimiento de grupos aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio de problema. El proceso de ponderación de interdependencia de atributos aplica cuando se requiere identificar cuáles son los factores con mayor incidencia (o frecuencia de ocurrencia) sobre un determinado resultado del problema. El proceso de descubrimiento de reglas de pertenencia a grupos aplica cuando se requiere identificar cuáles son las condiciones de pertenencia a cada una de las clases en una partición desconocida “a priori”, pero presente en la masa de información disponible sobre el dominio de problema. El proceso de ponderación de reglas de comportamiento o de la pertenencia a grupos aplica cuando se requiere identificar cuáles son las condiciones con mayor incidencia (o frecuencia de ocurrencia) sobre la obtención de un determinado resultado en el dominio del problema, sean éstas las que en mayor medida inciden sobre un comportamiento o las que mejor definen la pertenencia a un grupo.

Cada proceso de explotación de información tiene asociado una familia de técnicas de minería de datos [García-Martínez *et al.*, 2013] que son útiles para descubrir los patrones de conocimiento que están siendo buscados. Varias de estas técnicas vienen del campo del Aprendizaje Automático [García-Martínez *et al.*, 2003].

Una vez que el problema de explotación de información es identificado, el Ingeniero de Explotación de Información selecciona la secuencia de procesos de explotación de

información a ser ejecutados para resolver el problema de descubrimiento de patrones de conocimiento.

1.1.3. Explotación de Información Educativa

La Explotación de Información Educativa (EIE) consiste en aplicar técnicas de minería de datos a los datos generados a partir de los procesos de enseñanza y aprendizaje. El objetivo es poder analizar este tipo de datos con el propósito de ayudar a resolver problemas de investigación educativa [Barnes *et al.*, 2009]. La EIE está relacionada con el desarrollo de métodos para explorar tipos únicos de datos en entornos educativos, utilizando estos métodos para comprender mejor a los estudiantes y los entornos en los que aprenden [Baker, 2010]. Por un lado el crecimiento tanto en software educativo como en bases de datos con información de los estudiantes, han creado un gran repositorio de datos, reflejando la forma en que los estudiantes aprenden [Koedinger *et al.*, 2008]. Por otro lado, el uso de internet en la educación ha creado un nuevo contexto conocido como *e-learning* o educación basada en la web, en la cual se genera gran cantidad de información sobre la interacción enseñanza-aprendizaje [Castro *et al.*, 2007]. La EIE busca usar estos repositorios de datos para comprender mejor a los estudiantes y profesores, y desarrollar aproximaciones computacionales que combinen datos y teoría para mejorar la forma en que se realiza la práctica.

La EIE se ha convertido en un área de investigación en los últimos años para los investigadores de todo el mundo abarcando temáticas como [Romero y Ventura, 2010]:

- Educación Presencial: en este ámbito se transmiten conocimientos y habilidades basadas en el contacto cara a cara y también se estudia psicológicamente cómo el humano aprende (se han aplicado técnicas psicométricas y estadísticas a datos obtenidos desde los EVEAs, relacionados con el comportamiento y rendimiento de los estudiantes).

- Educación a Distancia y EVEAs: la educación a distancia soportada por EVEAs además de brindar aprendizaje en línea provee herramientas de comunicación, colaboración y administración (se han aplicado técnicas de minería web a los datos que estos sistemas guardan en archivos *logs* y base de datos sobre las actividades realizadas por los estudiantes).
- Tutoría Inteligente: consiste en sistemas que adaptan el aprendizaje a las necesidades de cada estudiante en particular (la explotación de información ha sido aplicada a los datos generados por estos sistemas, como los archivos de *logs*, modelos de usuarios, etc.).

Avalando el uso de explotación de información en el campo de la educación se han desarrollado experiencias tanto a [a] nivel internacional como a [b] nivel nacional:

[a] Desde una visión de marco teórico, en [Luan, 2012; Baker y Yacef, 2009] se revisa la historia y las tendencias en el campo de la Explotación de Información Educativa. Se formulan reflexiones sobre el perfil metodológico de la investigación en EIE; y se discuten los cambios y tendencias llevadas por la comunidad. En particular, se discute el mayor énfasis en la predicción y el surgimiento de trabajo utilizando los modelos existentes para hacer descubrimiento de conocimiento en el ámbito educativo. Por otra parte, en [Romero y Ventura, 2007] analizan la aplicación de explotación de información en los sistemas de enseñanza tradicionales. Se realiza una comparación de la explotación de información aplicadas a sistemas de *e-learning* con los de *e-commerce*. En un trabajo posterior [Romero y Ventura, 2010] formulan una revisión bibliográfica sobre estudios relevantes realizados en EIE. Allí se describen los diferentes grupos de usuarios, tipos de entornos educativos y los datos que estos entornos generan. También se enumeran aplicaciones típicas de la explotación de información en el ámbito educativo mencionando las técnicas de minería de datos utilizadas. En [Romero *et al.*, 2008] se describe un proceso para explotación de información de datos de *e-learning*, así como la forma de aplicar las técnicas de minería de datos a la

obtención de reglas de comportamiento en el espacio virtual provisto por la plataforma Moodle.

Desde una visión de aplicativa, en [Minaei-Bidgoli *et al.*, 2003] se clasifican a los estudiantes prediciendo su calificación final basándose en características extraídas de los *logs* de acceso al EVEA. Se diseñan una serie de patrones de clasificación y se comparan entre sí. En [Bravo y Ortigosa, 2009] se evalúan aplicaciones de *e-learning* a través de los datos generados por los estudiantes, proponiendo un acercamiento para detectar potenciales síntomas de baja performance en cursos de *e-learning* utilizando el proceso de descubrimiento de reglas. En [Dekker *et al.*, 2009] se presenta un caso de estudio en el cual, se utilizan procesos de explotación de información para predecir el abandono de estudiantes de una carrera de ingeniería luego del primer semestre de sus estudios, o antes de entrar a la carrera, e identifica los factores de éxito para esta carrera.

[b] En [Jimenez-Rey *et al.*, 2008; Britos *et al.*, 2008] se utilizan procesos de explotación de información para ayudar a los docentes a diagnosticar las dificultades de aprendizaje de los alumnos (y sus causas) relacionadas a sus errores en programación con base en un proceso de descubrimiento del conocimiento de tres pasos. En [Britos *et al.*, 2008] se presenta una solución al problema de la selección del estilo de enseñanza utilizando el proceso de explotación de información de descubrimiento de grupos, para encontrar los patrones de estudiantes con características comunes partiendo de los datos de estilos de aprendizaje y personales, todos ellos relevados sobre cada individuo. Estos patrones de conocimiento descubiertos son utilizados para determinar el protocolo pedagógico, conforme a los descriptos por la Teoría Uno [Perkins, 1995], que mejor aplica a cada estudiante. En [Kuna *et al.*, 2009; Kuna *et al.*, 2010] se aborda el estudio del abandono de los estudios universitarios en el nivel de pregrado buscando posibles causas de deserción, utilizando los procesos de explotación de información: descubrimiento de reglas y

descubrimiento de atributos significativos para caracterizar el abandono e identificar las variables con mayor incidencia en la deserción, a partir de la información disponible en el Sistema SIU-Guarani [SIU-Guarani, 2014]. En [Saavedra-Martinez *et al.*, 2012] se define una metodología basada en procesos de explotación de información: descubrimiento de reglas y descubrimiento de atributos significativos, que permiten al docente: (1) identificar los errores de aprendizaje de los alumnos en instancias evaluativas y (2) diagramar los conceptos enseñados en pos de minimizar, en tanto sea posible, dichos errores.

1.2. Síntesis del Trabajo Realizado

Durante el desarrollo de esta tesis se abordará la investigación de temas tales como los patrones de comportamiento de comunidades educativas, los procesos de explotación de información, la creación de procesos de explotación de información aplicables a la identificación de comportamientos sociales y la propuesta de procesos de identificación de patrones de comportamientos de poblaciones educativas utilizando explotación de información.

De este modo, y a partir de lo expresado, se espera arribar a la definición de procesos de explotación de información que permitan estudiar el comportamiento de los estudiantes a lo largo de un curso o una carrera. Y así, identificar y resolver distintos problemas asociados a la apropiación de los conceptos y al favorable desarrollo de la materia, brindando una herramienta que favorezca la comprensión del curso y la posibilidad de implementar medidas que permitan detectar alumnos con dificultades en el curso e intervenir de manera temprana, pudiendo corregir factores que impacten en el resultado de la cursada.

1.3. Estructura de la Tesis

En el capítulo “Estado de la Cuestión” se presenta una introducción general a la explotación de información, luego se abordan los diferentes procesos de explotación

de información, se presentan los antecedentes de la explotación de información en el ámbito educativo, y se exponen diferentes casos de estudio.

En el capítulo “Problema” se abordan los problemas comunes relacionados con el rendimiento de los estudiantes tanto en la educación presencial como a distancia, luego se proponen las líneas de investigación con el objetivo de ayudar a resolver los problemas mencionados los diferentes procesos de explotación de información.

En el capítulo “Solución” se presentan los marcos de la solución propuesta y luego se plantea la solución a los problemas desarrollados en el capítulo anterior.

En capítulo “Validación” se describen tres casos; el primero de ellos de Procesos de Identificación de Comportamiento de Recursantes Utilizando EVEAs, el segundo consiste en Identificación Conceptos Críticos de una Cursada, y por último Identificación de Contenidos Débilmente Apropriados por Estudiantes.

En el capítulo “Conclusiones” se exponen los aportes realizados por la tesis y se presentan las futuras líneas de investigación.

1.4. Producción Científica

Durante el desarrollo de esta tesis se han publicados resultados parciales de la misma.

Artículos en Revistas con Referato:

- Cigliuti, P., Martins, S., García-Martínez, R. 2014. *Proceso de Identificación de Comportamiento de Estudiantes Recursantes Utilizando EVEAs*. Revista Iberoamericana de Educación en Tecnología y Tecnología en Educación, 13: 61-71. ISSN 1850-9959.

Comunicaciones a Congresos con Referato:

- Cigliuti, P., Pollo-Cattaneo, F., García-Martínez, R. 2012. *Procesos de Identificación de Comportamiento de Comunidades Educativas Centradas en EVEAs*. Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación. Pág. 954-956. ISBN 978-950-766-082-5.

- Cigliuti, P., Martins, S., Garcia-Martinez, R. 2014. *Identificación del Comportamiento de Estudiantes Basado en Explotación de Información*. Memorias IV Jornadas de Enseñanza de la Ingeniería. ISSN 2313-9056.
- Cigliuti, P., Martins, S., Rodríguez, D., García-Martínez, R. 2014. *Identificación de Contenidos Débilmente Apropriados por Estudiantes Basada en Explotación de Información*. Caso de Estudio. XII Workshop Tecnología Informática Aplicada en Educación. Proceedings XX Congreso Argentino de Ciencias de la Computación. Universidad Nacional de la Matanza. ISBN 978-987-3806-05-6.

2. ESTADO DE LA CUESTION

En este capítulo se presenta una introducción general a la explotación de información (sección 2.1), luego se abordan los diferentes procesos de explotación de información (sección 2.2), se presentan los antecedentes de la explotación de información en el ámbito educativo (sección 2.3), y se exponen diferentes casos de estudio (sección 2.4).

2.1. Introducción General a la Explotación de Información

Se denomina inteligencia de negocio al conjunto de estrategias y herramientas [Reinschmidt y Allison, 2000] enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización [Negash y Gray, 2008]. Involucra el uso de los datos de una organización para facilitarle a las personas que realizan la toma de decisiones estratégicas del negocio, la comprensión del funcionamiento actual y la anticipación de acciones para dar una dirección bien informada a la organización [Lönnqvist y Pirttimäki, 2006].

La explotación de información se ha definido como la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información [Grigori *et al.*, 2004]. Al hablar de explotación de información basada en sistemas inteligentes [Michalski *et al.*, 1998] se refiere específicamente a la aplicación de métodos de sistemas inteligentes, para descubrir y enumerar patrones presentes en la información. Los métodos basados en sistemas inteligentes [Kononenko y Cestnik, 1986], permiten obtener resultados de análisis de la masa de información que los métodos convencionales [Michalski, 1983] no logran tales como: los algoritmos TDIDT (Top Down Induction Decision Trees), los mapas auto organizados (SOM) y las redes bayesianas. Se ha señalado la necesidad de disponer de procesos [Chen *et al.*, 1996] que permitan obtener conocimiento [Chung *et al.*, 2005] a partir de las grandes masas de información disponible [Chau *et al.*, 2007], su caracterización [Golfarelli *et al.*, 2004] y tecnologías involucradas [Koubarakis y Plexousakis, 2000].

2.2. Procesos de Explotación de Información

Un proceso de explotación de información [Curtis *et al.*, 1992], puede definirse como un conjunto de tareas relacionadas lógicamente, que se ejecutan para lograr a partir de un conjunto de información con un grado de valor para la organización, otro conjunto de información con un grado de valor mayor que el inicial [Ferreira *et al.*, 2005].

Cada proceso de explotación de información define un conjunto de información de entrada, un conjunto de transformaciones y un conjunto de información de salida. Un proceso de explotación de información puede ser parte de un proceso mayor que lo abarque o bien puede incluir otros procesos de explotación de información que deban ser incluidos en él, admitiendo una visión desde varios niveles de granularidad [Kanungo, 2005].

Identificado el problema de inteligencia de negocio y las técnicas de explotación de información, un proceso de explotación de información describe cuáles son las tareas que hay que desarrollar para que aplicando las técnicas de explotación a la información que se tenga vinculada al negocio se obtenga una solución al problema de inteligencia de negocio [Musen *et al.*, 2003].

En este contexto [Britos y García-Martínez, 2009] proponen la siguiente caracterización de los problemas de explotación de información asociados a los problemas de inteligencia de negocios:

[a] Descubrimiento de Reglas de Comportamiento: el proceso de descubrimiento de reglas de comportamiento (figura 2.1) aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema. Para el descubrimiento de reglas de comportamiento definidos a partir de atributos clases en un dominio de problema que representa la masa de información disponible, se propone la utilización de algoritmos de inducción TDIDT [Britos *et al.*, 2008a] para descubrir las reglas de comportamiento de cada atributo clase. Los pasos a seguir para este proceso consisten en:

- a) Identificar todas las fuentes de información (bases de datos, archivos planos, entre otras), e integrarlas entre sí formando una sola fuente de información a la que se llamará datos integrados.
- b) Seleccionar un atributo clase con base en los datos integrados.
- c) Aplicar el algoritmo de inducción TDIDT al atributo clase. A partir de esto se obtiene un conjunto de reglas que definen el comportamiento de dicha clase.

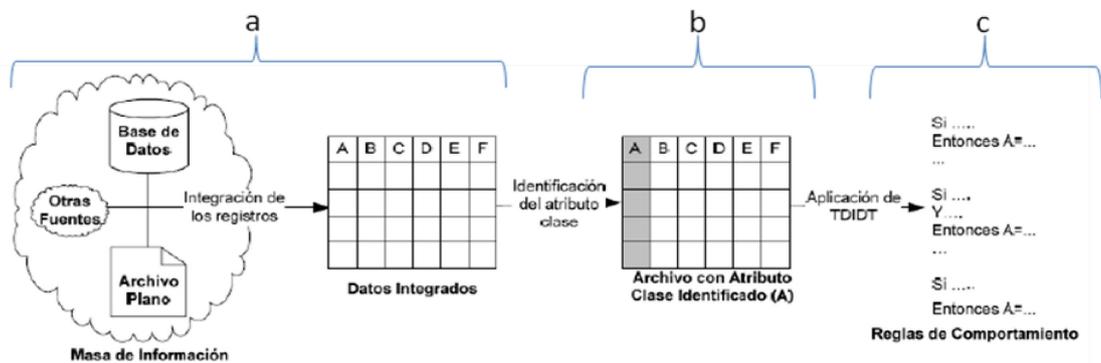


Figura 2.1-Proceso de Descubrimiento de Reglas de Comportamiento

[b] Descubrimiento de Grupos: se aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio de problema. Para el descubrimiento de grupos, a partir de masas de información del dominio de problema sobre las que no se dispone ningún criterio de agrupamiento "a priori", se propone la utilización de Mapas Auto Organizados de Kohonen. El uso de esta tecnología busca descubrir si existen grupos que permitan una partición representativa del dominio de problema que la masa de información disponible representa. Este proceso y sus subproductos pueden ser visualizados gráficamente en la figura 2.2 y consiste en:

- a) Identificar todas las fuentes de información (bases de datos, archivos planos, entre otras), e integrarlas entre sí formando una sola fuente de información a la que se llamará datos integrados.

- b) Aplicar mapas auto-organizados con base en los datos integrados. Como resultado de la aplicación de SOM se obtiene una partición del conjunto de registros en distintos grupos a los que se llamará grupos identificados.
- c) Generar el archivo correspondiente para cada grupo identificado.

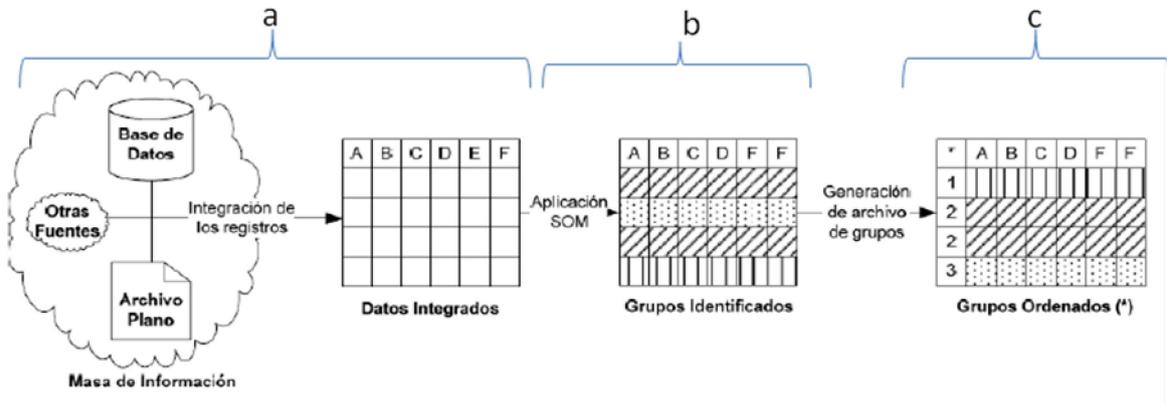


Figura 2.2- Proceso de Descubrimiento de Grupos

[c] Ponderación de Interdependencia de Atributos: este proceso (Figura 2.3) se aplica cuando se requieren identificar cuáles son los factores con mayor incidencia sobre un determinado resultado o problema. Los pasos de este proceso son los siguientes:

- a) Identificar todas las fuentes de información (bases de datos, archivos planos, entre otras), e integrarlas entre sí formando una sola fuente de información a la que se llamará datos integrados.
- b) Seleccionar el atributo clase con base en los datos integrados.
- c) Aplicar Redes Bayesianas al archivo con el atributo clase identificado. A partir de esto se obtiene el árbol de aprendizaje.
- d) Aplicar el aprendizaje predictivo al árbol de aprendizaje obtenido en c). Así, se obtiene el árbol de ponderación de interdependencias que tiene como raíz al atributo clase y como hojas a los otros atributos con la frecuencia (incidencia) sobre el atributo clase.

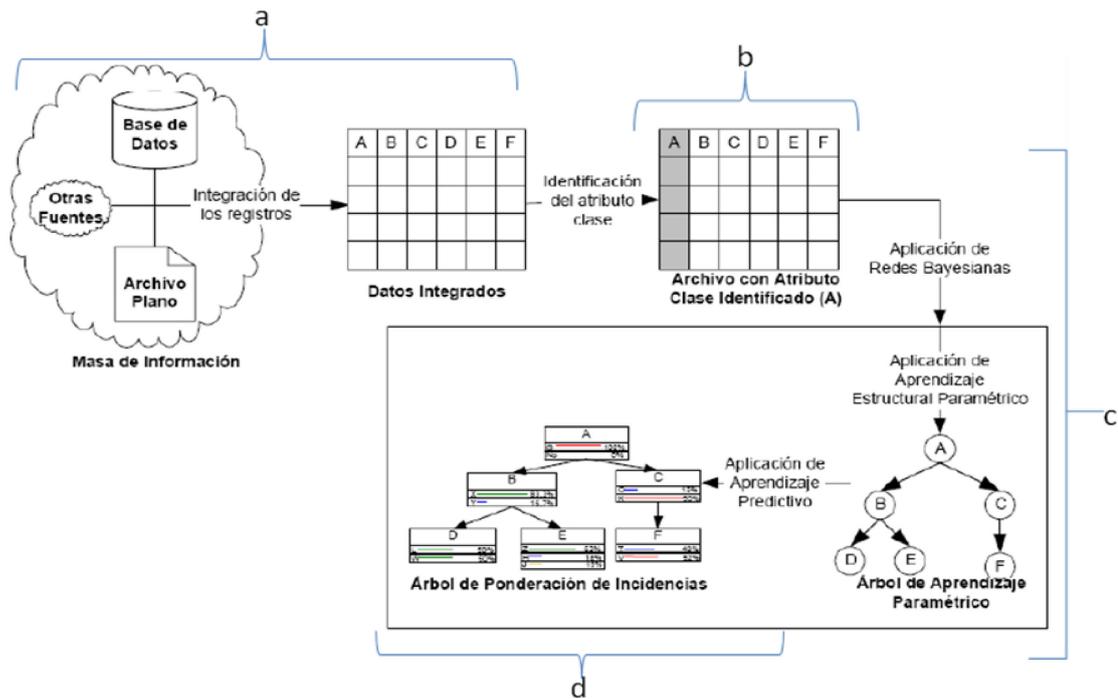


Figura 2.3- Proceso de Ponderación de Interdependencia de Atributos

[d] Descubrimiento de Reglas de Pertenencia a Grupos: este proceso (Figura 2.4) aplica cuando se requiere identificar cuáles son las condiciones de pertenencia a cada una de las clases en una partición desconocida “a priori”, pero presente en la masa de información disponible sobre el dominio de problema. Para el descubrimiento de reglas de pertenencia a grupos se propone la utilización de mapas auto-organizados para el hallazgo de los mismos y; una vez identificados los grupos, la utilización de algoritmos de inducción para establecer las reglas de pertenencia a cada uno [Britos *et al.*, 2008b]. Los pasos para realizar este proceso son:

- Identificar todas las fuentes de información (bases de datos, archivos planos, entre otras), e integrarlas entre sí formando una sola fuente de información a la que se llamará datos integrados.
- Aplicar mapas auto-organizados con base en los datos integrados. A partir de esto se obtiene una partición del conjunto de registro en distintos grupos. A estos grupos se los llama grupos identificados.

- c) Generar los archivos asociados a cada grupo identificado. A este conjunto de archivos se lo llama grupos ordenados. El atributo “grupo” de cada grupo ordenado se identifica como el atributo clase de dicho grupo, constituyéndose este en un archivo con atributo clase identificado (GR).
- d) Aplicar el algoritmo de inducción TDIDT al atributo clase de cada grupo GR. De esta manera, se obtiene un conjunto de reglas que definen el comportamiento de cada grupo.

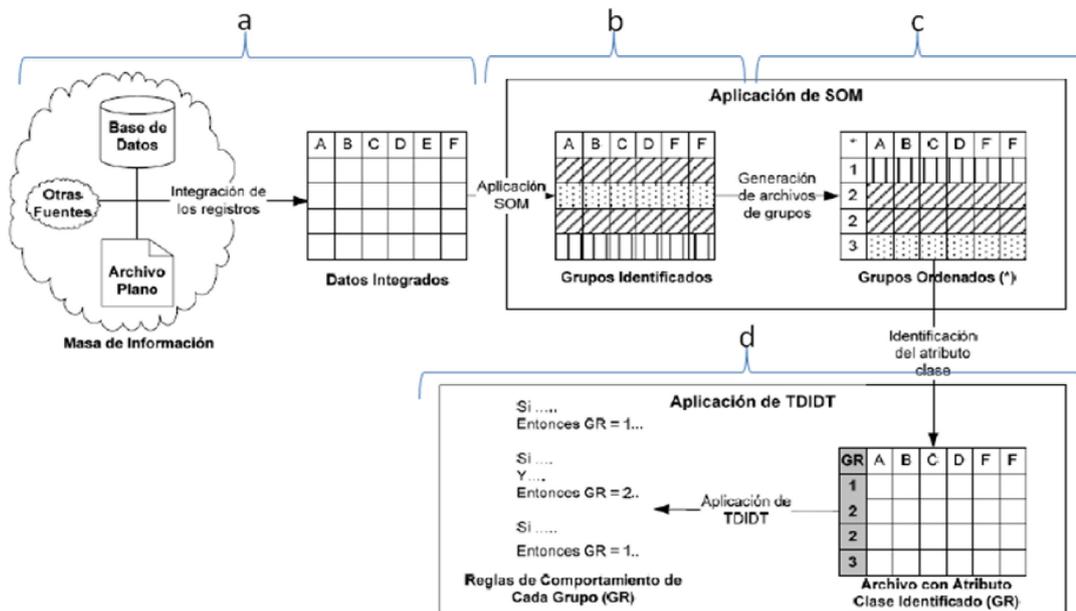


Figura 2.4-Proceso de Descubrimiento de Reglas de Pertenencia a Grupos

[e] Ponderación de Reglas de Comportamiento o de Pertenencia a Grupos: el proceso de ponderación de reglas de comportamiento (Figura 2.5) se aplica cuando se requiere identificar cuáles son las condiciones con mayor incidencia (o frecuencia de ocurrencia) sobre la obtención de un determinado resultado en el dominio del problema, sean estas las que en mayor medida inciden sobre un comportamiento o las que mejor definen la pertenencia a un grupo. Para la ponderación de reglas de comportamiento o de pertenencia a grupos se propone la utilización de redes bayesianas [Britos *et al.*, 2008c]. Esto puede hacerse a partir de dos procedimientos dependiendo de las características del problema a resolver: cuando no hay clases/grupos identificados; o cuando hay

clases/grupos identificados. El procedimiento a aplicar cuando hay clases/grupos identificados consiste en la utilización de algoritmos de inducción TDIDT [Britos *et al.*, 2008a] para descubrir las reglas de comportamiento de cada atributo clase y posteriormente se utiliza redes bayesianas para descubrir cuál de los atributos establecidos como antecedentes de las reglas tiene mayor incidencia sobre el atributo establecido como consecuente. Los pasos para aplicar este proceso son los siguientes:

a) Identificar todas las fuentes de información (bases de datos, archivos planos, entre otras), e integrarlas entre sí formando una sola fuente de información a la que se llamará datos integrados. Con base en los datos integrados se puede aplicar dos procedimientos posibles: 1) cuando hay clases/grupos identificados, y 2) cuando no hay clases/grupos identificados.

1) Clases o grupos identificados:

1.a) Aplicar el algoritmo de inducción TDIDT al atributo clase, de manera de obtener un conjunto de reglas que definen el comportamiento de dicha clase.

1.b) Construir un archivo con los atributos antecedentes y consecuentes identificados por la aplicación del algoritmo TDIDT.

2) Clases o grupos no identificados:

2.a) Aplicar la técnica de mapas auto organizados. Como resultado de la aplicación de SOM se obtiene una partición del conjunto de registros en distintos grupos a los que se llamará grupos identificados.

2.b) Generar para cada grupo identificado el archivo correspondiente. A este conjunto de archivos se lo llama grupos ordenados. El atributo “grupo” de cada grupo ordenado se identifica como el atributo clase de dicho grupo, constituyéndose este en un archivo con atributo clase identificado (GR).

- b) Aplicar el aprendizaje estructural a los archivos generados en 1.b) o 2.b) según corresponda, para de esta forma obtener el árbol de aprendizajes.
- c) Aplicar el aprendizaje predictivo al árbol de aprendizajes. De este modo, se obtiene el árbol de ponderación de interdependencias que tiene como raíz al atributo clase (en este caso el atributo consecuente) y como nodos hojas a los atributos antecedentes con la frecuencia (incidencia) sobre el atributo consecuente.

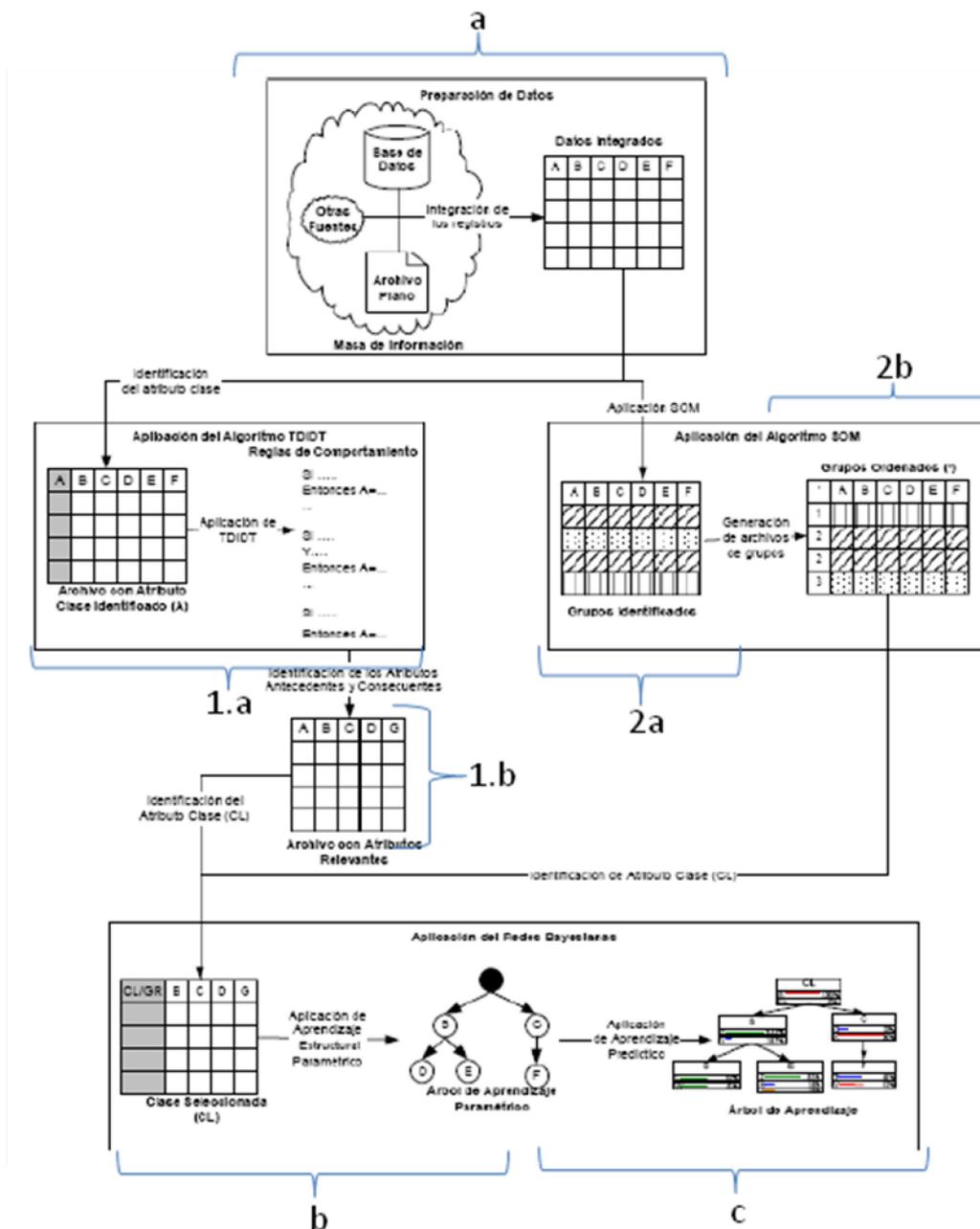


Figura 2.5- Proceso de Ponderación de Reglas de Comportamiento o de Pertenencia a Grupos

2.3. Explotación de Información en el Ámbito Educativo

Con el objetivo de proveer un ambiente de aprendizaje más efectivo se pueden utilizar procesos de explotación de información sobre la información generada por los EVEAs. El proceso de explotación de información aplicado a la educación involucra diferentes grupos de usuarios o participantes. Los diferentes grupos analizan la información relacionada con la educación desde diferentes ángulos de acuerdo a la misión, visión y objetivos de cada uno en cuanto a la utilización de los procesos de explotación de información [Hanna, 2004]. En la tabla 2.1 se clasifican posibles grupos a los cuales se les puede aplicar explotación de información [Romero y Ventura, 2010].

Usuario / Rol	Objetivo en la utilización de procesos de explotación
Estudiantes	Personalización del aprendizaje; para recomendar actividades, recursos y tareas de aprendizaje que puedan mejorar su aprendizaje; para sugerir experiencias interesantes de aprendizaje; para recomendar cursos; discusiones relevantes; libros, etc.
Profesores / Educadores	Para recibir retroalimentación objetiva sobre el instructor; para analizar el aprendizaje y comportamiento de los estudiantes; para detectar que estudiantes requieren soporte; para agrupar estudiantes en distintos grupos; para encontrar los errores cometidos con más frecuencia, etc
Desarrolladores de cursos / Investigadores de la educación	Para evaluar y mantener los cursos; para mejorar el aprendizaje de los estudiantes; para evaluar la estructura de los contenidos del curso y su efectividad en el proceso de aprendizaje; para construir automáticamente modelos de estudiantes y de tutores; para comparar técnicas de minería de datos con el objetivo de recomendar la más adecuada para cada tarea, etc.
Organizaciones / Universidades / Empresas de cursos	Para mejorar el proceso de decisiones en la escuela superior; para lograr objetivos específicos; para recomendar ciertos cursos que pueden ser mejores para cierta clase de estudiantes; para encontrar el costo más efectivo para retener alumnos y graduarlos, etc.
Administradores / Directores de colegios / Administradores de redes / Administradores de Sistemas	Para desarrollar la mejor forma de organizar los recursos institucionales (humanos y materiales) y su oferta educativa; para utilizar los recursos disponibles de forma más efectiva; para evaluar a los profesores y los planes de estudio, para configurar parámetros para mejorar la eficiencia de los sitios y adaptarlos a los usuarios, etc

Tabla 2.1 – Grupos de Aplicación de Procesos de Explotación de Información

La explotación de información ha sido aplicada al comercio electrónico en forma exitosa [Srivastava *et al.*, 2000] y ha comenzado a utilizarse en la educación a distancia con resultados prometedores. [Hanna, 2004] sostiene que, si bien los métodos utilizados en el comercio electrónico y en la educación mediada por EVEAs son similares, existen algunas diferencias importantes a resaltar:

- [a] Dominio: el propósito del comercio electrónico es guiar a los clientes en las compras, mientras el propósito del aprendizaje mediados por EVEAs es guiar a los estudiantes en el aprendizaje [Romero *et al.*, 2004].
- [b] Datos: en el comercio electrónico la utilización de los datos por lo general es simplemente acceder al *log* de acceso, pero en el aprendizaje a través de entornos virtuales hay más volumen de información sobre las interacciones del estudiante [Pahl y Donnellan, 2003].
- [c] Objetivo: el objetivo de la explotación de información en el comercio electrónico es el de incrementar las ganancias, una ganancia tangible y medible en términos de dinero, número de clientes, fidelidad del cliente, y demás. Mientras que el objetivo de la explotación de información aplicada a la educación tiene como objetivo, entre otras cosas, mejorar el aprendizaje. Por supuesto que, en materia de educación, es mucho más subjetivo y difícil de medir.
- [d] Técnicas: los EVEAs tienen características especiales que requieren de diferentes tratamientos. Como consecuencia de esto, algunos procesos de explotación de información hay que adaptarlos a necesidades particulares del proceso de aprendizaje [Li y Zaiane, 2004; Pahl y Donnellan, 2003]. Así, algunos pueden ser adaptados mientras que otros no.

La aplicación de la extracción de conocimiento en sistemas educativos con el objetivo de mejorar el aprendizaje puede ser vista como una técnica de evaluación formativa. Una técnica de evaluación formativa es la evaluación de un programa de

educación mientras está en desarrollo [Arruabarrena *et al.*, 2002], y tiene el propósito de una mejora continua.

En la figura 2.6 se puede observar a los profesores y responsables académicos que están encargados de diseñar, planificar, construir, y mantener los sistemas educativos. Los estudiantes utilizan e interactúan con estos sistemas. A partir de la información disponible en los cursos, y la interacción y uso de la misma por parte de los alumnos, se utilizan diferentes procesos de explotación de información para descubrir conocimiento el cual puede ayudar no solo a los educadores, sino también a los estudiantes [Zorrilla *et al.*, 2005].

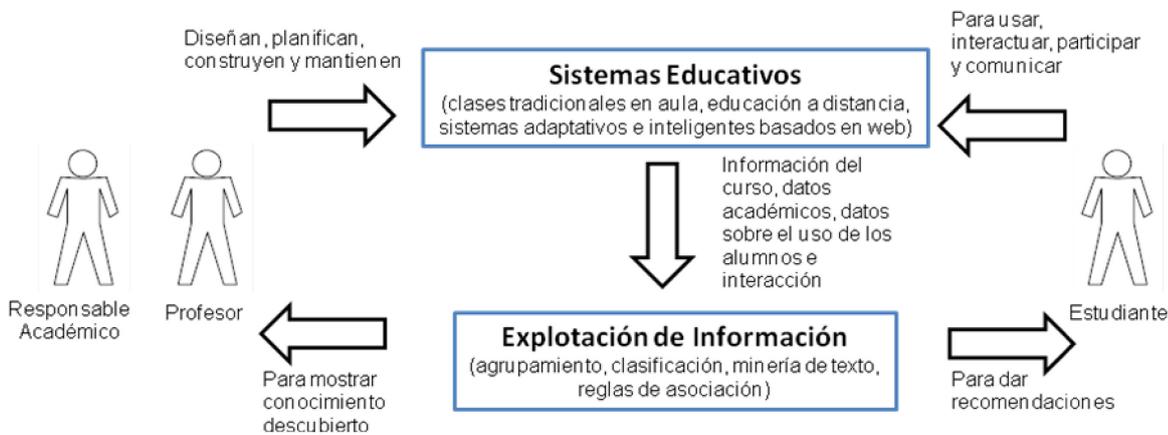


Figura 2.6 - El ciclo de la aplicación de los procesos de explotación de información en los sistemas educativos [Romero y Ventura, 2007].

[Romero y Ventura, 2007] sostienen que la explotación de información aplicada a la educación se puede aplicar en el ámbito de [a] la clase presencial y en [b] educación a distancia.

[a] Las clases en modalidad presencial son las más utilizadas actualmente en el sistema educativo y están basadas en un contacto cara a cara entre alumnos y docentes. Hay varios subtipos de educación presencial: privada, pública, primaria, secundaria, terciaria, universitaria, etc. Este tipo de educación ha sido muy criticada porque muchas veces pasa por alto las diferencias individuales y necesidades de los diferentes estudiantes, y no presta atención a la resolución

de problemas, pensamiento crítico u otras habilidades de pensamiento de orden superior [Johnson *et al.*, 2000].

En la clase presencial los docentes intentan mejorar sus clases a través de la visualización del rendimiento de los estudiantes por medio de la observación y los registros escritos, como exámenes y ejercicios prácticos. También utilizan información de asistencia, información sobre el curso, objetivos de la materia, etc. Una institución educativa posee diversas y variadas fuentes de información [Ma *et al.*, 2000]: bases de datos tradicionales (con información de los estudiantes, de los docentes, de las clases, y de la planificación de estas), información en línea (páginas web y cuenta de *twitter* de las institución, contenido de la materia), base de datos multimedia, etc.

La explotación de información puede ayudar a cualquier actor del proceso de aprendizaje (profesores, estudiantes, responsables académicos, entre otros). Las instituciones se interesan por saber qué estudiante se podría anotar a qué curso, y qué estudiante necesita ayuda para poder graduarse. Al administrador le podrían interesar predecir los cantidad de estudiantes que se van a anotar a un curso. El alumno podría querer saber cuál es el mejor curso para anotarse de acuerdo a su perfil e incluso predecir su rendimiento. Los docentes podrían querer saber cuáles son las experiencias que más contribuyen a resultados generales de aprendizaje, por qué una clase supera a otra, grupos de estudiantes similares, etc.

[b]La educación a distancia consiste en proveer acceso a programas de educación para estudiantes que están separados por tiempo y espacio del docente. Existen varias formas de implementar la educación a distancia, pero hoy en día una de las más difundidas es mediante Entornos Virtuales de Educación y Aprendizajes (EVEAs). En su mayoría estos entornos pueden por lo general grabar los accesos de los estudiantes, los cuales proveen por donde fue navegando estudiante. Si bien los *logs* pueden registrar las actividades de los estudiantes, también poseen varias limitaciones, como por ejemplo: se conoce

desde qué máquina se conecta el usuario pero no se sabe qué usuario lo hizo, se registran los clics que realiza cada uno pero no hay información sobre la actividad a la que estaban relacionados, además tampoco capturan información del contexto. Por este motivo varios autores plantean otras soluciones. Así, [Yu y Own, 2001] propone otra forma de almacenar las actividades de los estudiantes. La misma incluye el camino de aprendizaje, los cursos preferidos, el tiempo de aprendizaje, y demás. Por otro lado [Li y Zaïane, 2004] utiliza más canales de comunicación para modelar el comportamiento de navegación de los usuarios: *logs* de acceso, estructura de los sitios web visitados y sus contenidos. En cambio, [Ingram, 1999] combina los datos con otros métodos de investigación, tales como charlas informales con los estudiantes, encuestas, o devolución escrita sobre el sitio web.

En el ámbito de educación a distancia, la explotación de información puede ser utilizada, por ejemplo, para saber cómo los estudiantes hacen uso del curso, cómo una estrategia pedagógica impacta en los diferentes estudiantes, en qué orden los estudiantes estudian las diferentes unidades de las materias, cuánto tiempo el estudiante permanece en una página.

La explotación de información en el ámbito educativo se clasifica en diferentes grupos según el objetivo que ésta persiga. En [Romero y Ventura, 2010] se propone la siguiente clasificación:

- [a] Análisis y visualización de datos: se resalta la información útil y se da soporte a la toma de decisiones mediante técnicas gráficas que ayudan a las personas a analizar y entender los datos. En ambientes educativos, por ejemplo, podría ser de gran ayuda a los docentes para analizar las actividades realizadas en un curso y la información utilizada para tener una vista general del aprendizaje del estudiante.
- [b] Proveer conocimiento para dar soporte a los instructores: se genera retroalimentación de las actividades que realizan los estudiantes para que esto ayude a los profesores en la toma de decisiones (sobre mejorar el aprendizaje

de los estudiantes y organizar el material de enseñanza de manera más eficiente, entre otras cosas) y en la realización de acciones proactivas y/o correctivas.

[c] Recomendaciones para estudiantes: se hacen recomendaciones a los estudiantes, en forma personalizada, sobre sus actividades, enlaces a visitar, el próximo problema a realizar, etc. También se pueden adaptar contenidos a cada estudiante en particular.

[d] Predecir el rendimiento de los estudiantes: se estima el valor desconocido de las variables que describen al alumno. En educación a distancia por lo general estos valores se utilizan para predecir el desempeño, conocimiento, y notas de los estudiantes.

[e] Modelado del estudiante: se desarrollan modelos cognitivos para los estudiantes, incluyendo el modelado de sus habilidades y conocimiento declarativo. Se utiliza la explotación de información para examinar de forma automática características del estudiante y su comportamiento de aprendizaje con el objetivo de poder generar modelos de estudiantes en forma automática [Frías-Martínez, 2006].

[f] Detectar comportamientos no deseados de los estudiantes: se descubren aquellos estudiantes que tienen algún tipo de problema o comportamiento inusual tales como: acciones erróneas, baja motivación, abandono, etc.

[g] Agrupamiento de estudiantes: se crean grupos de estudiantes de acuerdo características particulares, funciones personalizadas, etc. Cada uno de estos grupos puede ser utilizado luego por los docentes para construir un sistema de aprendizaje personalizado, para promover aprendizaje efectivo al grupo, para adaptar contenidos, etc.

[h] Análisis de redes sociales: se busca el estudio de las relaciones entre individuos, en lugar de atributos o propiedades individuales. Una red social es

un grupo de personas que están conectados por relaciones sociales, ya sean de amistad, de cooperación o de intercambio de información [Freeman, 2006].

[i] Desarrollo de mapas conceptuales: se ayuda a los profesores a la construcción automática de mapas conceptuales. Un mapa conceptual es un gráfico que muestra relaciones entre conceptos y expresa una estructura jerárquica del conocimiento [Novak y Cañas, 2006].

[j] Construcción de cursos: se ayuda a los instructores y desarrolladores de cursos a distancia a llevar adelante el proceso de construcción y desarrollo del curso y de los contenidos de aprendizaje de forma automática. Por otro lado, también intenta promover la reutilización / intercambio de recursos de aprendizaje existentes entre los diferentes usuarios y sistemas.

[k] Planificación y organización: se mejora el proceso tradicional de educación a través de la planificación de cursos, ayudando con la organización de los estudiantes, planificando la asignación de recursos, colaborando en la admisión y desarrollando planes de estudio, entre otras cosas.

A diferencia de Romero y Ventura, [Baker y Yacef, 2009] proponen clasificar la explotación de información en el ámbito educativo de la siguiente forma:

[a] Mejoramiento de los modelos de estudiantes: un modelo de estudiante representa información sobre las características o estado del estudiante, tales como el conocimiento actual del mismo, motivación, y actitudes. Modelar las diferencias individuales de los estudiante sobre estas características permite responder a estas diferencias en forma individual, mejorando notablemente el aprendizaje del estudiante [Corbett, 2011].

[b] Estudio del apoyo pedagógico: se busca conocer qué tipo de apoyo pedagógico es más efectivo, ya sea en general, o para diferentes grupos de estudiantes, o en distintas situaciones.

[c] Evidencias empíricas: se buscan evidencias en base a las experiencias para refinar y ampliar las teorías de la educación y los fenómenos educativos ya

conocidos, a través de la obtención de un conocimiento más profundo de los factores claves que impactan en la educación, siempre teniendo en vista el mejoramiento de los sistemas de aprendizaje.

Así, luego del análisis de ambas clasificaciones en el presente trabajo se tomará una posición más cercana a los grupos definidos por [Romero y Ventura, 2010] y se hará foco en la identificación del comportamiento de las comunidades educativas a través de la explotación de información aplicada a la educación, pudiendo ayudar a actividades como:

[a] Evaluación del rendimiento del estudiante.

[b] Mejoramiento del proceso de aprendizaje.

[c] Modificación sobre el diseño de los cursos.

Para las diferentes aplicaciones mencionadas anteriormente, los distintos autores proponen la utilización de diversas técnicas de minería de datos, así [Romero y Ventura, 2007] utiliza las siguientes técnicas:

[a] Estadísticas y Visualización: si bien las estadística es a menudo la herramienta utilizada como punto de partida para el análisis de los datos generados por los sistemas de educación a distancia, no es considerada una técnica de minería de datos [Tsantis y Castellani, 2001]. Algunos ejemplos de estadísticas pueden ser indicadores tan simples como el número total de ingresos al EVEA, la cantidad de visitas por páginas, o la distribución a través del tiempo de cómo se fueron conectando los estudiantes a los entornos de aprendizaje. La visualización es una técnica que facilita el análisis de grandes volúmenes de información a través de gráficos (gráficos de línea, de dispersión, gráficos en tres dimensiones, etc.). Muchas veces, con esta técnica se representan los resultados de encuestas de admisión a las universidades, resultados de exámenes de los alumnos, etc [Shen *et al.*, 2002].

[b] Web Mining: la Minería Web [Srivastava *et al.*, 2000] es la aplicación de técnicas de minería de datos para extraer conocimiento a partir de los datos de

la web. Hay dos tipos de minería web aplicadas a la educación [Li y Zaïane, 2004]:

- Minería Web Off Line: se utiliza para descubrir patrones u otra información útil para ayudar a los educadores para validar los modelos de aprendizaje y reestructurar los EVEAs.
- Minería Web On Line: los patrones automáticamente descubiertos son introducidos en un sistema de software inteligente el cual podría ayudar a los estudiantes a llevar adelante sus tareas.

Se utilizan diferentes técnicas de datos para la minería web en sistemas educativos, pero casi todas ellas se pueden agrupar en:

- Agrupación, clasificación y detección de *outliers*: se trata de un proceso de agrupamiento físico o abstracto de objetos en clases de objetos similares.
- Reglas de asociación y patrones secuenciales: se trata de uno de los métodos de minería más estudiados en donde las normas se asocian con uno o más atributos de un conjunto de datos con otro atributo, produciendo una declaración sobre los valores de los atributos.
- Minería de texto: los métodos de minería de texto pueden ser vistos como una extensión de la minería de datos en datos de texto. Se trata de un área interdisciplinaria que involucra aprendizaje automático y minería de datos, estadísticas, información recuperación y procesamiento de lenguaje natural [Grobelnik, 2002]. La minería de textos puede trabajar con conjuntos de datos no estructurados o semi-estructurados (documentos, archivos HTML, correos electrónicos, etc.)

[Baker y Yacef, 2009] clasifican las diferentes técnicas de minería de datos para la explotación de información aplicada a la educación en cinco grupos:

- [a] Predicción: se utilizan técnicas de minería de datos como clasificación, regresión, y estimación de densidad.

[b] Agrupamiento: se utilizan técnicas de *clustering* las cuales consisten en algoritmos de agrupamiento.

[c] Minería de relación: se utilizan reglas de asociación, correlación, y patrones secuenciales.

[d] Destilación de los datos para el juicio humano.

[e] Descubrimiento con Modelos.

Las técnicas [a], [b], y [c] son las más conocidas y utilizadas por la mayoría de los investigadores. La técnica [d], puede ser comparada con las estadísticas y la técnica de visualización propuesta por [Romero y Ventura, 2007]. La técnica [e] es, quizás, en el ámbito de la explotación de información la menos común. En esta técnica se desarrolla el modelo de algún evento o hecho específico a través de algún tipo de proceso el cual puede ser validado de alguna forma (por lo general con predicción o ingeniería del conocimiento). Este modelo es luego utilizado como componente para otros análisis, como predicción o minería de relación. El descubrimiento con modelos soporta análisis sofisticados pudiendo responder a preguntas como: qué sub categoría del material de aprendizaje beneficiará mejor al estudiante [Beck and Mostow 2008], cómo los diferentes tipos de comportamiento de los estudiantes impacta en la forma de aprender de los estudiantes [Cocea et al. 2009], y cómo las variaciones en el diseño de un tutor inteligente impacta en el comportamiento de los estudiantes a través del tiempo [Jeong and Biswas 2008].

Las técnicas de minería de relación ha sido una de las técnicas más destacadas en lo concerniente a minería de datos aplicada a la educación. En la encuesta que realiza Romero y Ventura entre 1995 y 2005, de un total de 60 trabajos reportados, en el 43% estaba involucrada la minería de relación, mientras que un 28% utilizaba métodos de predicción. En la figura 2.7 se puede observar la distribución de los diferentes métodos utilizados en los trabajos mencionados.

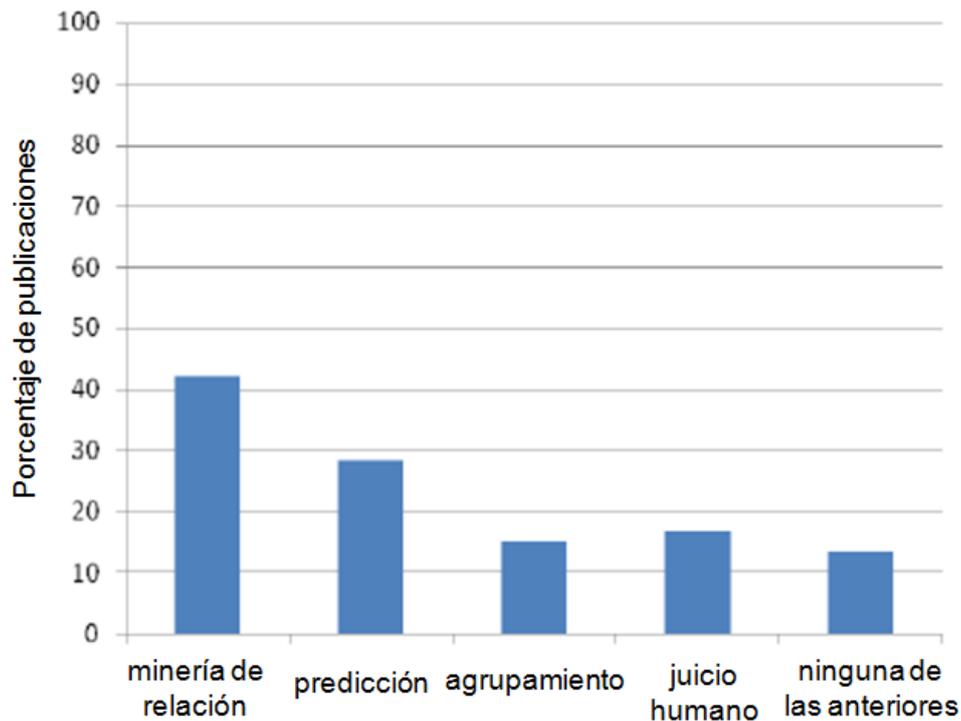


Figura 2.7–Publicaciones por Tipo de Técnica de Minería de Datos Utilizadas [Romero y Ventura, 2007]

Las técnicas mencionadas anteriormente son propias de la minería de datos. En este trabajo se propone utilizar los procesos de explotación de información definidos por [Britos y García-Martínez, 2009]. En la tabla 2.2 se realiza el mapeo de las diferentes técnicas con los procesos de explotación de información.

Proceso de explotación de información	Técnicas de minería de datos / Algoritmos
Descubrimientos de Grupos	<ul style="list-style-type: none"> • Agrupación • Estimación de densidad • Detección de Outliers
Descubrimiento de Reglas de Comportamiento	<ul style="list-style-type: none"> • Clasificación • Minería de Texto • Árboles de decisión
Ponderación de Interdependencia de Atributos	<ul style="list-style-type: none"> • Reglas de Asociación

Tabla 2.2 – Mapeo de Procesos de Explotación de Información con Técnicas de Minería de Datos.

2.4. Casos de Estudio

En esta sección se presenta los distintos casos de estudio de ámbito internacional (sección 2.4.1), y de ámbito nacional (sección 2.4.2).

2.4.1. Ámbito Internacional

A nivel internacional, los procesos de explotación de información aplicados a temáticas relacionadas con la educación llevan unos cuantos años. La cantidad de trabajos publicados sobre explotación de información aplicada a la educación ha incrementado notablemente año a año. En la figura 2.8 se puede apreciar la cantidad de publicaciones que ha realizado el grupo de investigación Working Group in EDM (Educational Data Mining) a lo largo de los años [Romero y Ventura, 2010].

Si bien se trata de las publicaciones de un solo grupo de investigación, esto puede tomarse como una muestra para apreciar la tendencia de crecimiento en cantidad de publicaciones.

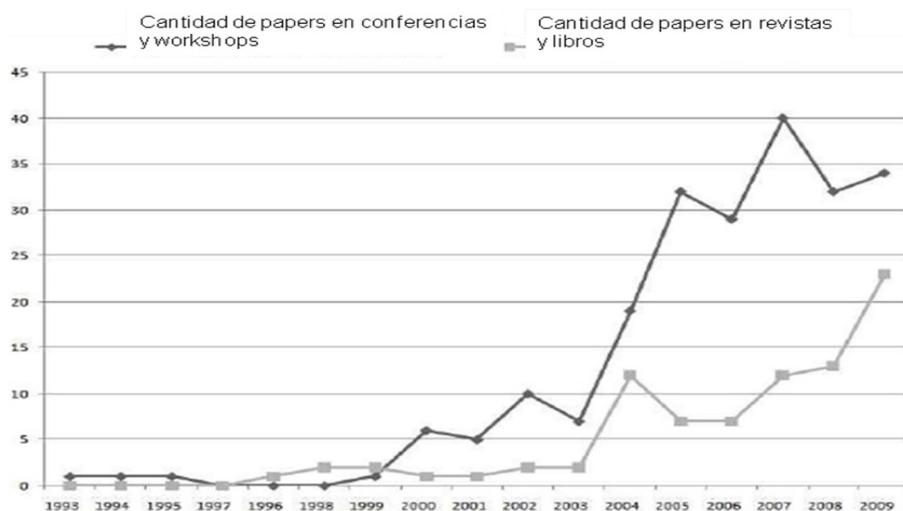


Figura 2.8 – Cantidad de trabajos publicados hasta el 2009 agrupados por año

[Romero y Ventura, 2010]

En [Luan, 2002] se propone monitorear y predecir la transferencia de estudiantes entre colegios a través de procesos de explotación de información. El modelo realizado ayuda a predecir la transferencia de cada estudiante inscripto y tiene como objetivo proveer el perfil de los estudiantes transferidos y predecir qué estudiantes

anotados en el colegio solicitarán una transferencia para, de esta forma, poder personalizar y planificar sus iteraciones e intervenciones con estos estudiantes, quienes probablemente necesiten de asistencia para realizar estos trámites. Se utilizaron datos de tres universidades diferentes sumando un total de treinta y dos mil estudiantes. Los datos utilizados incluyen algunas de las siguientes características: demográficas, ayuda financiera, estado de la transferencia, cantidad de transferencias, total de unidades ganadas y calificaciones por tipo de curso.

Para preparar los datos se proponen cuatro pasos basados en la metodología CRISP-DM:

- [a] Investigar la posibilidad de aplicar procesos de explotación de información directamente desde el *Data Warehouse*. Esto evita posibles errores en nombre de campos, cambios inesperados en los tipos de datos, esfuerzo extra en llenar diferentes repositorios de datos.
- [b] Seleccionar una sólida herramienta de consulta para poder dar forma a los archivos sobre los cuales se trabajarán.
- [c] Visualización y validación de información. Esto significa examinar la distribución de esta información a través de histogramas, gráficos y otros tipos de herramientas. Este paso da a los investigadores una primera impresión del contenido de los datos y cómo aprovecharlos en el análisis.
- [d] Explotar los datos. Se utilizaron redes neuronales, y algoritmos de inducción C5.0 y CR&T para comparar ambos modelos y complementar los resultados.

Una vez obtenidos los resultados se concluye:

- [a] La utilización de explotación de información para la investigación en los colegios secundarios en un gran desafío.
- [b] A través de la utilización de algoritmos como la máquina de aprendizaje y la inteligencia artificial para discernir reglas, asociaciones y la probabilidad de eventos, la explotación de información tiene un profundo significado de aplicación.

[c] A través del uso de explotación de información se pudo convertir una base de datos poco manejable y caótica en una base para planificar el programa y para resolver cuestiones operativas.

[d] La potencialidad de la explotación de información radica en el hecho que simultáneamente mejora las salidas y reduce costos. La doctrina del 1% [Luan, 2000] dice que un cambio del 1% significa una unidad de ganancia y una unidad de ahorro.

[e] La posibilidad de intervenir en aquellos estudiantes que van a abandonar o a cambiarse de colegio genera un valor agregado que va más allá de los costos y ahorro que se puedan generar.

[f] La explotación de información ayuda a predecir la probabilidad de personas interesadas en inscribirse a un colegio. Esto permite a los colegios anticiparse y enviarle al potencial estudiante material para que se asesore antes del ingreso.

[Romero *et al.*, 2008] proponen aplicar explotación de información sobre datos generados por un EVEA implementado sobre Moodle.

Moodle guarda en forma detallada todas las acciones que realizan los estudiantes [Rice, 2006]. En base a esas acciones guardadas en tablas de la base de datos del EVEAs se propone aplicar los siguientes procesos:

La aplicación de los procesos de información en EVEAs es un proceso iterativo [Romero *et al.*, 2006]. Se define un proceso de cuatro pasos que consiste en:

[a] Recolectar datos del EVEA,

[b] Procesar los datos,

[c] Aplicar procesos de explotación de información, e

[d] Interpretar y evaluar los resultados e implementarlos en el EVEA (Figura 2.9).

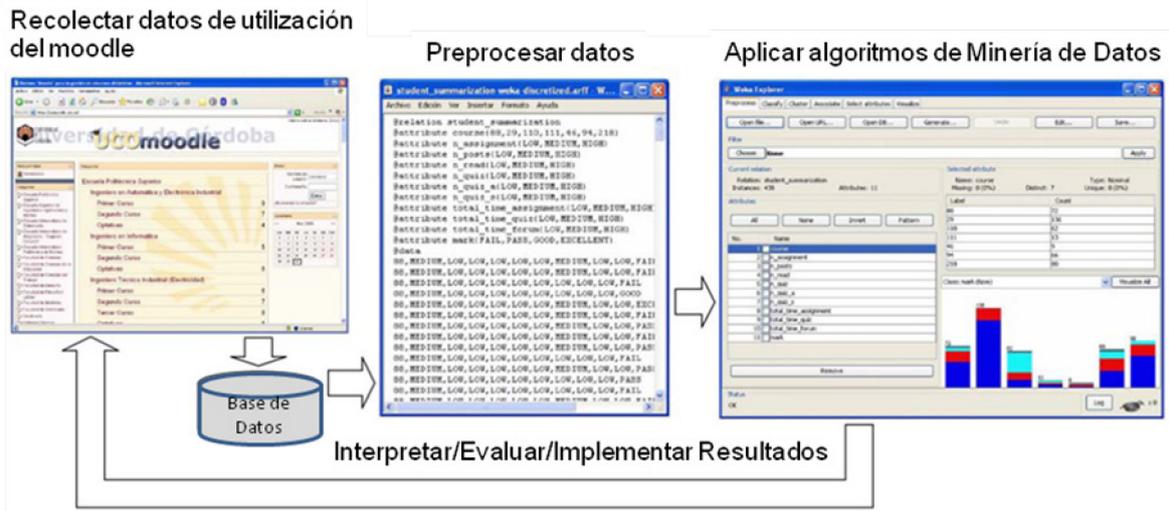


Figura 2.9 – Minería de Datos sobre Moodle [Romero *et al.*, 2008]

La aplicación de la explotación de información en la educación no dista mucho de su aplicación en otras áreas, aunque hay algunas características importantes que le son propias [Romero y Ventura, 2006]:

[a] Datos: en otros sistemas basados en web los datos utilizados para analizar, por lo general son sacados de un *log* de acceso, pero en los sistemas EVEAs hay mucha más información disponible sobre la interacción de los estudiantes. Los EVEAs pueden registrar cualquier actividad que el estudiante realice (leer, escribir, realizar exámenes, realizar diferentes tareas, e incluso comunicarse con sus compañeros). Normalmente este tipo de sistemas proveen una base de datos que almacena toda la información del sistema con la información personal de los usuarios y los datos de interacción de los usuarios.

[b] Objetivo: el objetivo de la explotación de información en cada área es diferente y puede ser más o menos objetiva. Por ejemplo, en comercio electrónico el objetivo es incrementar la rentabilidad, lo cual es tangible y puede ser medido en término de cantidades de dinero, números de clientes, y fidelidad de los clientes. Sin embargo, los objetivos de la explotación de información son los de mejorar el proceso de aprendizaje y guiar a los estudiantes en ese proceso. Este objetivo es más subjetivo y más delicado de medir.

[c] Técnicas: los EVEAs tienen características especiales que requieren un tratamiento diferente de los problemas de minería. Algunas técnicas tradicionales pueden ser aplicadas directamente, pero otras no, y tienen que ser adaptadas a un problema de educación específico.

Esta actividad se ha realizado sobre una base de 438 alumnos donde los profesores utilizan actividades del EVEA como tareas, mensajes, foros y encuestas. A continuación se mencionan los pasos utilizados para la preparación de datos:

[a] Se creó una tabla de sumarización donde se puede almacenar la información al nivel requerido para el análisis, en este caso a nivel alumno. Esta tabla tendrá la suma por fila del total de cada actividad que ha realizado cada estudiante.

[b] Se pasaron los datos continuos a discretos para poder mejorar la interpretación y comprensión.

[c] Se transformaron los datos al formato necesario para poder aplicar las técnicas de minería de datos

Una vez preparados los datos y aplicadas diferentes técnicas de explotación de información se concluye que:

[a] Se puede observar cómo diferentes técnicas de explotación de información pueden ser utilizadas para mejorar el curso y el aprendizaje de los estudiantes.

[b] Estas técnicas pueden ser aplicadas por separado en un mismo sistema o todas juntas en un sistema híbrido.

[c] Si bien se describieron las técnicas más comunes utilizadas en la explotación de información aplicada a la educación, existen otras técnicas también aplicables.

[d] En el futuro será muy útil tener herramientas de explotación de información enfocadas en la educación. Actualmente, éstas son diseñadas pensando en potencia y flexibilidad, más que en simplicidad. Las mismas son demasiado complejas de utilizar para los educadores, entonces, deben tener una interfaz de

usuario más amigable. También es necesario que las herramientas de explotación de información sean integradas en los EVEAs y que éstas sean una herramienta más que el docente pueda utilizar para obtener resultados del curso y de los alumnos.

[Minaei-Bigdoli *et al.*, 2003] se propone utilizar la explotación de información para predecir la performance de los estudiantes, buscando responder las siguientes preguntas:

[a] ¿Se pueden detectar diferentes clases de estudiantes? Es decir, existen grupos de estudiantes que utilizan los recursos del EVEA en forma similares. ¿Se pueden identificar esas clases en cada estudiante? Contestando estas preguntas se podrá lograr que los estudiantes utilicen mejor los recursos, basado en el uso de los recursos por los otros estudiantes de la misma clase.

[b] ¿Se pueden clasificar los problemas que han resuelto los alumnos? En caso de poder, se podrá ver cómo diferentes tipos de problemas impactan en el logro de cada alumno y permitirá ayudar a los profesores a realizar los ejercicios de tarea de forma más efectiva y eficiente.

Este trabajo se realiza sobre el EVEA de la Universidad del Estado de Michigan (MSU) [MSU, 2014]. Se utilizan los datos de un curso introductorio de física, compuesto por 12 tareas con un total de 184 problemas por resolver, donde se matricularon 261 alumnos y terminaron 221. El EVEA almacena diferentes características de las actividades que realiza el alumno de las cuales se utilizarán las siguientes:

[a] Número total de respuestas correctas

[b] Resolución del problema en la primera oportunidad en contraposición a la resolución del problema en un número alto de oportunidades.

[c] Total de intentos para realizar el problema en forma correcta.

[d] El tiempo total desde que se empezó a resolver el problema, hasta que se lo resolvió correctamente.

[e] El tiempo que se tardó en abordar por primera vez el ejercicio, más allá de que se haya resuelto o no.

[f] Resolución del problema tras la participación en el mecanismo de comunicación en contraposición a la resolución del problema en forma aislada.

Para este problema se utiliza el proceso de descubrimiento de reglas de comportamiento y se concluye que:

[a] La optimización exitosa de la clasificación de estudiantes demostró los méritos de utilizar los datos del EVEA para predecir las notas finales de los estudiantes basadas en sus características extraídas de las tareas.

[b] La explotación de información puede ser aplicada en forma exitosa por los docentes para incrementar el aprendizaje de los estudiantes.

[Bravo y Ortigosa, 2009] sostienen que la tecnología ha ayudado a la adopción de la educación a distancia mediante EVEAs. A través de estas plataformas podemos obtener, entre otros beneficios, mayor interacción con el alumno, enriquecer la experiencia de aprendizaje, flexibilidad de acceso, etc. Sin embargo el diseño y la evaluación de la educación a distancia mediada por sistemas EVEAs aún es una tarea compleja que requiere tiempo y experiencia que muchos docentes todavía no poseen, y quienes por lo general necesitan ayuda de expertos para poder diseñar sus cursos. Por lo tanto, se propone evaluar los sistemas EVEA's en base al análisis de los datos que los mismos registran en su *log*. Uno de los desafíos para un instructor que evalúa los EVEAs es la falta de evidencia de los estudiantes, ya que los mismos tienen solo acceso a las interacciones con el material de estudio y no se puede observar su comportamiento o recibir *feedback* de ellos en el momento. Dado que el comportamiento de los estudiantes está oculto dentro de sus interacciones, los instructores deberían analizarlos para evaluar performance de los estudiantes y

evaluar el software de *e-learning*. Lo que es más, los sistemas EVEAs generan gran cantidad de datos en sus *logs*. Se considera que la explotación de información es lo más adecuado para realizar esta tarea, ya que es una tecnología que combina métodos de análisis de datos tradicionales con algoritmos complejos para procesar grandes volúmenes de datos con el objetivo de descubrir información útil [Tan *et al.*, 2006]. El objetivo de este trabajo es encontrar patrones o síntomas que puedan indicar la baja performance de los estudiantes, para esto se analizan los *logs* generados por los EVEAs. Entonces, por ejemplo, se podrían detectar a aquellos estudiantes que en un ejercicio superan la cantidad de errores que se comenten en promedio.

Es difícil para los instructores detectar este tipo de síntomas sin la ayuda de la explotación de información. Al mismo tiempo, es beneficioso para el profesor poder ser advertido de ciertos problemas que están teniendo los estudiantes con un conjunto de ejercicios en particular y, por lo tanto, poder tomar una acción a tiempo. ¿Cómo ocurren estos síntomas? ¿Qué factores los disparan? Este trabajo propone responder a todas esas preguntas proveyendo un método para detectar síntomas de baja performance de los estudiantes en un curso.

Naturalmente, cada síntoma observado en los datos, significa una variación de performance en el curso. Es útil identificar aquellos patrones que puedan causar caídas en el rendimiento, como así también identificar cuáles son los más importantes. Para esto se propone un método que ayuda a filtrar y armar un ranking de los síntomas detectados basado en Descubrimiento de Reglas de Comportamiento.

[Bravo y Ortigosa, 2009] trabaja sobre datos por la Escuela de Ciencias de la Información de la Universidad de Pittsburgh (Pittsburgh, EEUU) los cuales fueron recolectados durante el año 2007 del EVEAs de la materia Introducción a la Programación en C, generando un registro de análisis de 55 alumnos, con un total de 52734 interacciones. Las actividades desarrolladas para esta materia en el sistema EVEA se basan en cuestionarios. Cuando un estudiante responde una pregunta del cuestionario se genera un nuevo registro en la tabla de *log* del sistema. Para el análisis se realizan tres pasos:

- [a] Se selecciona el algoritmo a utilizar. Se eligen árboles de decisión, en especial el algoritmo C4.5 dado los buenos resultados que ha arrojado en trabajos previos [Bravo *et al*, 2008].
- [b] Se definen los atributos del archivo de *log* a ser analizado, y cuál de ellos se utilizará como clase.
- [c] Se aplica el algoritmo para poder obtener los resultados e interpretarlos.
- [d] Se analizan los resultados y se deja la rama del árbol donde están todos los casos en los cuales los alumnos fallaron al responder la encuesta.
- [e] Se filtran las reglas con el objetivo de solo quedarse con aquellas reglas relevantes. Para esto se calcula la banda inferior como el 10% de los casos utilizados en las reglas.

Luego de aplicados los pasos mencionados se concluye que:

- [a] Se detectaron tres reglas que indican la caída de rendimiento para varios estudiantes.
- [b] Esta información puede ser relevante para el instructor o diseñador del curso ya que puede realizar mejoras generando nuevas actividades o modificando las existentes o, incluso, modificando la estructura del curso.
- [c] Los árboles de decisión muestran cierta debilidad, sus resultados están fuertemente relacionados con la distribución de los datos y la proporción de las clases. En consecuencia, para futuros trabajos aplicarán reglas de asociación.
- [d] Por último, se resalta la importancia de probar y verificar, con otros tipos de datos, el análisis propuesto. En este contexto, en futuros trabajos se procurará utilizar datos de otros EVEAs con el objetivo de mejorar el método de detección.

[Dekker *et al*, 2009] propone la utilización de la explotación de información para predecir el abandono de estudiantes. El monitoreo y apoyo a los estudiantes de primer

año es un tema que se considera sumamente importante en varias instituciones de educación. En algunas facultades la inscripción de los estudiantes recién graduados del bachillerato puede ser menor a lo deseado y, cuando se combina con un alto ratio de abandono, ponen de manifiesto la necesidad de dar forma a enfoques efectivos para predecir el abandono de los estudiantes, así como la identificación de los factores que afectan a todo ello. Entonces, se propone utilizar la explotación de información para predecir el abandono de los estudiantes. En el departamento de ingeniería eléctrica de la Universidad de Tecnología de Eindhoven el porcentaje de abandono de los nuevos estudiantes es de un 40% [Dekker *et al.*, 2009]. Aparte de que el departamento intenta disminuir este porcentaje, existe otra razón para identificar el éxito o fracaso de un estudiante en la etapa temprana (antes del primer semestre): la carrera de Ingeniería Eléctrica de la universidad vecina acepta alumnos provenientes de otra universidad antes del mes de enero. Por otro lado, existe siempre un grupo de estudiantes que el departamento lo considera un grupo de riesgo como, por ejemplo, estudiantes que necesitan atención extra o algún cuidado especial que los ayude a aprobar la materia. Detectar estos grupos en una etapa temprana le permite al departamento de ingeniería eléctrica enfocar sus recursos hacia los estudiantes que más lo necesitan y de esta forma poder evitar el abandono de la materia.

2.4.2. Ámbito Nacional

La aplicación de procesos de explotación de información aplicados a la educación, a nivel nacional, es algo reciente. Se han realizado algunas experiencias en temas tales como: identificación de errores de apropiación [Saavedra *et al.*, 2012b], detección de problemas de aprendizaje [Saavedra *et al.*, 2012b], e identificación de comportamiento en comunidades educativas [Cigliuti *et al.*, 2012].

En [Jiménez Rey *et al.*, 2008] se propone la utilización de herramientas de explotación de información como ayuda a los docentes para diagnosticar las dificultades de aprendizaje de los alumnos (y sus causas) relacionadas con la programación, para el cual se utilizan herramientas de bases de datos de sistemas

inteligentes: inducción de reglas por algoritmos TDIDT y redes bayesianas, con base en un descubrimiento del conocimiento de tres pasos:

- [a] Construir una base de datos sobre una caracterización estándar de cada estudiante y conceptos mal aprendidos de programación. Esta caracterización se focaliza en aspectos de: metodología de desarrollo del programa escrito por el estudiante, funcionalidad del programa desarrollado por el estudiante y calidad del diseño del programa creado por el estudiante.
- [b] Proceder al descubrimiento de reglas (mediante algoritmos TDIDT) las cuales establecen una relación entre conceptos no comprendidos de programación y sus posibles causas. Las reglas obtenidas son aplicadas para revisar la estructura propuesta para el curso y la hipótesis de los profesores. Los resultados experimentales indican que aplicando este paso, el profesor puede descubrir posibles causas de los conceptos mal aprendidos por los estudiantes.
- [c] Descubrir el peso que cada una tiene sobre cada concepto mal aprendido (mediante redes bayesianas). Esto permite establecer el rango de importancia de las causas de conceptos no comprendidos para proponer estrategias de enseñanzas remediales.

[Saavedra *et al*, 2012] proponen la utilización de procesos de explotación de información para la identificación de errores de apropiación de conceptos. El proyecto busca la obtención de una metodología que le permita al docente:

- [a] identificar los errores de aprendizaje de los alumnos en instancias educativas y;
- [b] diagramar los conceptos enseñados en pos de minimizar, en tanto sea posible, dichos errores.

Para esto, se utilizan los procesos de explotación de información [Britos, 2008: Britos y García-Martínez, 2009]: descubrimiento de reglas de comportamiento y ponderación de interdependencia de atributos. Se tuvieron en cuenta dos casos de

estudio, uno con la asignatura Análisis de Sistemas correspondiente a la carrera de Ingeniería en Sistemas de Información de la Universidad Tecnológica Nacional Facultad Regional Buenos Aires (UTN-FRBA) y otro con la asignatura Sistemas y Organizaciones de la misma casa de altos estudios. Para ambos se utilizó el siguiente proceso de cinco pasos para identificar errores de apropiación de conceptos:

Paso 1: conceptualización del dominio

- [a] Seleccionar los conceptos enseñados que se consideran importantes para la determinación del aprendizaje del individuo.
- [b] Construir un glosario de términos
- [c] Construir el mapa de dependencia o precedencia conceptual

Paso 2: identificación de Subdominio de Análisis

- [a] Identificar el submapa candidato para el análisis
- [b] Refinar los conceptos del submapa candidato en subconceptos.
- [c] Identificar los atributos de evaluación vinculados al submapa candidato.

Paso 3: preparación de datos

- [a] Identificar los valores de los atributos de evaluación, vinculados a los subconceptos del submapa candidato en las instancias de evaluación individual.
- [b] Construir la Base de Datos usando como campos los atributos de evaluación y, como instancias, los valores de estos atributos para cada sujeto de la población en estudio.

Paso 4: Explotación de Información

- [a] Aplicar el Descubrimiento de Reglas de Comportamiento [Britos, 2008], tomando como clase el atributo que se encuentra en la raíz del subárbol asociado al submapa candidato.

[b]Aplicar el proceso Ponderación de Interdependencia de Atributos [Britos, 2008], tomando a los atributos hoja de las reglas obtenidas en el Subpaso 4.1 y atributo clase

Paso 5: interpretar los resultados

[a] Interpretar los resultados generados a partir de los procesos de explotación de información del Paso 4.

Luego de realizada las pruebas de concepto con las materias mencionadas anteriormente se concluye que [Saavedra *et al.*, 2012]:

[a] La apropiación de conceptos por parte del alumno no es un proceso unidireccional. Si bien requiere un esfuerzo del alumno, es fundamental la participación del docente no solo en la transmisión pura del conocimiento sino en el proceso que esa transmisión implica.

[b] Para el docente es de gran ayuda contar con una metodología o proceso que le permita determinar qué errores en la apropiación de conceptos de los alumnos provienen de la implementación del proceso de transmisión mencionado. Entonces, se plantea una metodología que le permita a un docente revisar, en caso de ser necesario, el proceso utilizado en el dictado de sus clases para intentar minimizar los errores de apropiación de conceptos detectados en los alumnos.

[c] El proceso propuesto constituye una nueva herramienta de diagnóstico en el área; y si bien los resultados obtenidos en este estadio del proyecto son prometedores, no son concluyentes.

[Cigliuti *et al.*, 2012] sostienen que los procesos de explotación de información pueden ser utilizados también, para identificar el comportamiento de comunidades educativas, en este caso, las centradas en entornos virtuales de enseñanza.

Para cumplir con los objetivos de este estudio se llevan a cabo los siguientes pasos:

[a] Realizar una Investigación documental, estudio y síntesis de los marcos teóricos de:

[1] Educación mediada por EVEAs,

[2] Explotación de información y

[3] Comportamientos de poblacionales con comunicación mediada por tecnología.

[b] Identificar las variables utilizadas para el modelado de comportamientos poblacionales con comunicación mediada por EVEAs.

[c] Crear y validar los procesos de explotación de información que identifiquen comportamientos de poblaciones con comunicación mediadas por EVEA.

[d] Realizar una prueba de concepto de los procesos de explotación creados utilizando datos reales de poblaciones con comunicación mediadas por EVEA.

A través de los pasos mencionados el proyecto busca:

[a] La comprensión de comportamientos de comunidades mediada por tecnología con énfasis en comunidades educativas; y

[b] La formulación de una propuesta de identificación de patrones de comportamientos poblacionales en comunidades educativas mediadas por EVEAs utilizando procesos de explotación de información.

[Deroche *et al.*, 2013] proponen utilizar la explotación de información para la mejora de la asignatura de grado de Inteligencia Artificial de la Universidad Tecnológica Nacional. Dicha materia es aprobada por el 65% de los alumnos que la cursan. Si bien actualmente la mayoría de los alumnos aprueba el parcial, desde el 2011 se observa una tendencia negativa que proyectada al año 2013 bajaría a menos del 50%. Entonces, se estudian alternativas en el desarrollo de la clase con el objetivo de aumentar el porcentaje de aprobación.

Los temas que forman parte de la asignatura son los siguientes:

- Teoría (TEORIA), donde se incluyen preguntas teóricas de todos los conceptos de la asignatura.
- Modelos de Arquitecturas de Sistemas Inteligentes (ARQ) que incluye preguntas teóricas sobre conceptos y tipos de sistemas inteligentes (Redes Neuronales, Sistemas Expertos y Algoritmos Genéticos).
- Métodos de Búsqueda (BUSQ), Emparrillado (EMP), Análisis de Protocolos (AP), Deducción Natural (DEDNAT) y Traducción Lógica (TRAD), que engloban ejercicios prácticos para cada tema.

Se propone como solución la aplicación de un proceso de explotación de información aplicándolo sobre la información proveniente de exámenes parciales ya evaluados de la asignatura en cuestión, para así estudiar los temas relevantes para la evaluación de los exámenes parciales. El proceso de explotación utilizado es el de Descubrimiento de Reglas de Comportamiento [Britos y García-Martínez, 2009] a través de los siguientes pasos:

- Se prepara la base de datos con un total de 229 registros (cada uno representado por un parcial realizado por un alumno) los cuales están formados por 25 atributos y una clase.
- Se aplica el algoritmo de inducción C4.5 de la familia TDIDT [Quinlan, 1990] y como resultado se obtienen 11 reglas.
- Se interpretan los resultados

Realizados estos pasos se concluye que:

- La solución propuesta un ejemplo de herramienta de diagnóstico en el área satisfactorio.
- Se pudo observar en detalle el grado de importancia de los resultados finales permitiendo al docente mejorar el dictado de la asignatura haciendo hincapié en aquellos temas críticos para su aprobación.

- Estos resultados son concluyentes y prometedores para ser analizados en un futuro y evaluar su incidencia en la fase de aprobación de los alumnos.

3. PROBLEMA

En este capítulo se abordan los problemas comunes relacionados con el rendimiento de los estudiantes tanto en la educación presencial como a distancia (sección 3.1), luego se proponen las líneas de investigación con el objetivo de ayudar a resolver los problemas mencionados los diferentes procesos de explotación de información (sección 3.2).

3.1. Generalidades del Problema

La institución u organización es la que enseña (organización que ayuda la llama [Delling, 1987] y organización de apoyo lo denomina [Holmberg, 1989]), no el profesor. [Aretio, 1999] sostiene que en la enseñanza presencial es el docente el que habitualmente diseña, produce, distribuye, desarrolla y evalúa el proceso de enseñanza-aprendizaje, por lo que el aprendizaje del alumno suele estar en función de que le haya correspondido un buen o un mal docente. En la enseñanza a distancia, sin embargo, el docente nunca es uno, son multitud los agentes que intervienen en el proceso de enseñar y aprender, hasta tal punto que solemos reconocer a la institución como la portadora de la responsabilidad de enseñar. Por tanto, son el conjunto de agentes que conforman la institución, más que el profesor, los que diseñan, producen, distribuyen, desarrollan o tutelan el proceso de aprendizaje de los estudiantes. Pero este aprendizaje ha de permitir al estudiante ser protagonista en cuanto al tiempo, espacio y ritmo de aprender, es decir, el proceso de enseñanza diseñado por la institución propicia el aprendizaje flexible del estudiante. Y esa flexibilidad es facilitada a través de la comunicación o diálogo didáctico mediado entre institución y estudiante. Son, en efecto, los medios los que permiten la flexibilidad antes referida [Aretio, 1999].

Al tratar el caso de un nuevo paradigma de enseñanza basado en los entornos de aprendizaje virtual, existe la tendencia a centrarse en las posibilidades positivas. Sin embargo, hay otra cara de esta visión. [Sherman y Judkins, 1992] en su investigación,

tienen que enfrentarse a problemas causados por las diferencias horarias y de cultura que son inherentes a la interacción entre países. Algunos de los problemas que han encontrado eran predecibles, como el hecho de que algunos profesores se sintieran amenazados por la tecnología y el hecho de que los estudiantes quisieran tener contacto real unos con otros. Otras dificultades resultaron más inesperadas. Se produjeron problemas de intransigencia de la propia tecnología y modelos de conducta inesperados, como una exagerada comunicación por correo electrónico.

[Aretio, 1987] sostiene que uno de los problemas que más acusan los alumnos de esta modalidad de enseñanza es el de la soledad y alejamiento del profesor y de los compañeros de estudio. La necesidad de relacionarse con los otros se convierte a veces en determinante para el logro de resultados de aprendizaje. Entonces, el reto de la educación a distancia consiste en mantener un sistema comunicacional no presencial ya que se prescinde de la habitual presencia cara a cara entre profesores y alumnos. Si se logra una buena comunicación el alumno no se sentirá tan solo y contará con la orientación y motivación del profesor y de sus propios compañeros.

Para [Tiffin y Ragasingham, 1997] uno de los principales escollos con los que topa el desarrollo de nuevos sistemas de enseñanza a distancia basados en internet es la ausencia de un marco teórico adecuado. En efecto, por una parte, las instituciones dedicadas tradicionalmente a la enseñanza a distancia encuentran dificultades para cambiar, rápidamente, sus métodos y procedimientos por causa de resistencias en parte internas de la propia institución y en parte en la mentalidad de sus usuarios, sean éstos docentes o estudiantes. Por otra parte, las instituciones que nacen asentadas sobre una base tecnológica innovadora tienen el riesgo de comprometerse excesivamente con una determinada tecnología, hasta el punto de verse en dificultades en el momento en que dicha tecnología pase a ser obsoleta [Tiffin y Ragasingham, 1997].

Lo que distingue la enseñanza a distancia del resto es el carácter diferido en el tiempo de buena parte de las ayudas pedagógicas, contenidas ya en los materiales didácticos [Holmberg, 1989]. Quienes participan en la elaboración de dichos materiales

acostumbran a preocuparse no sólo del contenido sino también, de la disposición de las actividades que deben ayudar al estudiante a relacionar los con sus conocimientos previos. Así, a diferencia de otros procesos de enseñanza en que toda interacción se da en una coincidencia espacio-temporal, aquí se parte del convencimiento que esa coincidencia o no es posible o bien debe reducirse al mínimo imprescindible. Las nuevas tecnologías ayudan decididamente a que los materiales didácticos sean más ricos en las posibilidades de diferir las ayudas pedagógicas y, por supuesto, de diversificarlas de forma individualizada [Pedró, 1997].

La función docente, en este contexto, debe considerarse como la organización de las ayudas pedagógicas y la gestión de los ajustes derivados de la individualización del proceso [Alsina, 1997]. Por consiguiente, la gestión del aprendizaje debe ser considerada el resultado de un proceso en el que participan tanto el docente como el estudiante.

El desarrollo tradicional de los cursos de e-learning es una actividad laboriosa [Herin *et al.*, 2002] en el que el promotor (por lo general el profesor del curso) tiene que elegir los contenidos que se mostrarán, decidir sobre la estructura de los contenidos, y determinar los elementos de contenido más adecuadas para cada tipo de usuario potencial del curso. Debido a la complejidad de estas decisiones, un diseño de una sola vez es poco factible, incluso cuando se hace con cuidado. En su lugar, será necesario en la mayoría de los casos para evaluar y posiblemente modificar los contenidos del curso, la estructura y la navegación basada en la información de uso de los estudiantes, de preferencia, incluso siguiendo un enfoque continuo de evaluación empírica [Ortigosa y Carro, 2003]. Para facilitar esto, se necesitan métodos y herramientas de análisis de datos para observar el comportamiento de los estudiantes y para ayudar a los maestros en la detección de posibles errores, deficiencias y posibles mejoras. El análisis de datos tradicional en e-learning es hipótesis o suposición conducido [Gaudioso y Talavera , 2006] en el sentido que el usuario comienza a partir de una pregunta y explora los datos para confirmar la intuición. Mientras que esto puede ser útil cuando un número moderado de factores y

datos están involucrados, puede ser muy difícil para el usuario encontrar patrones más complejos que se relacionan diferentes aspectos de los datos. Una alternativa al análisis de datos tradicional es utilizar la minería de datos como un enfoque inductivo para descubrir automáticamente información oculta presente en los datos. La minería de datos, en contraste, es el descubrimiento impulsado en el sentido de que la hipótesis se extrae automáticamente de los datos y por lo tanto es impulsado por los datos en lugar de basados en la investigación o inducido por el hombre [Tsantis y Castellani, 2001]. La minería de datos se construye modelos analíticos que descubren patrones y tendencias interesantes de la información de uso de los estudiantes que pueden ser utilizados por el profesor para mejorar el aprendizaje y por supuesto el mantenimiento del estudiante.

La utilización de EVEAs ha crecido exponencialmente en los últimos años, estimulado por el hecho de que ni los estudiantes ni los profesores están obligados a una ubicación específica y que esta forma de educación basada en la computadora es prácticamente independiente de cualquier plataforma de hardware específicos [Brusilovsky and Peylo, 2003]. Las herramientas de colaboración y comunicación se están utilizado cada vez más en contextos educativos así, como resultado, los Entornos de Aprendizaje y Enseñanza Virtuales se instalan cada vez más por las universidades, colegios comunitarios, escuelas, empresas e incluso también son utilizados por instructores individuales con el fin de agregar tecnología web para sus cursos y para complementar los cursos tradicionales cara a cara [Cole, 2005]. Estos sistemas pueden ofrecer una gran variedad de canales y espacios de trabajo para facilitar el intercambio de información y la comunicación entre los participantes en un curso, para permitir que los educadores de distribuir información a los estudiantes, producir material de contenido, preparar las tareas y exámenes, participar en debates, gestionar las clases a distancia y permitir la colaboración aprendizaje con foros, chats, áreas de almacenamiento de archivos, servicios de noticias, etc. Estos entornos de aprendizaje acumulan una gran cantidad de información que es muy valiosa para el análisis del comportamiento de los estudiantes y podrían crear una mina de oro de

datos educativos [Mostow y Beck, 2006]. Se puede grabar cualquier actividad en la que los estudiantes están involucrados, tales como leer, escribir, realizar exámenes, la realización de diversas tareas, e incluso comunicarse con los compañeros [Mostow *et al.*, 2005]. Por lo general, estas plataformas no proporcionan a los educadores las herramientas específicas que les permitan realizar un seguimiento a fondo, la posibilidad de evaluar todas las actividades realizadas por sus alumnos, ni tampoco cuentan con un mecanismo para evaluar la estructura, el contenido del curso, y su eficacia en el proceso de aprendizaje. Aunque algunas plataformas ofrecen herramientas como reportes de gestión, cuando existe un gran número de estudiantes, la tarea de extraer información útil se vuelve difícil [Zorrilla *et al.*, 2005]. Entonces, un área muy prometedora para el logro de este objetivo, es el uso de minería de datos [Zaïane y Luo, 2001].

El EVEA registra todas las instancias de comunicación interpersonales. Tales registros podrían utilizarse como base para modelar el comportamiento de los alumnos de educación a distancia.

En la educación presencial los docentes intentan mejorar sus clases a través de la visualización del rendimiento de los estudiantes por medio de la observación. Dicho método puede ser muchas veces subjetivo e incompleto. También existen registros escritos, como exámenes y ejercicios prácticos en los cuales los docentes se pueden apoyar para mejorar sus clases. También utilizan información de asistencia, información sobre el curso, objetivos de la materia, etc. Una institución educativa posee diversas fuentes de información y variadas [Ma *et al.*, 2000]: bases de datos tradicionales (con información de los estudiantes y de los docentes, de las clases, y la programación de las mismas).

3.2. Pregunta de Investigación

La explotación de información permite identificar conocimiento oculto en los datos registrados. El proceso de aprendizaje está compuesto por una serie de hitos que

permiten al docente identificar puntos críticos para favorecer la comprensión de los temas expuestos. En [Cigliuti *et al*, 2012] se plantea la línea de investigación cuyos objetivos son:

[a] La comprensión de comportamientos de comunidades mediadas por tecnología con énfasis en comunidades educativas.

[b] La formulación de una propuesta de identificación de patrones de comportamientos poblacionales en comunidades educativas mediadas por EVEAs utilizando procesos de explotación de información.

En este contexto, se plantea la siguiente pregunta de investigación:

¿Se puede desarrollar un proceso de explotación de información que ayude a la comprensión del comportamiento de comunidades educativas a partir de los resultados académicos?

4. SOLUCIÓN

En el presente capítulo se abordan los marcos de la solución propuesta (sección 4.1), y luego se presenta la solución a los problemas planteados en el capítulo anterior (sección 4.2).

4.1. Marco de la Solución

Un Proceso de Explotación de Información se define como un grupo de tareas relacionadas lógicamente [García-Martínez *et al.*, 2013a] que, a partir de un conjunto de información con un cierto grado de valor para la organización, se ejecuta para lograr otro, con un grado de valor mayor que el inicial. Adicionalmente, existe una variedad de técnicas de minería de datos, en su mayoría provenientes del campo del aprendizaje automático [García-Martínez *et al.*, 2013b], susceptibles de utilizar en cada uno de estos procesos.

[Britos, 2008] propone los siguientes procesos de explotación de información: descubrimiento de reglas de comportamiento, descubrimiento de grupos, descubrimiento de atributos significativos, descubrimiento de reglas de pertenencia a grupos y ponderación de reglas de comportamiento o de pertenencia. El proceso de descubrimiento de reglas de comportamiento aplica cuando se requiere identificar cuáles son las condiciones para obtener determinado resultado en el dominio del problema. El proceso de descubrimiento de grupos aplica cuando se requiere identificar una partición en la masa de información disponible sobre el dominio de problema. El proceso de ponderación de interdependencia de atributos aplica cuando se requiere identificar cuáles son los factores con mayor incidencia (o frecuencia de ocurrencia) sobre un determinado resultado del problema. El proceso de descubrimiento de reglas de pertenencia a grupos aplica cuando se requiere identificar cuáles son las condiciones de pertenencia a cada una de las clases en una partición desconocida “a priori”, pero presente en la masa de información disponible sobre el dominio de problema. El proceso de ponderación de reglas de

comportamiento o de la pertenencia a grupos aplica cuando se requiere identificar cuáles son las condiciones con mayor incidencia (o frecuencia de ocurrencia) sobre la obtención de un determinado resultado en el dominio del problema, sean estas las que en mayor medida inciden sobre un comportamiento o las que mejor definen la pertenencia a un grupo.

En particular, las metodologías de explotación de información destacan la importancia de la comprensión del negocio y la vinculación entre el problema de negocio con el patrón o regularidad a descubrir en la información a partir del proceso de explotación de información a utilizar. Sin embargo, es un problema abierto determinar de manera sistemática cual es el proceso de explotación de información que se debe aplicar para resolver el problema de negocio [Martins, 2014].

4.2. Solución Propuesta

Como respuesta al problema planteado [Martins, 2014] se propone un procedimiento que permite derivar el proceso de explotación de información a partir del modelado del dominio de negocio (Tabla 4.1).

Los formalismos de ingeniería de conocimiento que se utilizan para modelar el dominio de negocio y de explotación de información son las redes semánticas [Sowa, 1992] y las tablas concepto - atributo - valor [García-Martínez y Britos, 2004].

Con base en los formalismos señalados, se proponen doce formalismos para el modelado del dominio:

- Tabla concepto - atributo - valor del dominio (CAVD): se utiliza para la representación sistemática de los conceptos del dominio, con detalles de los atributos y valores.
- Tabla concepto - relación del dominio (CRD): se utiliza para modelar las relaciones entre los conceptos del dominio.

- Red semántica del modelo de negocio (RSMN): modela el dominio del negocio.
- Tabla concepto - atributo - valor del problema de explotación de información (CAVPEI): es una representación sistemática de los conceptos del problema de explotación de información, con detalles de los atributos y valores.
- Tabla concepto - relación del problema de explotación de información (CRPEI): busca modelar las relaciones entre los conceptos del problema de explotación de información.
- Tabla de entradas/salidas del problema de explotación de información (E/S - PEI): es un listado de entradas/salidas del problema de explotación de información.
- Red semántica del problema de explotación de información (RSPEI): representación sistemática del problema de explotación de información.
- Red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI - DE/S): modela el problema de explotación de información con los datos identificados de entrada y salida.
- Tabla concepto - atributo - valor Integrada (CAVI): representación sistemática integrada de los conceptos del dominio y del problema de explotación de información, con detalles de los atributos y valores.
- Tabla concepto - relación Integrada (CRI): modela de manera integrada las relaciones entre los conceptos del dominio y del problema de explotación de información.
- Red semántica integrada modelo negocio - explotación de información (RSI - MN - EI): representación sistemática integrada del modelo de negocio y del problema de explotación de información.
- Red semántica integrada modelo negocio - explotación de información con datos de entrada y salida identificados (RSI - MN - EI - DE/S): representación

sistemática integrada del modelo de negocio y del problema de explotación de información con los datos identificados de entrada y salida.

Procedimiento de derivación	
Paso 1:	Identificación de conceptos del dominio.
Paso 2:	Identificación de relaciones entre conceptos.
Paso 3:	Conceptualización del dominio.
Paso 4:	Identificación de conceptos y relaciones del problema de explotación de información.
Paso 5:	Conceptualización del problema de explotación de información.
Paso 6:	Identificación de los datos de entradas y resultados del problema de explotación de información.
Paso 7:	Integración de las novaciones que introduce el problema de explotación de información al dominio.
Paso 8:	Integración de los conceptos del dominio y del problema de explotación de información.
Paso 9:	Identificación Integrada de los datos de entrada y objetivo.
Paso 10:	Derivación del proceso de explotación de información.

Tabla 4.1-Procedimiento de derivación del proceso de explotación de información a partir del modelado del dominio de negocio

El procedimiento propuesto por [Martins, 2014] conceptualiza el dominio a partir de la identificación de los conceptos del dominio y las relaciones entre estos conceptos (pasos 1 a 3). En los pasos siguientes (4 a 6) se conceptualiza el problema de explotación de información con base en la identificación de conceptos y relaciones del problema de explotación de información y la identificación de los datos de entradas y resultados del problema de explotación de información. En la representación lograda hasta el momento se producen novaciones que introduce el problema de explotación de información al dominio, los mismos se incorporan a la representación del dominio. En este punto se pueden integrar los conceptos del dominio y del problema de explotación de información; e identificar de manera integral los datos de entrada y objetivo (pasos 7 a 9). A partir de este punto, y con base en los sub pasos del paso 10 que se presentan en la Tabla 4.2, se deriva el proceso de explotación de información.

Detalle del paso 10		
Derivación del proceso de explotación de información		
Subpaso 10.1:	SI	En la Red semántica integrada modelo negocio - explotación de información con datos de entrada y salida identificados (RSI - MN - EI - DE/S) se puede identificar un resultado intermedio.
	ENTONCES	Ir a Subpaso 10.5.
	SINO	Ir a Subpaso 10.2.
Subpaso 10.2:	SI	El objetivo de la red semántica integrada modelo negocio - explotación de información con datos de entrada y salida identificados (RSI - MN - EI - DE/S) es un único nodo cuya arista se identifica con las palabras: pertenece, integra, es, es miembro de, u otras palabras de similar significado.
	ENTONCES	Aplicar proceso descubrimiento de grupos.
	SINO	Ir a Subpaso 10.3.
Subpaso 10.3:	SI	En la red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI - DE/S) se busca detectar los valores que más se destacan o que aparecen con mayor frecuencia del conjunto de datos de salida
	ENTONCES	Aplicar proceso ponderación de interdependencia de atributos.
	SINO	Ir a Subpaso 10.4.
Subpaso 10.4:	SI	En la red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI - DE/S) se busca identificar, a partir del conjunto de datos de salida, el valor de un atributo determinado
	ENTONCES	Aplicar proceso descubrimiento de reglas de comportamiento.
	SINO	Representación errónea.
Subpaso 10.5:	SI	La Red semántica integrada modelo negocio - explotación de información con datos de entrada y salida identificados (RSI - MN - EI - DE/S) tiene un concepto resultado intermedio, para el cual el arco que llega se rotula con las palabras: "pertenece", "integra", "es", "es miembro de" u otras palabras de similar significado
	ENTONCES	Ir a Subpaso 10.6.
	SINO	Representación errónea.
Subpaso 10.6:	SI	El resultado intermedio de la red semántica integrada modelo negocio - explotación de información con datos de entrada y salida identificados (RSI - MN - EI - DE/S), cuya relación se define con las palabras pertenece,

		integra, es, es miembro de, u otras palabras de similar significado, es un nodo existente en la red semántica del modelo de negocio (RSMN).
	ENTONCES	Aplicar proceso ponderación de reglas de pertenencia a grupos (grupos definidos).
	SINO	Ir a Subpaso 10.7.
Subpaso 10.7:	SI	En la Red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI - DE/S) se busca detectar aquellos valores de los atributos pertenecientes al concepto identificado como resultado intermedio, que más se destacan o que aparecen con mayor frecuencia en el conjunto de datos de salida.
	ENTONCES	Aplicar proceso ponderación de reglas de pertenencia a grupos (grupos no definidos).
	SINO	Ir a Subpaso 10.8.
Subpaso 10.8:	SI	En la red semántica del problema de explotación de información con datos de entrada y salida identificados (RSPEI - DE/S) se busca definir el identificador del resultado intermedio, de acuerdo con los atributos de salida
	ENTONCES	Aplicar proceso descubrimiento de reglas de pertenencia a grupos.
	SINO	Representación errónea.

Tabla 4.2. Detalle de los subpasos de derivación del proceso de explotación de información

Ante la falta de herramientas por parte de los educadores para evaluar a los alumnos, las estructuras, el contenido del curso y su eficacia en el proceso de aprendizaje, este trabajo propone como solución un recurso. El mismo permite estudiar el comportamiento de los estudiantes a lo largo de un curso o una carrera, identificando y resolviendo los distintos problemas asociados al aprendizaje y al favorable desarrollo de la materia. De esta manera, los docentes tendrán a su disposición una herramienta que favorece la comprensión del curso y la posibilidad de implementar medidas que les permitan detectar alumnos con dificultades para, así, intervenir de manera temprana con la posibilidad de corregir factores que impacten en el resultado de la cursada.

El proceso propuesto hace uso de los procesos de explotación de información y del de derivación del problema de explotación de información definido mencionados anteriormente.

A continuación se presenta el proceso propuesto compuesto por los siguientes 5 pasos:

PROCESO PROPUESTO	
Paso 1:	Conceptualización del Dominio
	<i>Subpaso 1.1:</i> Identificación de los elementos relevantes del dominio
	<i>Subpaso 1.2:</i> Identificación de las relaciones entre elementos del dominio
	<i>Subpaso 1.3:</i> Representación conceptual del dominio.
Paso 2:	Conceptualización del Problema de Explotación de Información
	<i>Subpaso 2.1:</i> Identificación del problema de negocio y su traducción al problema de explotación de información
	<i>Subpaso 2.2:</i> Identificación de los conceptos pertenecientes al problema de explotación de información
	<i>Subpaso 2.3:</i> Identificación de las relaciones entre elementos del problema de explotación de información
	<i>Subpaso 2.4:</i> Identificación de dependencias entre los elementos conceptuales del problema
	<i>Subpaso 2.5:</i> Representación del problema de explotación de información
Paso 3:	Identificación del Proceso de Explotación de Información
	Derivación del proceso de explotación de información
Paso 4:	Preparación de los Datos
	<i>Subpaso 4.1</i> Selección de los campos asociados al problema de explotación de información
	<i>Subpaso 4.2</i> Conversión de los datos
	<i>Subpaso 4.3</i> Generación de la base de datos
Paso 5:	Implementación
	<i>Subpaso 5.1</i> Implementación
	<i>Subpaso 5.2</i> Interpretación de los resultados

Tabla 4.3. Proceso Propuesto para el Estudio del Comportamiento de Comunidades Educativas

El primer paso se orienta a la comprensión y conceptualización del negocio a analizar (reconociendo los distintos elementos que conforman al estudio del curso a analizar), identificando aquellos conceptos dependientes.

El segundo paso se orienta a la identificación del problema de negocio, entendiendo como problema de negocio a aquel objetivo que se quiere resolver, para el cual se requiere obtener información relevante que soporte la toma de decisión, su traducción

a lenguaje técnico (problema de explotación de información), la identificación de los elementos que lo componen y la conceptualización del mismo.

El tercer paso consiste en identificar el proceso de explotación a utilizar, que es el que define las técnicas de minería de datos a implementar para dar solución al problema de negocio.

El cuarto paso consiste en generar la base de datos y los datos acorde al proceso de explotación de información identificado y de los algoritmos de minería de datos a utilizar, convirtiendo los datos en base a las necesidades de los mismos.

El último paso es el que consiste en implementar e interpretar los resultados obtenidos, aportando información relevante para la solución del problema de negocio previamente identificado.

5. VALIDACIÓN

En este capítulo se describen tres casos de validación. El primero de Identificación de Comportamiento de Recursantes Utilizando EVEAs (sección 5.1), el segundo consiste en la Identificación del Conceptos Críticos de una Cursada (sección 5.2), y por último de Identificación de Contenidos Débilmente Apropriados por Estudiantes (sección 5.3).

5.1. Proceso de Identificación de Comportamiento de Estudiantes Recursantes Utilizando EVEAs

En esta sección se describe el caso de estudio (sección 5.1.1), luego se aplica el proceso (sección 5.1.2), y por último se describe la conclusión (sección 5.1.3).

5.1.1. Descripción del Caso

Desde el año 2007, la Secretaría de Gestión Académica de la Facultad Regional Buenos Aires (UTN-FRBA) lidera un proyecto para la creación de aulas virtuales como apoyatura a la presencialidad, desarrollado en el marco del PROMEI (Programa de Mejoramiento para la Enseñanza de la Ingeniería) [PROMEI, 2014]. En el mismo se involucran distintas asignaturas homogéneas del plan de las Carreras de Ingeniería con la utilización de entornos virtuales, integrándolos paulatinamente al proceso de aprendizaje de los alumnos [Cuzzani, 2014].

Un antecedente desarrollado en paralelo a este proyecto es la experiencia del curso de ingreso, en donde se pusieron en juego usos educativos de las TIC que se desarrollaron en el Entorno Virtual de Enseñanza y Aprendizaje (EVEA) que proveyó la Universidad para cada Facultad Regional.

A inicios del año 2008, profundizando esta línea de acción y atendiendo a la Ordenanza de la Universidad Tecnológica Nacional N° 1129 “Cursado Intensivo” [Ord1129, 2014], la Secretaría de Gestión Académica de la F.R.B.A. presentó una nueva propuesta denominada “Mejoramiento de la calidad educativa favoreciendo la

retención estudiantil mediante el uso de las nuevas tecnologías” cuyo objetivo General fue el de profundizar las acciones de Educación a Distancia en la Carrera de Grado en el Marco de la Ordenanza antes mencionada. Dentro de los objetivos específicos de dicha presentación se encontraban [Cuzzani, 2014]:

- Implementar el régimen de cursado intensivo para la aprobación de asignaturas de la carrera de grado en el ámbito de la Facultad Regional Buenos Aires.
- Integrar significativamente las nuevas tecnologías a la formación aprovechando las potencialidades que ofrecen.
- Capacitar a docentes/tutores en el uso de las herramientas de gestión del campus virtual de la Facultad.
- Preparar material didáctico apto para la intervención pedagógica en entornos virtuales.

Bajo estas circunstancias la Secretaría de Gestión Académica junto con la Cátedra de Matemática Discreta, asignatura del primer Nivel de la Carrera Ingeniería en Sistemas de Información, se abocaron a la generación del proyecto del “Cursado Intensivo de Matemática Discreta” [Cuzzani, 2014].

La asignatura involucrada en esta investigación se dicta en forma cuatrimestral para la Carrera Ingeniería en Sistemas de Información que dicta la Facultad Regional Buenos Aires. Se pensó, en el momento de la implementación en cursos que se desarrollen con la modalidad a distancia, pero que tienen previsto instancias presenciales, dichas instancias tienen la siguiente característica:

- a) No obligatoria: clases de consulta
- b) Obligatoria: encuentro para la toma de parciales y recuperatorios.

Con la puesta en marcha de este Proyecto, la Facultad apunta a resolver, entre otros, problemas relacionados con la deserción y ofrecerle a los alumnos que tienen que recurrir dicha asignatura, la posibilidad de cursarla bajo otra modalidad.

En la actualidad, para poder cursar de forma intensiva (Escuela de Verano presencial) una asignatura que la Facultad ofrece, los alumnos deben cumplir con las siguientes condiciones [Cuzzani, 2014]:

- a) Posean la condición de alumno regular al momento de la inscripción.
- b) Hayan sido cursantes de las asignaturas en cuestión en el Ciclo Lectivo que se realiza la inscripción a la Escuela de Verano en la UTN-FRBA y se hayan presentado, al menos, a dos instancias de evaluación.
- c) No hayan sido dados de baja por inasistencias (estén incluidos en el Acta de Trabajos Prácticos - TPA).
- d) Además, podrán inscribirse quienes no cumpliendo con los puntos b y c, hayan rendido por cuarta vez mal el final correspondiente (lo que los obligaría según Reglamento de Estudios a recurrar la asignatura).

Las condiciones antes enunciadas han sido estipuladas fundamentalmente por temas presupuestarios, esto lleva aparejado que un número importante de alumnos no puedan cursarla, aunque sea su deseo hacerlo.

El hecho de poder hacer la cursada en forma semi-presencial o a distancia lleva implícito, al entender de la Institución, varios elementos favorables a tener en cuenta:

- a) La cursada podrá comenzar antes y así dedicarle más tiempo que el actual curso. Esto es así dado que no será necesario el trabajo de demasiados docentes, teniendo en cuenta que en el período de verano se encuentran de vacaciones.
- b) Como ya se indicó, la institución ha establecido normas que restringen la inscripción, con el objeto de minimizar la cantidad de cursos a raíz de dos problemas centrales:
 - 1- No es fácil conseguir los docentes que dicten este curso en época estival por una remuneración, además, que es menor a la establecida para una cursada estándar. Todo ello a raíz de cuestiones presupuestarias.

2- El recurso "aulas" es limitado dado que conviven en el mismo momento tres actividades académicas diferentes: los cursos de ingreso a las diferentes carreras, los exámenes finales y la Escuela de Verano. Con esta nueva propuesta, este tipo de restricciones no serían necesarias, dado que optimizaríamos tanto el recurso docente como el de espacio físico y, por ende, más alumnos podrían aprovechar esta oportunidad. Los alumnos podrían organizar su horario de estudio de mejor forma, ya que en la actualidad para este tipo de cursos hay un solo horario de cursada.

- c) Teniendo en cuenta que los alumnos ya cursaron la asignatura en cuestión, cabe suponer que hay temas en los que solo necesitan repasar y otros a los que les debe dedicar más tiempo.
- d) Es un alto número de alumnos el que año tras año preguntan sobre la posibilidad de cursar estas asignaturas (la que conforman la escuela de verano) en la modalidad a distancia.

Por lo antes expuesto, el curso intensivo estaría dirigido a todos los alumnos que quieran cursar mediante esta modalidad y hayan sido cursantes, en este caso particular de la asignatura Matemática Discreta, en el Ciclo Lectivo que se realiza la inscripción en la UTN-FRBA o hayan rendido por cuarta vez mal el final correspondiente [Cuzzani, 2014].

Entre los objetivos de propuesta podemos enumerar:

Objetivo general:

- a) Implementar estrategias de innovación pedagógica basada en entornos virtuales.
- b) Ofrecer una nueva modalidad de cursada, atendiendo en una primera instancia a alumnos de los dos primeros niveles de estudio apuntando al área de matemática.

Objetivos específicos:

- a) Desarrollar entornos virtuales de aprendizaje para el dictado de asignaturas del Ciclo General de Conocimientos Básicos, como ser, Matemática Discreta.

- b) Capacitar a los docente para trabajar bajo esta modalidad
- c) Capacitar a los tutores para acompañar el proceso de aprendizaje de los alumnos.
- d) Generar materiales multimediales que acompañen el proceso en el entorno virtual.
- e) Crear herramientas de gestión acordes a este proceso de enseñanza-aprendizaje.
- f) Desarrollar en los alumnos capacidades inherentes al compromiso que deben asumir con su propia formación.

La materia se encuentra dividida en unidades temáticas en las cuales el alumno deberá ir presentando trabajos prácticos para cada una de ellas. Un alumno aprueba la cursada a distancia si cumple con la entrega del 75% de los trabajos prácticos. La materia en esta modalidad fue cursada desde el año 2010 al 2012 por un total de 363 alumnos.

La materia está implementada en un EVEA bajo la plataforma Moodle la cual cuenta con módulos donde el alumno podrá realizar diferentes actividades:

- 1- Bajar el material de lectura: el alumno podrá bajar el material de lectura correspondiente a las diferentes unidades de la materia.
- 2- Subir trabajos prácticos: el alumno podrá subir los trabajos prácticos para cada una de las unidades (actividades). Los trabajos prácticos podrán ser subidos más de una vez si el docente lo requiere.
- 3- Leer foros: el alumno podrá acceder a leer los foros visualizando los diferentes temas de discusión. Una vez adentro de algún foro podrá entrar a leer algún tema de discusión específico del foro. Existen tres tipos de foros:
 - 3.1. Aquellos relacionados con las diferentes unidades de las materias.
 - 3.2. El foro de novedades, donde el docente publicará las novedades referentes a la materia.
 - 3.3. Foro de alumnos, donde los alumnos podrán publicar cualquier tipo de consulta dirigida hacia los docentes o hacia el resto de los alumnos.

- 4- Crear nuevos temas de discusión en los foros. El alumno podrá crear nuevos temas de discusión en el foro 3.1 y el foro 3.3.
- 5- Agregar comentarios temas de discusión ya creados: el alumno podrá agregar algún comentario sobre algún tema de discusión de algún foro. Lo podrá realizar sobre los foros del tipo 3.1 y 3.3”

Las variables relacionadas con cada una de las actividades descriptas anteriormente se ven reflejadas en la tabla 5.1

Campo	Descripción
Id_curso	Es el número que identifica un único curso.
id_alumno	Es el número que identifica a un único alumno.
ingreso_curso	Cantidad de ingresos al EVEA.
bajo_material	Indica si el alumno bajó material de lectura de alguna de las unidades de la materia.
actividad_tp	Indica si el alumno subió actividades de tp.
subio_tp	Indica si subió más de un tp para alguna actividad.
vio_foro_unidad	Indica si el alumno entró a alguno de los foros correspondientes a las unidades de la materia.
vio_discusion_unidad	Indica si el alumno entró a ver alguna discusión particular de algún foro correspondiente a las unidades.
comento_discusion_unidad	Indica si el alumno realizó algún comentario en los foros correspondientes a las unidades de la materia.
creo_discusion_unidad	Indica si el alumno creó algún tema de discusión en un foro correspondiente a una unidad
vio_foro_alumno	Indica si el alumno entró a ver el foro de alumno.
vio_discusion_alumno	Indica si el alumno entró a ver alguna discusión particular del foro de alumnos.
comento_discusion_alumno	Indica si el alumno comentó alguna discusión del foro de alumnos.
creo_discusion_alumno	Indica si el alumno creó algún tema de discusión en el foro de alumnos.
vio_foro_novedades	Indica si el alumno entró a ver el foro de novedades.
vio_discusion_novedades	Indica si el alumno entró a ver alguna discusión particular del foro de novedades.
aprueba_cursada	Indica si el alumno aprobó la cursada.

Tabla 5.1-Listado de Campos Registrados

Cada uno de los campos previamente mencionados, tienen como valor posible los caracteres “SI” y “NO”, con excepción del campo “Id_curso” e “Id_alumno”, los cuáles son de tipo numérico.

El problema de negocio presentado consiste en estudiar a los alumnos con respecto a su participación en el curso y la relación que la misma pudiera tener con la aprobación o desaprobación del curso.

5.1.2. Aplicación del Proceso

Paso 1: Conceptualización del dominio

La conceptualización del dominio del negocio se compone de 3 subpasos:

- Subpaso 1.1. Identificación los elementos relevantes del dominio

Para la identificación de los elementos relevantes del dominio de negocio, se hace uso de las técnicas propuestas en [Martins, 2014]. Los formalismos resultantes de ejecutar dichas técnicas en el caso de estudio propuesto son la tabla término-categoría-definición del dominio (tabla 5.2) y la tabla concepto-atributo-relación-valor del dominio (tabla 5.3).

- Subpaso 1.2. Identificación de las relaciones entre elementos del dominio

Para la identificación de las relaciones entre elementos del dominio de negocio, se hace uso la técnica propuesta en [Martins, 2014], la tabla concepto-relación del dominio. Como resultado de aplicar la técnica tabla concepto-relación del dominio al caso de estudio propuesto se obtiene la tabla 5.4.

- Subpaso 1.3. Representación conceptual del dominio

Para la conceptualización del dominio del negocio, se hace uso de la técnica red semántica del modelo de negocio propuesto en [Martins, 2014]. El resultado de aplicar dicha técnica al caso de estudio propuesto, se presenta en la figura 5.1

Paso 2: Conceptualización del problema de explotación de información

La conceptualización del problema de explotación de información está compuesta por 5 subpasos:

- Subpaso 2.1. Identificación del problema de negocio y su traducción al problema de explotación de información.

Término	Categoría	Definición
actividad_tp	Atributo	Indica si el alumno subió actividades de tp
Alumno	Concepto	Persona que realiza un curso
bajada_material	Atributo	Indica si el alumno bajó material de lectura de alguna de las unidades de la materia.
comentario_discusion_unidad	Atributo	Indica si el alumno realizó algún comentario en foros correspondientes a unidades
comentarios_discusion_alumno	Atributo	Indica si el alumno comentó alguna discusión del foro de alumnos
creacion_discusion_alumno	Atributo	Indica si el alumno creó algún tema de discusión en el foro de alumnos
creacion_discusion_unidad	Atributo	Indica si el alumno creó algún tema de discusión en un foro correspondiente a una unidad
Cursada	Atributo	Indica si el alumno aprobó la cursada
Curso	Concepto	Materia realizada en un periodo del año
Id_alumno	Atributo	Identificación unívoca de cada alumno
Id_curso	Atributo	Identificación unívoca de cada curso
observaciones_discusion_foro_alumno	Atributo	Indica si el alumno entró a ver alguna discusión particular del foro de alumnos
observaciones_discusion_foro_novedades	Atributo	Indica si el alumno entró a ver alguna discusión particular del foro de novedades
observaciones_discusion_foro_unidad	Atributo	Indica si el alumno entró a ver alguna discusión particular de algún foro correspondiente a las unidades
observaciones_foro_alumno	Atributo	Indica si el alumno entró a ver el foro de alumno
observaciones_foro_novedades	Atributo	Indica si el alumno entró a ver el foro de novedades
Observaciones_foro_unidad	Atributo	Indica si el alumno entró a alguno de los foros correspondientes a las unidades de la materia.
Obtuvo	Relación	El alumno obtuvo como resultado de la cursada
Realiza	Relación	el alumno realiza un curso
Realizó	Relación	El alumno realizó la descarga de algún material de lectura del curso
Realizó	Relación	El alumno realizó más de una subida de alguna actividad del curso
Realizó	Relación	El alumno realizó visualizaciones de algún foro de las unidades de la materia
Realizó	Relación	El alumno realizó visualizaciones de alguna discusión de algún foro correspondiente a una unidad
Realizó	Relación	El alumno realizó algún comentario en un foro de unidad
Realizó	Relación	El alumno realizó la creación de algún tema de discusión en el algún foro de unidad
Realizó	Relación	El alumno realizó visualizaciones en el foro de alumno
Realizó	Relación	El alumno realizó visualizaciones en alguna discusión del foro de alumnos
Realizó	Relación	El alumno realizó comentarios en alguna discusión del foro de alumnos
Realizó	Relación	El alumno realizó la creación de algún tema de discusión en el foro de alumnos
Realizó	Relación	El alumno realizó visualizaciones en el foro de novedades
Realizó	Relación	El alumno realizó visualizaciones en alguna discusión del foro de novedades
subida_tp	Atributo	Indica si subió más de un tp para alguna actividad
Subió	Relación	El alumno subió actividades del curso

Tabla 5.2 - Término-Categoría-Definición del dominio aplicada al caso de estudio

Concepto	Atributo	Relación	Valor
Curso	Id curso	identifica	Numérico
	Id alumno	identifica	Numérico
Alumno	bajada material	realizó	SI - NO
	actividad_tp	subió	SI - NO
	subida_tp	realizó	SI - NO
	observaciones foro unidad	realizó	SI - NO
	observaciones_discusion foro alumno	realizó	SI - NO
	comentario_discusion_unidad	realizó	SI - NO
	creacion_discusion_unidad	realizó	SI - NO
	observaciones_discusion foro novedades	realizó	SI - NO
	observaciones_discusion foro unidad	realizó	SI - NO
	comentarios_discusion_alumno	realizó	SI - NO
	creacion_discusion_alumno	realizó	SI - NO
	observaciones foro alumno	realizó	SI - NO
	observaciones foro novedades	realizó	SI - NO
	cursada	obtuvo	SI - NO

Tabla 5.3-Concepto-Atributo-Relación-Valor del dominio aplicado al caso de estudio

Concepto	Concepto asociado	relación	Descripción
Curso	Alumno	realiza	el alumno realiza un curso

Tabla 5.4-Concepto-Relación del dominio aplicado al caso de estudio

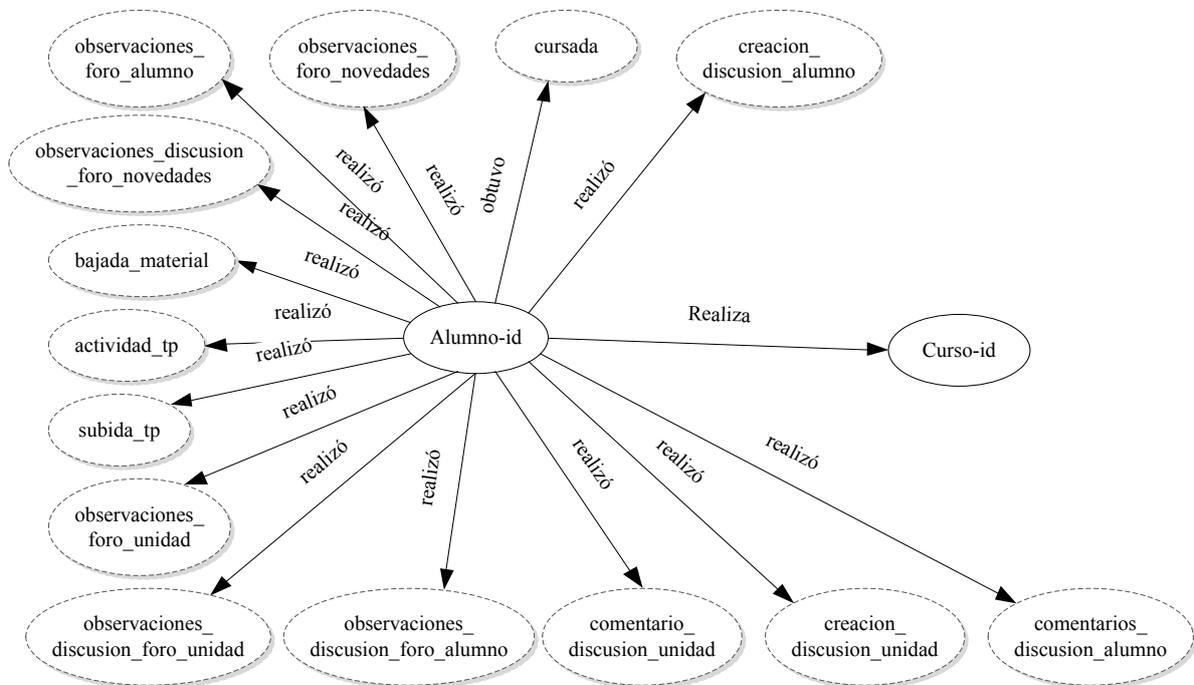


Figura 5.1-Red Semántica del Modelo de Negocio aplicada al caso de estudio

A partir del problema de negocio definido en la descripción del caso de estudio (sección 5.1), se obtiene como problema de explotación de información:

“Identificar aquellas características relevantes del alumno que definen la aprobación o no del curso”.

- Subpaso 2.2. Identificación de los conceptos pertenecientes al problema de explotación de información

A partir del problema de explotación de información se procede a seleccionar y determinar cuáles elementos están relacionados con el mismo. Para ello se implementa las técnicas tabla Término-Categoría-Definición del Problema de Explotación de Información y Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información, propuestas en [Martins, 2014]. El resultado de aplicar cada una de ellas al caso de estudio propuesto, se presenta en las tablas 5.5 y 5.6 respectivamente.

- Subpaso 2.3. Identificación de las relaciones entre elementos del problema de explotación de información

Para la identificación de las relaciones entre elementos del problema de explotación de información, se implementa la técnica propuesta en [Martins, 2014]: la tabla concepto-relación del problema de explotación de información. El resultado de aplicar dicha tabla al caso de estudio propuesto se presenta en la tabla 5.7.

- Subpaso 2.4. Identificación de dependencias entre los elementos conceptuales del problema

El objetivo de este paso es identificar aquellos elementos que presenten dependencias cognitivas con otros elementos los cuales brinden información redundante o afecten la comprensión del conocimiento

Término	Categoría	Descripción
Alumno	Concepto	Persona que realiza un curso
bajada_material	Atributo	Indica si el alumno bajó material de lectura de alguna de las unidades de la materia.
Características relevantes	Concepto	Las características relevantes definen la aprobación de la cursada
comentario_discusion_unidad	Atributo	Indica si el alumno realizó algún comentario en los foros correspondientes a las unidades de la materia
comentarios_discusion_alumno	Atributo	Indica si el alumno comentó alguna discusión del foro de alumnos
creacion_discusion_alumno	Atributo	Indica si el alumno creó algún tema de discusión en el foro de alumnos
creacion_discusion_unidad	Atributo	Indica si el alumno creó algún tema de discusión en un foro correspondiente a una unidad
Cursada	Atributo	Indica si el alumno aprobó la cursada
Id_alumno	Atributo	Identificación unívoca de cada alumno
Id_características_relevantes	Atributo	Identificación unívoca de las características
observaciones_discusion_foro_alumno	Atributo	Indica si el alumno entró a ver alguna discusión particular del foro de alumnos
observaciones_discusion_foro_novedades	Atributo	Indica si el alumno entró a ver alguna discusión particular del foro de novedades
observaciones_discusion_foro_unidad	Atributo	Indica si el alumno entró a ver alguna discusión particular de algún foro correspondiente a las unidades
observaciones_foro_alumno	Atributo	Indica si el alumno entró a ver el foro de alumno
observaciones_foro_novedades	Atributo	Indica si el alumno entró a ver el foro de novedades
Observaciones_foro_unidad	Atributo	Indica si el alumno entró a alguno de los foros correspondientes a las unidades de la materia.
Según	Relación	Las características relevantes del alumno se definen según el resultado de la cursada
Subconjunto	Relación	bajada del material es un elemento del conjunto de características relevantes
Subconjunto	Relación	Observar discusiones del foro de alumnos es un elemento del conjunto de características relevantes
Subconjunto	Relación	Observar discusiones en el foro de unidad es un elemento del conjunto de características relevantes
Subconjunto	Relación	Observar discusiones del foro de unidades es un elemento del conjunto de características relevantes
Subconjunto	Relación	Observar el foro alumno es un elemento del conjunto de características relevantes
Subconjunto	Relación	Observar el foro novedades es un elemento del conjunto de características relevantes
Subconjunto	Relación	Observar el foro de las unidades es un elemento del conjunto de características relevantes
Subconjunto	Relación	Comentario en foro de unidad es un elemento del conjunto de características relevantes
Subconjunto	Relación	Comentarios sobre discusiones del foro de alumnos es elemento conjunto de características relevantes
Subconjunto	Relación	Crear un tema de discusión en foro de alumnos es elemento del conjunto de características relevantes
Subconjunto	Relación	Crear un tema de discusión en foro de unidad es un elemento del conjunto de características relevantes
Subconjunto	Relación	Subir más de un tp es un elemento del conjunto de características relevantes
subida_tp	Atributo	Indica si subió más de un tp para alguna actividad

Tabla 5.5-Término-Categoría-Definición del Problema de Explotación de Información aplicada al caso de estudio

oculto en los datos. Para ello se propone la tabla de dependencias cuyo objetivo es identificar las dependencias existentes entre los datos, y eliminar aquellas que se consideren perjudiciales para el desarrollo del proceso. La tabla de dependencias está compuesta por tres columnas: en la primera se registra el elemento que no posea dependencia, en la segunda el elemento dependiente del elemento anteriormente registrado, y en la tercera columna se registra la necesidad o no de eliminar el elemento dependiente. En la tabla 5.8 se presentan los elementos identificados en el caso de estudio propuesto.

- Subpaso 2.5. Representación del problema de explotación de información

Para la conceptualización del problema de explotación de información, se hace uso de la técnica red semántica del problema de explotación propuesta en [Martins, 2014]. El resultado de aplicar dicha técnica al caso de estudio propuesto, se presenta en la figura 5.2.

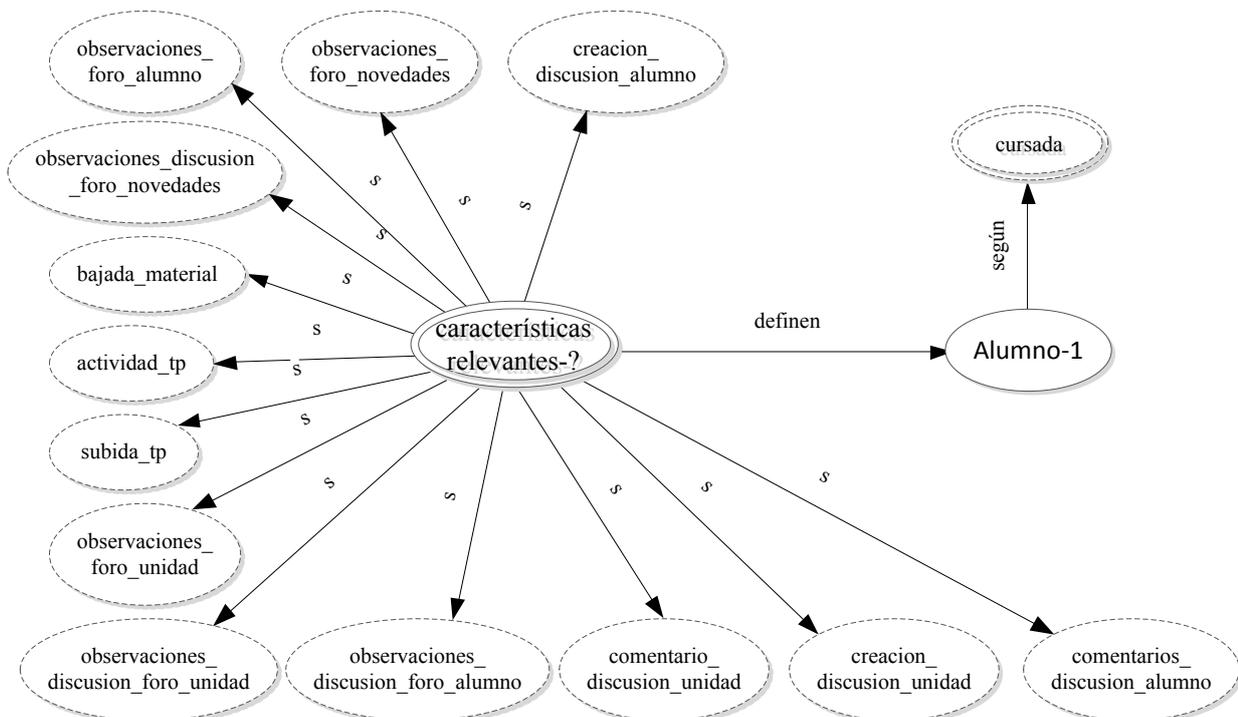


Figura 5.2-Red Semántica del Problema de Explotación de Información aplicada al caso de estudio

Concepto	Atributo	Relación	Entrada/ Salida	Valor
Características relevantes	Identificador	identifica	Salida	Numérico
	bajada_material	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	actividad_tp	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	subida_tp	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	observaciones_foro_unidad	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	observaciones_discusion_foro_alumno	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	comentario_discusion_unidad	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	creacion_discusion_unidad	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	observaciones_discusion_foro_novedades	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	observaciones_discusion_foro_unidad	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	comentarios_discusion_alumno	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	creacion_discusion_alumno	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	observaciones_foro_alumno	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
	observaciones_foro_novedades	subconjunto	<i>Elemento Conjunto Salida</i>	SI - NO
Alumno	Id_alumno	identifica	<i>No relevante</i>	Numérico
	cursada	según	Entrada	SI - NO

Tabla 5.6-Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información aplicada al caso de estudio

Concepto	Concepto asociado	Relación	Descripción
Características relevantes	Alumno	definen	Las características relevantes definen la cursada del alumno

Tabla 5.7-Concepto-Relación del Problema de Explotación de Información aplicada al caso de estudio

Elemento	Elemento dependiente	Eliminar
actividad_tp	subida_tp	NO

Tabla 5.8- Dependencias aplicadas al caso de estudio.

Paso 3: Identificación del proceso de explotación de información

A partir de las conceptualizaciones del dominio del negocio y del problema de explotación previamente presentadas, se establece el proceso de explotación de información a implementar, a través de aplicar el algoritmo de derivación del proceso de explotación de información

propuesto en [Martins, 2014]. El proceso de explotación de información obtenido es el proceso de Ponderación de Interdependencia de atributos.

Paso 4: Preparación de los datos

La preparación de los datos para la ejecución del proceso de explotación de información previamente detectado, consta de 3 subpasos:

- Subpaso 4.1: Selección de los campos asociados al problema de explotación de información

A partir de conceptualizar los distintos elementos involucrados se seleccionan aquellos campos asociados al problema de explotación de información previamente establecido, obteniéndose el listado de datos que conformarán la base de datos a implementar. En el caso de estudio propuesto, los campos a utilizar son:

- bajo_material
- actividad_tp
- subio_tp
- vio_foro_unidad
- vio_discusion_unidad
- comento_discusión_unidad
- creo_discusion_unidad
- vio_foro_alumno
- vio_discusion_alumno
- comento_discusión_alumno
- creo_discusion_alumno
- vio_foro_novedades
- vio_discusion_novedades

- aprueba_cursada

- Subpaso 4.2: Conversión de los datos

A partir de los algoritmos de minería de datos seleccionados, en base al proceso de explotación de información determinado, puede ser necesario convertir el tipo de dato de alguno de los campos seleccionados. Para el caso de estudio propuesto, no fue necesario convertir ningún dato.

- Subpaso 4.3: Generación de la base de datos

Una vez identificados y convertidos los datos, es necesario realizar un proceso que integre todos los campos seleccionados en una única y misma fuente de datos. El resultado de aplicar el subpaso actual es la generación de la base de datos en la cual se implementará el proceso de explotación de información.

Paso 5: Implementación

El paso implementación está conformado por 2 subpasos:

- Subpaso 5.1: Implementación

Este subpaso tiene como objetivo implementar el proceso de explotación de información identificado (mediante la ejecución de los algoritmos de minería de datos), en la base de datos generada en el paso anterior.

- Subpaso 5.2: Interpretación de los resultados

A partir de los resultados obtenidos en el subpaso anterior, se realiza una interpretación de los mismos cuyo objetivo es identificar información relevante, descartando aquella que no sea fiable, y presentar la información obtenida. La información obtenida del caso de estudio presentado es:

Los atributos de mayor incidencia para la desaprobación del curso son:

- comento_discusion_unidad = NO, con un 91% de incidencia.

- `vio_discusion_novedades` = NO, con un 86% de incidencia.
- `comento_discusion_alumno` = NO, con un 81% de incidencia.
- `actividad_tp` = NO, con un 73% de incidencia.

Adicionalmente, los atributos de mayor incidencia para la aprobación del curso son:

- `actividad_tp` = SI, con un 100% de incidencia.
- `vio_discusion_unidad` = SI, con un 100% de incidencia.
- `vio_foro_unidad` = SI, con un 100% de incidencia.

Es decir, que los atributos que se relacionan más fuertemente con la no aprobación de la cursada por parte del alumno son: si no comenta en algún foro de alguna unidad, si no observa las discusiones en los foros de novedades, sino comenta en las discusiones en los foros de alumnos y sino realiza las actividades. Mientras que los que tienen mayor relación con la aprobación de la cursada son: que realice las actividades, que vea las discusiones y los foros de las unidades.

5.1.3. Conclusión del Caso de Validación

En el presente caso de estudio se ha propuesto un proceso basado en explotación de información orientado a encontrar patrones sobre el comportamiento de estudiantes que cursan asignaturas universitarias mediadas por EVEA.

Se ha ilustrado el uso del proceso propuesto con una prueba de concepto en el dominio de alumnos que recursan en la modalidad virtual una materia del área de ciencias básicas cuya primera cursada fue presencial. La población estudiada estuvo integrada por 363 estudiantes, que al término del curso virtual aprobaron 31,95%, lo desaprobaron el 41,32%, y lo abandonaron el 26,72%.

Los resultados obtenidos son intuitivamente válidos y no presentan indicios que contraríen las creencias del equipo docente sobre las condiciones de aprobación del

curso en esta modalidad. Sin embargo, con relación a las condiciones de desaprobación y considerando que la población estudiada es pequeña, el hecho que los atributos de mayor incidencia para la desaprobación presenten porcentajes sensiblemente por debajo del 100% se considera un indicador que permite postular que la replicación del proceso sobre una población por lo menos mayor en un orden de magnitud, puede hacer emerger atributos específicos que permitan refinar en distintos grupos las causales de desaprobación.

Los resultados derivados de este caso de estudio dan indicios que el proceso propuesto constituye una nueva herramienta de diagnóstico en el área; y si bien los resultados obtenidos en este estadio son consistentes, sólo sirven para ilustrar la aplicación de la solución propuesta.

5.2. Identificación de Conceptos Críticos de una Cursada

En esta sección se describe el caso de estudio (sección 5.2.1), luego se aplica el proceso (sección 5.2.2), y por último se describe la conclusión (sección 5.2.3).

5.2.1. Descripción del Caso

La carrera Ingeniería en Sistemas de Información de la de la Universidad Tecnológica Nacional Facultad Regional Buenos Aires (UTN-FRBA) tiene un programa académico el cual divide la cursada de materias a lo largo de cinco años. Análisis de Sistemas [DISI, 2014] es una de las materias de segundo año y troncales de la carrera la cual se focaliza en la enseñanza de la herramienta: diagrama de Flujo de Datos (DFD). El aprendizaje de esta herramienta para el alumno resulta dificultosa generando una cantidad de desaprobados en los exámenes parciales, incluso generando a veces, que el alumno termine recursando la materia. Por tal motivo se desea describir la calificación final de los alumnos, en base a los resultados obtenidos en cada una de las áreas de evaluación. Cuando se evalúa al alumno, para aprobar DFD el alumno debe tener bien el Diagrama de Contexto, la Tabla de Eventos y el

DFD propiamente dicho. Se registra un identificador por alumno, además de los atributos de evaluación vinculados que se identifican son (si el alumno es capaz de): definir alcance sistema, identificar entidades, identificar entradas, identificar salidas, identificar funciones, identificar activación, identificar respuestas, graficar procesos, asignar activación, asignar entradas, asignar salidas, definir demoras e identificar comunicación entre procesos. Los atributos pueden tomar los valores: bien (B), Regular (R) y Mal (M). Cada uno de los atributos de evaluación mencionados anteriormente corresponde a alguna de las herramientas que conforman el DFD. Por último se registra un campo que define el resultado de la cursada del alumno, cuyos valores son Aprobó (A) o Desaprobó (D) la cursada. La base de datos cuenta con 101 casos para su estudio.

5.2.2. Aplicación del Proceso

Paso 1: Conceptualización del dominio

La conceptualización del dominio del negocio se compone de 3 subpasos:

- Subpaso 1.1. Identificación los elementos relevantes del dominio.

Los formalismos resultantes de ejecutar el paso actual son las tablas término-categoría-definición del dominio y concepto-atributo-relación-valor del dominio (tabla 5.9 y 5.10 respectivamente).

- Subpaso 1.2. Identificación de las relaciones entre elementos del dominio.

La ejecución de la técnica tabla concepto-relación del dominio en el caso de estudio propuesto no identifica relaciones, dado que solo existe un único concepto.

- Subpaso 1.3. Representación conceptual del dominio

El resultado de aplicar dicha técnica al caso de estudio, se presenta en la figura 5.3.

	Término	Categoría	Definición
1	Alumno	Concepto	Estudiante de la materia Análisis de Sistemas
2	Asigna las entradas y salidas a cada proceso	Atributo	Posibilidad del alumno de Asignar las entradas y salidas de cada proceso en el DFD
3	Define las demoras	Atributo	Posibilidad del alumno de definir las demoras
4	es capaz de	Relación	El alumno es capaz de identificar el alcance del sistema
5	es capaz de	Relación	El alumno es capaz de Identifica entidades externas
6	es capaz de	Relación	El alumno es capaz de Identifica entradas y salidas
7	es capaz de	Relación	El alumno es capaz de Puede identificar funciones?
8	es capaz de	Relación	El alumno es capaz de Identifica la forma de activación
9	es capaz de	Relación	El alumno es capaz de Identifica las respuestas
10	es capaz de	Relación	El alumno es capaz de Grafica los procesos de nivel 1 según TE
11	es capaz de	Relación	El alumno es capaz de Asigna las entradas y salidas a cada proceso
12	es capaz de	Relación	El alumno es capaz de Marca la activación de los procesos
13	es capaz de	Relación	El alumno es capaz de Identifica la comunicación entre procesos
14	es capaz de	Relación	El alumno es capaz de Define las demoras
15	Grafica los procesos de nivel 1 según TE	Atributo	Posibilidad del alumno de graficar los procesos de nivel 1 según la TE
16	Identifica	Relación	El alumno se identifica con un código único
17	Identifica el alcance del sistema	Atributo	Posibilidad del alumno de identificar el alcance del sistema
18	Identifica entidades externas	Atributo	Posibilidad del alumno de identificar las entidades externas que interactúan con el sistema
19	Identifica entradas y salidas	Atributo	Posibilidad del alumno de identificar las entradas y salidas del sistema
20	Identifica la comunicación entre procesos	Atributo	Posibilidad del alumno de identificar la comunicación entre los procesos
21	Identifica la forma de activación	Atributo	Posibilidad del alumno de identificar la forma en la cual se activa el sistema
22	Identifica las respuestas	Atributo	Posibilidad del alumno de identificar las respuestas que el sistema provee a las entidades externas
23	identificador	Atributo	Código Único que permite identificar a un alumno
24	Marca la activación de los procesos	Atributo	Posibilidad del alumno de identificar la activación de los procesos
25	Obtuvo	Relación	El alumno obtuvo un resultado
26	Puede identificar funciones?	Atributo	Posibilidad del alumno de identificar las funcionalidades del sistema
27	Resultado	Atributo	Calificación que obtuvo el alumno

Tabla 5.9-Término-Categoría-Definición del dominio aplicada al caso de estudio

Concepto	Atributo	Relación	Valor
Alumno	identificador	identifica	Numérico
	Identifica el alcance del sistema	es capaz de	B - R - M
	Identifica entidades externas	es capaz de	B - R - M
	Identifica entradas y salidas	es capaz de	B - R - M
	Puede identificar funciones?	es capaz de	B - R - M
	Identifica la forma de activación	es capaz de	B - R - M
	Identifica las respuestas	es capaz de	B - R - M
	Grafica los procesos de nivel 1 según TE	es capaz de	B - R - M
	Asigna las entradas y salidas a cada proceso	es capaz de	B - R - M
	Marca la activación de los procesos	es capaz de	B - R - M
	Identifica la comunicación entre procesos	es capaz de	B - R - M
	Define las demoras	es capaz de	B - R - M
	Resultado	Obtuvo	D - A

Tabla 5.10-Concepto-Atributo-Relación-Valor del dominio aplicada al caso de estudio

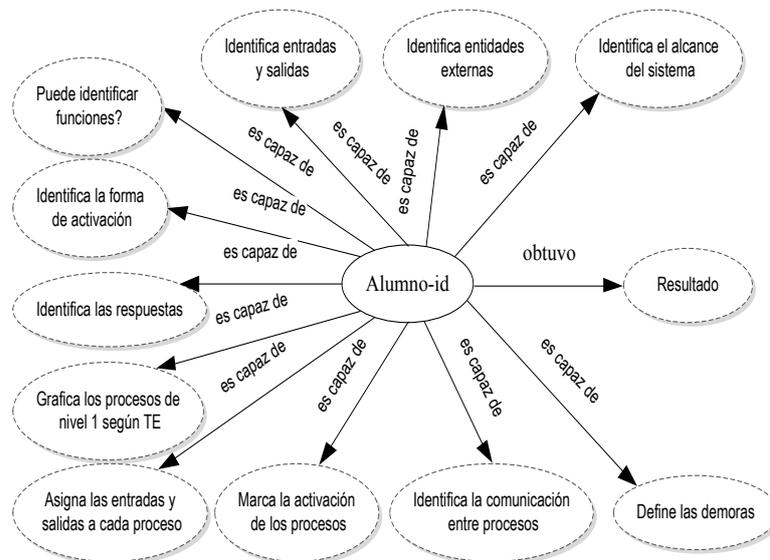


Figura 5.3-Representación Conceptual del Dominio

Paso 2: Conceptualización del problema de explotación de información

La conceptualización del problema de explotación de información está compuesta por 5 subpasos:

- Subpaso 2.1. Identificación del problema de negocio y su traducción al problema de explotación de información

A partir del problema de negocio definido en la sección 5.1, se define el siguiente Problema de Explotación de Información: “Identificar las

reglas que describan el resultado del alumno en base a las condiciones de evaluación.”.

- Subpaso 2.2. Identificación de los conceptos pertenecientes al problema de explotación de información

A partir del problema de explotación de información se procede a seleccionar y determinar cuáles elementos están relacionados con el mismo, implementando las tablas Término-Categoría-Definición del Problema de Explotación de Información y Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información (tablas 5.11 y 5.12 respectivamente).

- Subpaso 2.3. Identificación de las relaciones entre elementos del problema de explotación de información

Para la identificación de las relaciones entre elementos del problema de explotación de información, se implementa la técnica tabla concepto-relación del problema de explotación de información (tabla 5.13).

- Subpaso 2.4. Identificación de dependencias entre los elementos conceptuales del problema

El objetivo de este paso es identificar aquellos elementos que presenten dependencias cognitivas con otros elementos los cuales brinden información redundante o afecten la comprensión del conocimiento oculto en los datos. En el caso de estudio propuesto, no se identifican datos dependientes que afecten los resultados del análisis.

- Subpaso 2.5. Representación del problema de explotación de información

Para la conceptualización del problema de explotación de información, se hace uso de la técnica red semántica del problema de explotación, cuyo resultado se presenta en la figura 5.4.

	Termino	Categoría	Descripción
1	Alumno	Concepto	Estudiante de la materia Análisis de Sistemas
2	Asigna las entradas y salidas a cada proceso	Atributo	Posibilidad del alumno de Asignar las entradas y salidas de cada proceso en el DFD
3	Define las demoras	Atributo	Posibilidad del alumno de definir las demoras
4	Grafica los procesos de nivel 1 según TE	Atributo	Posibilidad del alumno de graficar los procesos de nivel 1 según la TE
5	Identifica	Relación	El alumno se identifica con un código único
6	Identifica	Relación	Una regla se identifica a partir de un código único
7	Identifica el alcance del sistema	Atributo	Posibilidad del alumno de identificar el alcance del sistema
8	Identifica entidades externas	Atributo	Posibilidad del alumno de identificar las entidades externas que interactúan con el sistema
9	Identifica entradas y salidas	Atributo	Posibilidad del alumno de identificar las entradas y salidas del sistema
10	Identifica la comunicación entre procesos	Atributo	Posibilidad del alumno de identificar la comunicación entre los procesos
11	Identifica la forma de activación	Atributo	Posibilidad del alumno de identificar la forma en la cual se activa el sistema
12	Identifica las respuestas	Atributo	Posibilidad del alumno de identificar las respuestas que el sistema provee a las entidades externas
13	Identificador	Atributo	Código Único que permite identificar a un alumno
14	Identificador	Atributo	Código Único que permite identificar a una regla
15	Marca la activación de los procesos	Atributo	Posibilidad del alumno de identificar la activación de los procesos
16	Puede identificar funciones?	Atributo	Posibilidad del alumno de identificar las funcionalidades del sistema
17	Regla	Concepto	Descripción del resultado obtenido por el alumno en base a los elementos de evaluación
18	Resultado	Atributo	Calificación que obtuvo el alumno
19	Según	Relación	Las reglas que definen al alumno se definen según los elementos de evaluación
20	Subconjunto	Relación	Identifica el alcance del sistema es un elemento del conjunto a evaluar
21	Subconjunto	Relación	Identifica entidades externas es un elemento del conjunto a evaluar
22	Subconjunto	Relación	Identifica entradas y salidas es un elemento del conjunto a evaluar
23	Subconjunto	Relación	Puede identificar funciones? es un elemento del conjunto a evaluar
24	Subconjunto	Relación	Identifica la forma de activación es un elemento del conjunto a evaluar
25	Subconjunto	Relación	Identifica las respuestas es un elemento del conjunto a evaluar
26	Subconjunto	Relación	Grafica los procesos de nivel 1 según TE es un elemento del conjunto a evaluar
27	Subconjunto	Relación	Asigna las entradas y salidas a cada proceso es un elemento del conjunto a evaluar
28	Subconjunto	Relación	Marca la activación de los procesos es un elemento del conjunto a evaluar
29	Subconjunto	Relación	Identifica la comunicación entre procesos es un elemento del conjunto a evaluar

30	Subconjunto	Relación	Define las demoras es un elemento del conjunto a evaluar
----	-------------	----------	--

Tabla 5.11 - Término-Categoría-Definición del Problema de Explotación de Información aplicada al caso de estudio

Concepto	Atributo	Relación	Entrada/Salida	Valor
Regla	Identificador	identifica	Salida	Numérico
	Identifica el alcance del sistema	subconjunto	Elemento Conjunto Salida	B - R - M
	Identifica entidades externas	subconjunto	Elemento Conjunto Salida	B - R - M
	Identifica entradas y salidas	subconjunto	Elemento Conjunto Salida	B - R - M
	Puede identificar funciones?	subconjunto	Elemento Conjunto Salida	B - R - M
	Identifica la forma de activación	subconjunto	Elemento Conjunto Salida	B - R - M
	Identifica las respuestas	subconjunto	Elemento Conjunto Salida	B - R - M
	Grafica los procesos de nivel 1 según TE	subconjunto	Elemento Conjunto Salida	B - R - M
	Asigna las entradas y salidas a cada proceso	subconjunto	Elemento Conjunto Salida	B - R - M
	Marca la activación de los procesos	subconjunto	Elemento Conjunto Salida	B - R - M
	Identifica la comunicación entre procesos	subconjunto	Elemento Conjunto Salida	B - R - M
	Define las demoras	subconjunto	Elemento Conjunto Salida	B - R - M
Alumno	Id_alumno	identifica	No relevante	Numérico
	Resultado	según	Entrada	D - A

Tabla 5.12. - Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información aplicada al caso de estudio

Concepto	Concepto asociado	Relación	Descripción
Regla	Alumno	define	La regla define el resultado que obtuvo el alumno

Tabla 5.13. - Concepto-Relación del Problema de Explotación de Información aplicada al caso de estudio

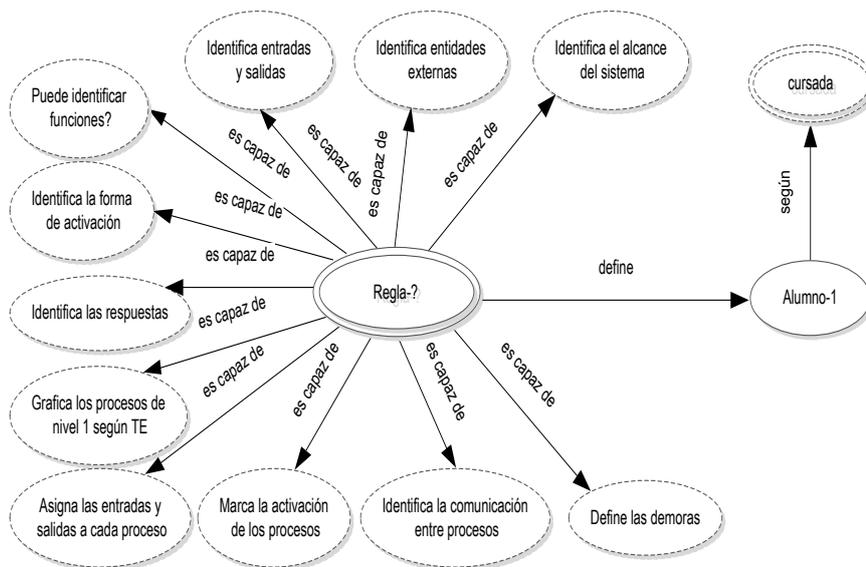


Figura 5.4 - Red Semántica del Problema de Explotación de Información aplicada al caso de estudio

Paso 3: Identificación del proceso de explotación de información

A partir de las conceptualizaciones del dominio del negocio y del problema de explotación previamente presentadas, se establece el proceso de explotación de información a implementar, a través de aplicar el algoritmo de derivación del proceso de explotación de información propuesto en [Martins, et. Al., 2014]. El proceso de explotación de información obtenido es el Proceso de Descubrimiento de Reglas de Comportamiento.

Paso 4: Preparación de los datos

La preparación de los datos para la ejecución del proceso de explotación de información previamente detectado, consta de 3 subpasos:

- Subpaso 4.1: Selección de los campos asociados al problema de explotación de información

Se presenta el listado de datos a utilizar para la ejecución del proceso de explotación de información previamente definido.

Para el caso de estudio propuesto, solo se excluirá el identificador del alumno, dada su irrelevancia en el problema a analizar.

- Subpaso 4.2: Conversión de los datos

A partir de los algoritmos de minería de datos seleccionados, en base al proceso de explotación de información determinado, puede ser necesario convertir el tipo de dato de alguno de los campos seleccionados.

Para el caso de estudio propuesto, no fue necesario convertir ningún dato.

- Subpaso 4.3: Generación de la base de datos

Una vez identificados y convertidos los datos, es necesario realizar un proceso que integre todos los campos seleccionados en una única y misma fuente de datos. El resultado de aplicar el subpaso actual es la

generación de la base de datos sobre la cual se implementará el proceso de explotación de información.

Paso 5: Implementación

El paso implementación está conformado por 2 subpasos:

- Subpaso 5.1: Implementación

Este subpaso tiene como objetivo implementar el proceso de explotación de información identificado (mediante la ejecución de los algoritmos de minería de datos), en la base de datos generada en el paso anterior.

- Subpaso 5.2: Interpretación de los resultados

A partir de los resultados obtenidos en el subpaso anterior, se descartan aquellos patrones irrelevantes, informando aquellas piezas de conocimientos útiles para dar soporte a la toma de decisiones.

En el caso de estudio propuesto, se identificaron dos reglas que definen al comportamiento de los alumnos:

[a] Los alumnos que aprueban, se caracterizan por realizar bien la asignación de las entradas y salidas de cada proceso e identificar las respuestas de cada proceso de manera bien (regla soportada por el 62.5% de los alumnos aprobados);

[b] Los alumnos que aprueban, se caracterizan por realizar bien la asignación de las entradas y salidas de cada proceso e identificar las respuestas de cada proceso de manera regular (regla soportada por el 30% de los alumnos aprobados); y

[c] Los alumnos que desaprueban, se caracterizan por realizar regular o mal la asignación de las entradas y salidas de cada proceso (regla soportada por el 100% de los alumnos desaprobados). Se identifica que la comprensión de la asignación de las entradas y salidas junto

con la identificación de las respuestas de cada proceso son claves para la aprobación del tema, siendo apoyado por más de un 90% de los casos. Asimismo, se resalta que no realizar de forma correcta (Bien) la asignación de las entradas y salidas de cada proceso, garantiza la desaprobación del alumno.

5.2.3. Conclusión del Caso de Validación

El proceso propuesto, permite identificar los conceptos fundamentales para la aprobación de la cursada, permitiendo generar acciones para fortalecer a los alumnos en dicho tema, garantizando un mayor grado de aprobados.

5.3. Identificación de Conceptos Débilmente Apropriados por Estudiantes

En esta sección se describe el caso de estudio (sección 5.3.1), luego se aplica el proceso (sección 5.3.2), y por último se describe la conclusión (sección 5.3.3).

5.3.1. Descripción del Caso

La carrera Ingeniería en Sistemas de Información de la de la Universidad Tecnológica Nacional Facultad Regional Buenos Aires (UTN - FRBA) tiene un programa académico el cual divide la cursada de materias a lo largo de cinco años. Sistemas y Organizaciones es una de las materias de primer año y troncales de la carrera [DISI, 2014]. Dicha materia es el primer acercamiento del alumno con los conceptos de sistemas de información y circuitos administrativos siendo de gran importancia la comprensión de estos temas por parte del alumno y la aprobación de la materia. Por tal motivo se desea poder entender el desempeño de los alumnos en esta materia en base a los exámenes finales.

Para el desarrollo del actual trabajo se han utilizado 449 exámenes finales correspondientes al período comprendido entre octubre 2012 y febrero 2013. Los temas evaluados en dichos exámenes se corresponden con los siguientes temas:

- Metodología de Sistemas de Información (Teoría de Metodología): el alumno debe poder describir el objetivo, las principales actividades, las técnicas y herramientas de cada una de las etapas de la Metodología de Sistemas de Información.
- Teoría General (V/F Teoría): abarca preguntas teóricas de todos los conceptos de la asignatura.
- Circuitos Administrativos (Teoría de Circuitos): comprende teoría sobre los circuitos básicos de una organización genérica junto con las normas de control interno.
- Cursograma (Práctica de Cursograma): consiste en señalar los errores de un ejercicio resuelto mediante esta técnica gráfica.

Las condiciones de aprobación del examen son tener al menos la mitad de cada ejercicio respondido correctamente. Las preguntas no respondidas al igual que las respondidas erróneamente, no suman ni restan en la nota final. Además se debe considerar que los finales pueden ser de dos tipos de acuerdo a los temas evaluados: Cursograma (347 finales) o Circuitos Administrativo (102 finales). En la siguiente tabla se pueden los tipos de ejercicios y las cantidades según tipo de final [Raus *et al.*, 2013].

Las variables relacionadas con el caso de estudio propuesto, se presentan en la tabla 5.14.

5.3.2. Aplicación del Proceso

Paso 1: Conceptualización del dominio

La conceptualización del dominio del negocio se compone de 3 subpasos:

- Subpaso 1.1. Identificación los elementos relevantes del dominio

Para la identificación de los elementos relevantes del dominio de negocio, se hace uso de las técnicas propuestas en [Martins, 2014]. Los formalismos resultantes de ejecutar dichas técnicas en el caso de estudio propuesto son la tabla término-categoría-definición del dominio (tabla 5.15) y la tabla concepto-atributo-relación-valor del dominio (tabla 5.16).

Campo	Descripción	Valores
MET_1	1º Pregunta Teórica de Metodología	0, 0.5 y 1
MET_2	2º Pregunta Teórica de Metodología	0, 0.5 y 1
ToF_1	1º Afirmación Verdadero o Falso	0, 0.5 y 1
ToF_2	2º Afirmación Verdadero o Falso	0, 0.5 y 1
ToF_3	3º Afirmación Verdadero o Falso	0, 0.5 y 1
ToF_4	4º Afirmación Verdadero o Falso	0, 0.5 y 1
Tipo	Es el tipo de final tomado	Circuito - Cursograma
CIR_1	1º Ejercicio para completar Teoría de Circuitos	-1: En caso de que no aplique 0, 0.5 o 1: En caso de que aplique
CIR_2	2º Ejercicio para completar Teoría de Circuitos	-1: En caso de que no aplique 0, 0.5 o 1: En caso de que aplique
CIR_3	3º Ejercicio para completar Teoría de Circuitos	-1: En caso de que no aplique 0, 0.5 o 1: En caso de que aplique
CIR_4	4º Ejercicio para completar Teoría de Circuitos	-1: En caso de que no aplique 0, 0.5 o 1: En caso de que aplique
CUR_1	1º Ejercicio cursograma	-1: En caso de que no aplique 0, 0.5 o 1: En caso de que aplique
CUR_2	2º Ejercicio cursograma	-1: En caso de que no aplique 0, 0.5 o 1: En caso de que aplique
CUR_3	3º Ejercicio cursograma	-1: En caso de que no aplique 0, 0.5 o 1: En caso de que aplique
CUR_4	4º Ejercicio cursograma	-1: En caso de que no aplique 0, 0.5 o 1: En caso de que aplique
CUR_5	5º Ejercicio cursograma	-1: En caso de que no aplique 0, 0.5 o 1: En caso de que aplique
Resultado	Indica si el alumno aprobó o no el final	A: Aprobado D: Desaprobado

Tabla 5.14-Término-Categoría-Definición del Problema de Explotación de Información aplicada al caso de estudio

- Subpaso 1.2. Identificación de las relaciones entre elementos del dominio

Para la identificación de las relaciones entre elementos del dominio de negocio, se hace uso la técnica propuesta en [Martins, 2014], la tabla concepto-relación del dominio. Como resultado de aplicar la técnica tabla concepto-relación del dominio al caso de estudio propuesto se obtiene la tabla 5.17.

Término	Categoría	Definición
Alumno	Concepto	Persona que realiza la cursada de SyO
CIR_1	Atributo	1° Ejercicio para completar Teoría de Circuitos
CIR_2	Atributo	2° Ejercicio para completar Teoría de Circuitos
CIR_3	Atributo	3° Ejercicio para completar Teoría de Circuitos
CIR_4	Atributo	4° Ejercicio para completar Teoría de Circuitos
contiene	Relación	El examen contiene una primer pregunta teórica de metodología
contiene	Relación	El examen contiene una segunda pregunta teórica de metodología
contiene	Relación	El examen contiene una primer pregunta de afirmación
contiene	Relación	El examen contiene una segunda pregunta de afirmación
contiene	Relación	El examen contiene una tercera pregunta de afirmación
contiene	Relación	El examen contiene una cuarta pregunta de afirmación
CUR_1	Atributo	1° Ejercicio cursograma
CUR_2	Atributo	2° Ejercicio cursograma
CUR_3	Atributo	3° Ejercicio cursograma
CUR_4	Atributo	4° Ejercicio cursograma
CUR_5	Atributo	5° Ejercicio cursograma
es	Relación	El examen es de un tipo posible
Examen	Concepto	Evaluación final de la cursada
id_alumno	Atributo	Identificador único del alumno
id_examen	Atributo	Identificador único del examen
identifica	Relación	el id_alumno identifica a un alumno
identifica	Relación	el id_examen identifica a un examen
MET_1	Atributo	1° Pregunta Teórica de Metodología
MET_2	Atributo	2° Pregunta Teórica de Metodología
Nota	Atributo	Indica si el alumno aprobó o no el final
obtiene	Relación	El alumno obtiene una calificación por el examen realizado
puede contener	Relación	El examen puede contener un primer ejercicio de teoría de circuitos
puede contener	Relación	El examen puede contener un segundo ejercicio de teoría de circuitos
puede contener	Relación	El examen puede contener un tercero ejercicio de teoría de circuitos
puede contener	Relación	El examen puede contener un cuarto ejercicio de teoría de circuitos
puede contener	Relación	El examen puede contener un primer ejercicio de cursograma
puede contener	Relación	El examen puede contener un segundo ejercicio de cursograma
puede contener	Relación	El examen puede contener un tercero ejercicio de cursograma
puede contener	Relación	El examen puede contener un cuarto ejercicio de cursograma
puede contener	Relación	El examen puede contener un quinto ejercicio de cursograma
Rinde	Relación	El alumno rinde el examen final
Tipo	Atributo	Es el tipo de final tomado
ToF_1	Atributo	1° Afirmación Verdadero o Falso
ToF_2	Atributo	2° Afirmación Verdadero o Falso
ToF_3	Atributo	3° Afirmación Verdadero o Falso
ToF_4	Atributo	4° Afirmación Verdadero o Falso

Tabla 5.15 - Término-Categoría-Definición del dominio aplicada al caso de estudio

- Subpaso 1.3. Representación conceptual del dominio

Para la conceptualización del dominio del negocio, se hace uso de la técnica red semántica del modelo de negocio propuesto en [Martins, 2014]. El resultado de aplicar dicha técnica al caso de estudio propuesto, se presenta en la figura 5.5.

Paso 2: Conceptualización del problema de explotación de información

La conceptualización del problema de explotación de información está compuesta por 5 subpasos:

- Subpaso 2.1. Identificación del problema de negocio y su traducción al problema de explotación de información

A partir del problema de negocio definido en la descripción del caso de estudio (sección 5.1), se obtiene como problema de explotación de información: “comprender cuales son los elementos (preguntas) relevantes para la aprobación o no del examen”.

Concepto	Atributo	Relación	Valor
Alumno	Id_Alumno	identifica	Numérico
	Nota	obtiene	A - D
Examen	Id_examen	identifica	Numérico
	MET_1	contiene	0, 0.5 y 1
	MET_2	contiene	0, 0.5 y 1
	ToF_1	contiene	0, 0.5 y 1
	ToF_2	contiene	0, 0.5 y 1
	ToF_3	contiene	0, 0.5 y 1
	ToF_4	contiene	0, 0.5 y 1
	Tipo	es	Circuito - Cursograma
	CIR_1	puede contener	0, 0.5,1 y -1
	CIR_2	puede contener	0, 0.5,1 y -1
	CIR_3	puede contener	0, 0.5,1 y -1
	CIR_4	puede contener	0, 0.5,1 y -1
	CUR_1	puede contener	0, 0.5,1 y -1
	CUR_2	puede contener	0, 0.5,1 y -1
	CUR_3	puede contener	0, 0.5,1 y -1
CUR_4	puede contener	0, 0.5,1 y -1	
CUR_5	puede contener	0, 0.5,1 y -1	

Tabla 5.16 - Concepto-Atributo-Relación-Valor del dominio aplicada al caso de estudio

Concepto	Concepto asociado	relación	Descripción
Alumno	Examen	rinde	el alumno rinde un examen

Tabla 5.17 - Concepto-Relación del dominio aplicada al caso de estudio

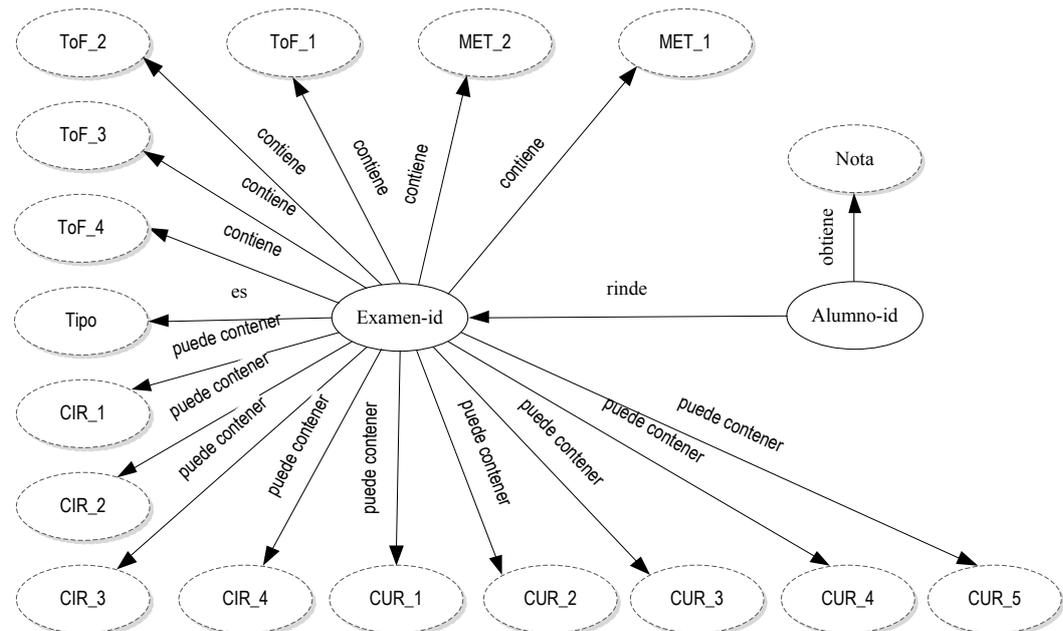


Figura 5.5-Red Semántica del Modelo de Negocio aplicada al caso de estudio

- Subpaso 2.2. Identificación de los conceptos pertenecientes al problema de explotación de información

A partir del problema de explotación de información se procede a seleccionar y determinar cuáles elementos están relacionados con el mismo. Para ello se implementa las técnicas tabla Término-Categoría-Definición del Problema de Explotación de Información y Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información, propuestas en [Martins, 2014]. El resultado de aplicar cada una de ellas al caso de estudio propuesto, se presenta en las tablas 5.18 y 5.19 respectivamente.

- Subpaso 2.3. Identificación de las relaciones entre elementos del problema de explotación de información

Para la identificación de las relaciones entre elementos del problema de explotación de información, se implementa la técnica propuesta en [Martins, 2014]: la tabla concepto-relación del problema de explotación de información. El resultado de aplicar dicha tabla al caso de estudio propuesto se presenta en la tabla 5.20.

- Subpaso 2.4. Identificación de dependencias entre los elementos conceptuales del problema

El objetivo de este paso es identificar aquellos elementos que presenten dependencias cognitivas con otros elementos los cuales brinden información redundante o afecten la comprensión del conocimiento oculto en los datos. Para ello se propone la tabla de dependencias cuyo objetivo es identificar las dependencias existentes entre los datos, y eliminar aquellas que se consideren perjudiciales para el desarrollo del proceso. La tabla de dependencias está compuesta por tres columnas: en la primera se registra el elemento que no posea dependencia, en la segunda el elemento dependiente del elemento anteriormente registrado, y en la tercera columna se registra la acción a realizar. En la tabla 5.21 se presentan los elementos identificados en el caso de estudio propuesto.

- Subpaso 2.5. Representación del problema de explotación de información

Para la conceptualización del problema de explotación de información, se hace uso de la técnica red semántica del problema de explotación propuesta en [Martins, 2014]. El resultado de aplicar dicha técnica al caso de estudio propuesto, se presenta en la figura 5.6.

Paso 3: Identificación del proceso de explotación de información

A partir de las conceptualizaciones del dominio del negocio y del problema de explotación previamente presentadas, se establece el proceso de explotación de información a implementar, a través de aplicar

el algoritmo de derivación del proceso de explotación de información propuesto en [Martins, 2014]. El proceso de explotación de información obtenido es: el Proceso de Ponderación de Reglas de Pertenencia a Grupos (Grupos definidos).

Paso 4: Preparación de los datos

La preparación de los datos para la ejecución del proceso de explotación de información previamente detectado, consta de 3 subpasos:

- Subpaso 4.1: Selección de los campos asociados al problema de explotación de información

A partir de conceptualizar los distintos elementos involucrados se seleccionan aquellos campos asociados al problema de explotación de información previamente establecido, obteniéndose el listado de datos que conformarán la base de datos a implementar. En el caso de estudio propuesto, los campos a utilizar son todos los datos identificados en la tabla 5.14.

Termino	Categoría	Descripción
Examen	Concepto	Evaluación final de la cursada SyO
CIR_1	Atributo	1° Ejercicio para completar Teoría de Circuitos
CIR_2	Atributo	2° Ejercicio para completar Teoría de Circuitos
CIR_3	Atributo	3° Ejercicio para completar Teoría de Circuitos
CIR_4	Atributo	4° Ejercicio para completar Teoría de Circuitos
CUR_1	Atributo	1° Ejercicio cursograma
CUR_2	Atributo	2° Ejercicio cursograma
CUR_3	Atributo	3° Ejercicio cursograma
CUR_4	Atributo	4° Ejercicio cursograma
CUR_5	Atributo	5° Ejercicio cursograma
elementos relevantes	Concepto	Conjunto de elementos del examen que definen la calificación
Es	Atributo	El examen es de un tipo
Id_Examen	Atributo	Identificador único del examen
Identifica	Relación	El id_examen identifica al examen
MET_1	Atributo	1° Pregunta Teórica de Metodología
MET_2	Atributo	2° Pregunta Teórica de Metodología
Nota	Atributo	Indica la calificación del examen
Pertenecen	Relación	Los elementos relevantes pertenecen al examen final
Según	Relación	La nota del examen se caracteriza según los resultados en elementos identificados
Subconjunto	Relación	Los elementos (preguntas) del examen son el subconjunto de elementos relevantes
Tipo	Atributo	Es el tipo de final tomado

ToF_1	Atributo	1° Afirmación Verdadero o Falso
ToF_2	Atributo	2° Afirmación Verdadero o Falso
ToF_3	Atributo	3° Afirmación Verdadero o Falso
ToF_4	Atributo	4° Afirmación Verdadero o Falso

Tabla 5.18-Término-Categoría-Definición del Problema de Explotación de Información aplicada al caso de estudio

concepto	Atributo	Relación	Entrada/Salida	Valor
características relevantes	MET_1	subconjunto	Elemento Conjunto Salida	0, 0.5 y 1
	MET_2	subconjunto	Elemento Conjunto Salida	0, 0.5 y 1
	ToF_1	subconjunto	Elemento Conjunto Salida	0, 0.5 y 1
	ToF_2	subconjunto	Elemento Conjunto Salida	0, 0.5 y 1
	ToF_3	subconjunto	Elemento Conjunto Salida	0, 0.5 y 1
	ToF_4	subconjunto	Elemento Conjunto Salida	0, 0.5 y 1
	Tipo	subconjunto	Elemento Conjunto Salida	Circuito - Cursograma
	CIR_1	subconjunto	Elemento Conjunto Salida	0, 0.5,1 y -1
	CIR_2	subconjunto	Elemento Conjunto Salida	0, 0.5,1 y -1
	CIR_3	subconjunto	Elemento Conjunto Salida	0, 0.5,1 y -1
	CIR_4	subconjunto	Elemento Conjunto Salida	0, 0.5,1 y -1
	CUR_1	subconjunto	Elemento Conjunto Salida	0, 0.5,1 y -1
	CUR_2	subconjunto	Elemento Conjunto Salida	0, 0.5,1 y -1
	CUR_3	subconjunto	Elemento Conjunto Salida	0, 0.5,1 y -1
	CUR_4	subconjunto	Elemento Conjunto Salida	0, 0.5,1 y -1
	CUR_5	subconjunto	Elemento Conjunto Salida	0, 0.5,1 y -1
Examen	Id_Examen	identifica	No relevante	Numérico
	Nota	según	Entrada	A - D

Tabla 5.19-Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información aplicada al caso de estudio

Concepto	Concepto asociado	Relación	Descripción
elementos relevantes	Examen	pertenecen	Los elementos relevantes pertenecen al examen final

Tabla 5.20-Concepto-Relación del Problema de Explotación de Información aplicada al caso de estudio

Elemento	Elemento dependiente	Acción
Tipo	CIR_1	Incorporar al elemento Tipo como variable del problema de explotación de información
Tipo	CIR_2	
Tipo	CIR_3	
Tipo	CIR_4	
Tipo	CUR_1	
Tipo	CUR_2	
Tipo	CUR_3	
Tipo	CUR_4	
Tipo	CUR_5	

Tabla 5.21-Dependencias aplicadas al caso de estudio.

- Subpaso 4.2: Conversión de los datos

A partir de los algoritmos de minería de datos seleccionados, en base al proceso de explotación de información determinado, puede ser necesario convertir el tipo de dato de alguno de los campos seleccionados. Para el caso de estudio propuesto, no fue necesario convertir ningún dato.

- Subpaso 4.3: Generación de la base de datos

Una vez identificados y convertidos los datos, es necesario realizar un proceso que integre todos los campos seleccionados en una única y misma fuente de datos. El resultado de aplicar el subpaso actual es la generación de la base de datos en la cual se implementará el proceso de explotación de información.

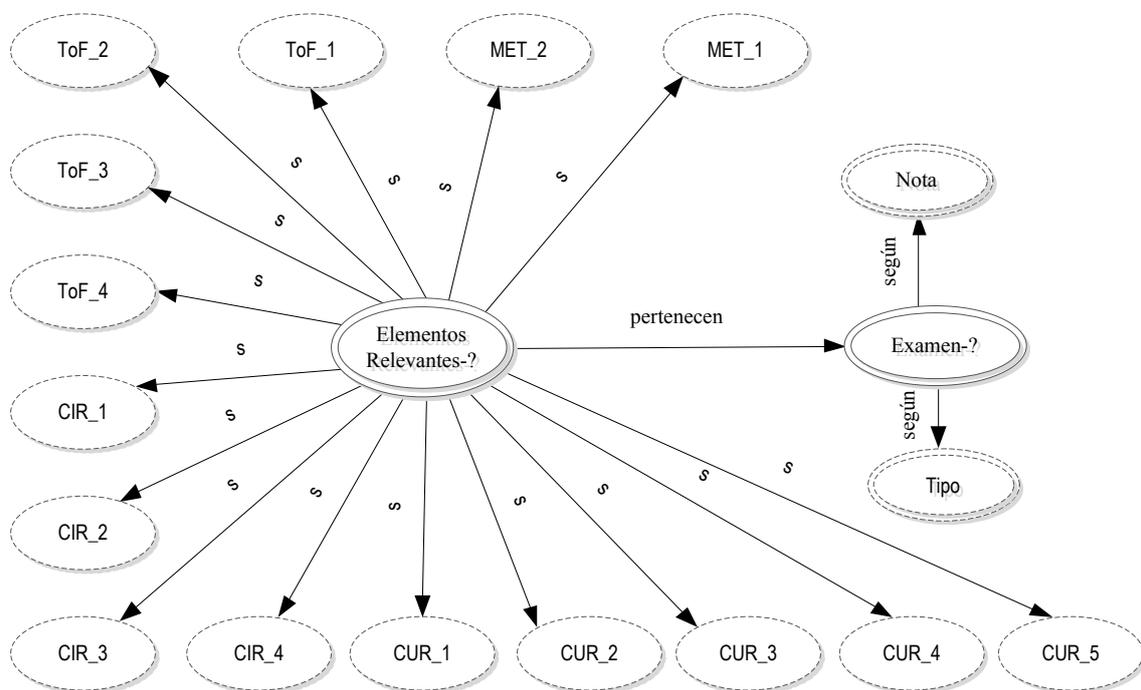


Figura 5.6. Red Semántica del Problema de Explotación de Información aplicada al caso de estudio

Paso 5: Implementación

El paso implementación está conformado por 2 subpasos:

- Subpaso 5.1: Implementación

Este subpaso tiene como objetivo implementar el proceso de explotación de información identificado (mediante la ejecución de los algoritmos de minería de datos), en la base de datos generada en el paso anterior.

- Subpaso 5.2: Interpretación de los resultados

A partir de los resultados obtenidos en el subpaso anterior, se realiza una interpretación de los mismos cuyo objetivo es identificar información relevante, descartando aquella que no sea fiable, y presentar la información obtenida. La información obtenida del caso de estudio presentado es:

Se identifican distintos elementos por tener en cuenta a partir de las dependencias conceptuales identificadas en el paso 2.4, es decir, que según el tipo de examen se identificó la relevancia de distintas preguntas.

Para el examen de cursograma, los elementos relevantes para la aprobación, o no, del mismo son las Preguntas Teóricas de Metodología, con una ponderación aproximadamente del 95%. Mientras que para el examen de Circuito, los ejercicios para completar de la Teoría de Circuitos presentan una ponderación del 81%.

A partir de los resultados obtenidos, puede afirmarse que los conceptos a reforzar para mejorar la calidad de los resultados de los estudiantes son:

- a) La teoría sobre la Metodología de Sistemas de Información, y
- b) La teoría sobre Circuitos Administrativos.

5.3.3. Conclusión del Caso de Validación

La Explotación de Información Educativa tiene un desarrollo de más de una década y emerge como una contribución a la sistematización del entendimiento de: comportamientos estudiantiles, fallas en los procesos de enseñanza,

contenidos mal apropiados, entre otros patrones de conocimiento identificables mediante los procesos y las técnicas que propone la Ingeniería de Explotación de Información. Si bien las primeras aplicaciones se centraron en la información provista por los espacios virtuales de enseñanza aprendizaje y los sistemas de gestión de calificaciones y perfiles de estudiantes; hace más de un lustro que se vienen desarrollando experiencias sobre aplicaciones a la evaluación de procesos de enseñanza-aprendizaje presenciales [Britos *et al.*, 2008a; 2008b; Saavedra-Martinez *et al.*, 2012].

En este trabajo, se ha propuesto un proceso que permite identificar contenidos débilmente apropiados por los estudiantes. En el caso de estudio estos son:

- [a] la teoría sobre la Metodología de Sistemas de Información, y
- [b] la teoría sobre Circuitos Administrativos.

Esta identificación, permite generar acciones para fortalecer los aprendizajes en dichos temas.

6. CONCLUSIONES

En este capítulo se exponen los aportes realizados por la tesis (sección 6.1), y se presentan las futuras líneas de investigación (sección 6.2).

6.1. Aportaciones de la Tesis

La Explotación de Información Educativa tiene un desarrollo de más de una década y emerge como una contribución a la sistematización del entendimiento de: comportamientos estudiantiles, fallas en los procesos de enseñanza, contenidos mal apropiados, entre otros patrones de conocimiento identificables mediante los procesos y las técnicas que propone la Ingeniería de Explotación de Información. Si bien los primeros resultados se centraron en la información provista por los espacios virtuales de enseñanza aprendizaje y los sistemas de gestión de calificaciones y perfiles de estudiantes; hace más de un lustro que se vienen desarrollado experiencias sobre aplicaciones a la evaluación de procesos de enseñanza-aprendizaje presenciales [Britos *et al.*, 2008a; 2008b; Saavedra-Martinez *et al.*, 2012].

En este trabajo se ha presentado un proceso basado en explotación de información que permite estudiar el comportamiento de los estudiantes durante el desarrollo de un curso o una carrera, logrando identificar y resolver distintos problemas asociados a la apropiación de los conceptos y al favorable desarrollo de la materia, brindando una herramienta que favorezca la comprensión del curso y la posibilidad de implementar medidas que permitan detectar alumnos con dificultades, e intervenir de manera temprana pudiendo corregir factores que impacten en el resultado de la cursada.

Este proceso se ha validado con tres casos de estudio: [a] procesos de identificación de comportamiento de estudiantes recursantes utilizando EVEAs, [b] identificación de conceptos críticos de una cursada, [c] identificación de contenidos débilmente apropiados por estudiantes.

En [a] se ilustra el uso del proceso propuesto con una prueba de concepto en el dominio de alumnos que recursan en la modalidad virtual una materia del área de

ciencias básicas cuya primera cursada fue presencial. Los resultados obtenidos son intuitivamente válidos y no presentan indicios que contraríen las creencias del equipo docente sobre las condiciones de aprobación del curso en esta modalidad. Sin embargo, con relación a las condiciones de desaprobación y considerando que la población estudiada era pequeña, el hecho que los atributos de mayor incidencia para la desaprobación presenten porcentajes sensiblemente por debajo del 100% se considera un indicador que permite postular que la replicación del proceso sobre una población por lo menos sea mayor en un orden de magnitud, permitiendo emerger atributos específicos sobre los distintos grupos de causales de desaprobación. En cambio en [b] y [c] se identificaron reglas que definen fuertemente el comportamiento de los estudiantes.

De lo anterior se desprende que el proceso propuesto constituye una nueva herramienta de diagnóstico en el área; y si bien en [a] los resultados obtenidos son consistentes pero sólo sirven para ilustrar la aplicación de la solución propuesta en [b] y [c] se refleja el comportamiento del alumno permitiendo identificar los conceptos fundamentales para la aprobación de la cursada, posibilitando la generación de acciones para fortalecer a los alumnos en dicho tema, garantizando un mayor grado de aprobados.

6.2 Futuras Líneas de Investigación

En este trabajo se propuso obtener procesos de comportamiento de comunidades educativas basadas en Entornos Virtuales de Enseñanza y Aprendizaje, dada la característica de los mismos que poseen toda la información de interacción alumno – profesor almacenada en un único repositorio de datos. De esta forma, se obtiene la información para el análisis del comportamiento al alcance, siendo de fácil acceso y procesamiento.

Los resultados obtenidos del proceso de adquisición de patrones en búsqueda de comportamiento a partir de la información disponible en EVEAs, el reducido tamaño

de las muestras utilizadas para la búsqueda de patrones si bien arrojó resultados promisorios, no son concluyentes y se planea continuar en esta línea de trabajo con poblaciones mayores.

De las otras validaciones ha surgido que los docentes usan explotación de información desde distintas visiones. Unos para revisar sus prácticas docentes, es decir, qué temas el docente cree que está enseñando bien, pero no es necesariamente así. En tanto, otros la utilizan para revisar e identificar aquellos conceptos críticos, es decir, qué conceptos tienen una complejidad tal que más allá de ser bien transmitidos por los docentes, no son correctamente apropiados por los alumnos. En este contexto, surgen como futuras líneas de trabajo: proponer extensiones al proceso propuesto que contribuyan a la gestión de la información disponible necesaria para evaluar cada una de estas dos visiones.

REFERENCIAS

- Alsina, C. (1997). *Tutoring versus academic counselling in open and distance learning. A case-study on the redefinition of tutoring in a virtual campus*. Paper submitted to the ICDE Conference, Pennsylvania State University.
- Aretio, L. G. (1987). *Hacia una definición de educación a distancia*. *Boletín informativo de la Asociación Iberoamericana de Educación Superior a Distancia*. Abril. Año 4, N° 18, 4pp.
- Aretio, L. G. (1999). *Fundamentos y componentes de la educación a distancia*.
- Arruabarrena, R., Pérez, T. A., López-Cuadrado, J., Vadillo, J. G. J. (2002). *On evaluating adaptive systems for education*. In *Adaptive hypermedia* (pp. 363–367).
- Baker, R. S., Yacef, K. (2009). *The state of educational data mining in 2009: A review and future visions*. *JEDM-Journal of Educational Data Mining*, 1(1), 3-17.
- Baker, R. (2010). *Data Mining for Education*. To appear in McGaw, B., Peterson, P., Baker, E. (Eds.) *International Encyclopedia of Education* (3rd edition). Oxford, UK: Elsevier.
- Barnes, T., Desmarais, M., Romero, C., Ventura, S. (2009). *Educational Data Mining 2009: 2nd International Conference on Educational Data Mining, Proceedings*. Cordoba, Spain.
- Beck, J.E., Mostow, J. (2008). *How who should practice: Using learning decomposition to evaluate the efficacy of different types of practice for different types of students*. In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems*, 353-362.
- Bravo, J., Ortigosa, A. (2009). *Detecting Symptoms of Low Performance Using Production Rules*. *International Working Group on Educational Data Mining*.
- Bravo, J., Vialardi, C., Ortigosa A. (2008). *Using Decision Trees for Discovering Problems on Adaptive Courses*. *Proceedings of E-Learn 2008: World Conference on ELearning in Corporate, Government, Healthcare & Higher Education*, 2008, p. 268-277.
- Britos, P. (2008). *Procesos de Explotación de Información Basados en Sistemas Inteligentes*. Tesis de Doctorado en Ciencias Informáticas. Facultad de Informática. Universidad Nacional de La Plata.
- Britos, P., Abasolo, M., García-Martínez, R., Perales, F. (2005). *Identification of MPEG-4 Patterns in Human Faces Using Data Mining Techniques*. *Proceedings 13th International*

- Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2005. Páginas 9-10.
- Britos, P., Felgaer, P., García-Martínez, R. (2008c). *Bayesian Networks Optimization Based on Induction Learning Techniques*. In Artificial Intelligence in Theory and Practice II, ed. M. Bramer, (Boston: Springer), en prensa.
- Britos, P., García-Martínez, R. (2009). *Propuesta de Procesos de Explotación de Información*. Proceedings XV Congreso Argentino de Ciencias de la Computación. Workshop de Base de Datos y Minería de Datos. Págs. 1041-1050. ISBN 978-897-24068-4-1.
- Britos, P., Grosser, H., Rodríguez, D., García-Martínez, R. (2008b). *Detecting Unusual Changes of Users Consumption*. In Artificial Intelligence in Theory and Practice II, ed. M. Bramer, (Boston: Springer), en prensa.
- Britos, P., Jiménez Rey, E., García-Martínez, E. (2008). *Work in Progress: Programming Misunderstandings Discovering Process Based On Intelligent Data Mining Tools*. Proceedings 38th ASEE/IEEE Frontiers in Education Conference. Session F4H: Assessing and Understanding student Learning. ISBN 978-1-4244-1970-8.
- Brusilovsky, P., Peylo, C., (2003). *Adaptive and intelligent web-based educational systems*. International Journal of Artificial Intelligence in Education, 13, 156-169.
- Castro, F., Vellido, A., Nebot, A. Mugica, F. (2007). *Applying Data Mining Techniques to e-Learning Problems*. In: Jain, L.C., Tedman, R. and Tedman, D. (eds.) Evolution of Teaching and Learning Paradigms in Intelligent Environment. Studies in Computational Intelligence, 62, Springer-Verlag, 183-221
- Charum, V, (2007). *Modelo de Evaluación de Plataformas Tecnológicas*. Tesis de Magíster en Telecomunicaciones. Escuela de Posgrado, Instituto Tecnológico de Buenos Aires.
- Chau, M., Shiu, B., Chan, I., Chen, H. (2007). *Redips: Backlink Search and Analysis on the Web for Business Intelligence Analysis*. Journal of the American Society for Information Science and Technology, 58(3): 351-365.
- Chen, M., Han, J., Yu, P. (1996). *Data Mining: An Overview from a Database Perspective*. IEEE Transactions on Knowledge and Data Engineering, 8(6): 866-883.
- Chung, W., Chen, H., Nunamaker, J. (2005). *A Visual Framework for Knowledge Discovery on the Web: An Empirical Study of Business Intelligence Exploration*. Journal of Management Information Systems, 21(4): 57-84.

-
- Cigliuti, P., Pollo Cattaneo, M. F., García Martínez, R. (2012). *Procesos de identificación de comportamiento de comunidades educativas centradas en EVEAs*. In XIV Workshop de Investigadores en Ciencias de la Computación.
- Cocea, M., HersHKovitz, A. and Baker, R.S.J.D. (2009). *The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate?* In Proceedings of the 14th International Conference on Artificial Intelligence in Education, 507-514.
- Cogliati, M., Britos, P., García-Martínez, R. (2006a). *Patterns in Temporal Series of Meteorological Variables Using SOM & TDIDT*. In IFIP International Federation for Information Processing, Volume 217, Artificial Intelligence in Theory and Practice, ed. M. Bramer, (Boston: Springer), Pág. 305-314.
- Cole, J. (2005). *Using Moodle*. O'Reilly.
- Corbertt, A.T. (2001). *Cognitive Computer Tutors: Solving the Two-Sigma Problem*. In Proceedings of the International Conference on User Modeling, 137-147.
- Curtis, B., Kellner, M., Over, J. (1992). *Process Modelling*. Communications of the ACM, 35(9): 75-90.
- Cuzzani, Karina. (2014). *Proceso de Evaluación Constante: "Una herramienta indispensable en el momento de re-definir Proyectos de Educación Semi-Presencial"*, Tesis de Maestría de Educación a Distancia. Universidad de Morón. Facultad de Filosofía, Ciencias de la Educación y Humanidades. Versión preliminar facilitado por la autora.
- Dekker, G. W., Pechenizkiy, M., & Vleeshouwers, J. M. (2009). *Predicting Students Drop Out: A Case Study*. International Working Group on Educational Data Mining.
- Delling, R.M. (1987). *Towards a theory of distance education*. ICDE Bulletin, 13, pp. 21-25.
- Deroche, A., Pytel, P., Pollo Cattaneo, M. F. (2013). *Propuesta de mejora en asignatura de grado mediante Explotación de Información*. In VIII Congreso de Tecnología en Educación y Educación en Tecnología.
- DISI. (2014). *Departamento de Ingeniería en Sistemas de Información*. Universidad Tecnológica Nacional – Facultad Regional Buenos Aires. <http://sistemas.frba.utn.edu.ar> Último acceso octubre 2014.
- Ferreira, J., Takai, O., Pu, C. (2005). *Integration of Business Processes with Autonomous Information Systems: A Case Study in Government Services*. Proceedings Seventh IEEE International Conference on Ecommerce Technology. Pp. 471-474.

-
- Freeman, L. (2006). *The Development of Social Network Analysis*. Empirical Press.
- Frias-Martinez, E., Chen, S. Y., Liu, X. (2006). *Survey of data mining approaches to user modeling for adaptive hypermedia*. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 36(6), 734-749.
- García - Martínez, R., Britos, P., Pesado, P., Bertone, R., Pollo - Cattaneo, F., Rodríguez, D., Pytel, P., Vanrell. J (2011). *Towards an Information Mining Engineering*. En Software Engineering, Methods, Modeling and Teaching. Sello Editorial Universidad de Medellín. ISBN 978 - 958 - 8692 - 32 - 6. Páginas 83 - 99
- García - Martínez, R., Britos, P., Rodríguez, D. (2013). *Information Mining Processes Based on Intelligent Systems*. Lecture Notes on Artificial Intelligence, 7906: 402 - 410. ISBN 978 - 3 - 642 - 38576 - 6
- García Martínez, R., Britos, P. (2004) *Ingeniería de Sistemas Expertos*. Editorial Nueva Librería. ISBN 987 - 1104 - 15 - 4
- García Martínez, R., Servente, M., Pasquini, D. (2003). *Sistemas Inteligentes*. Editorial Nueva Librería. ISBN 987-1104-05-7
- Gaudioso, E., Talavera, L. *Data mining to support tutoring in virtual learning communities: experiences and challenges*. Data mining in e-learning. Eds. Romero, C, Ventura, S., Southampton, UK: Wit Press. (pp. 207-226).
- Golfarelli, M., Rizzi, S., Cella, L. (2004). *Beyond data warehousing: what's next in business intelligence?*. Proceedings 7th ACM international workshop on Data warehousing and OLAP. Pág. 1-6.
- Grigori, D., Casati, F., Castellanos, M., Dayal, u., Sayal, M., Shan, M. (2004). *Business Process Intelligence*. Computers in Industry 53(3): 321-343.
- Grobelnik, M., Mladenic, D., Jermol, M. (2002). *Exploiting text mining in publishing and education*. In Proceedings of the ICML-2002 workshop on data mining lessons learned (pp. 34–39).
- Hanna, M. (2004). *Data mining in the e-learning domain*. In Campus-Wide Information Systems, Volume 21, Number 1, 29-34
- Herin, D. Sala, M. Pompidor, P. (2002). *Evaluating and Revising Courses from Web Resources Educational*. In Int. Conf. on Intelligent Tutoring Systems, Spain (pp. 208–218).
- Holmberg, B. (1989), *Theory and Practice of Distance Education*.

- Ingram, A. (1999). *Using web server logs in evaluating instructional web sites*. Journal of Educational Technology Systems, 28(2), 137–157.
- Jeong, H. and Biswas, G. (2008). *Mining Student Behavior Models in Learning-by-Teaching Environments*. In Proceedings of the 1st International Conference on Educational Data Mining, 127-136.
- Jiménez Rey, E. M., Rodríguez, D., Britos, P. V., García Martínez, R. (2008). *Identificación de problemas de aprendizaje de programación con explotación de información*. In XIV Congreso Argentino de Ciencias de la Computación.
- Johnson, S., Arago, S., Shaik, N., Palma-Rivas, N. (2000). *Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments*. Journal of Interactive Learning Research, 11(1), 29–49.
- Kanungo, S. (2005). *Using Process Theory to Analyze Direct and Indirect Value-Drivers of information Systems*. Proceedings of the 38th Annual Hawaii International Conference on System Sciences. Pp. 231-240.
- Koedinger, K., Cunningham, K., Skogsholm A., Leber, B. (2008). *An open repository and analysis tools for fine-grained, longitudinal learner data*. In 1st International Conference on Educational Data Mining, Montreal, 157-166.
- Kononenko, I., Cestnik, B. (1986). *Lymphography Data Set*. UCI Machine Learning Repository. <http://archive.ics.uci.edu/ml/datasets/Lymphography>. Último acceso 29 de Abril del 2008.
- Koubarakis, M., Plexousakis, D. (2000). *A Formal Model for Business Process Modeling and Design*. Lecture Notes in Computer Science, 1789: 142-156.
- Kuna, H., García Martínez, R., Villatoro, F. (2009). *Identificación de Causales de Abandono de Estudios Universitarios. Uso de Procesos de Explotación de Información*. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología 5: 39-44.
- Kuna, H., García Martínez, R. Villatoro, F. (2010). *Pattern Discovery in University Students Desertion Based on Data Mining*. Advances and Applications in Statistical Sciences Journal, 2(2): 275-286. ISSN 0974-6811.
- Langseth, J., Vivatrat, N. (2003). *Why Proactive Business Intelligence is a Hallmark of the Real-Time Enterprise: Outward Bound*. Intelligent Enterprise 5(18): 34-41.
- Larose, D. T. (2005). *Discovering Knowledge in Data, an introduction to Data Mining*. John Wiley & Sons

-
- Li, J., Zar'ane, O. (2004). *Combining usage, content, and structure data to improve web site recommendation*. In International conference on ecommerce and web technologies (pp. 305–315).
- Lönnqvist, A., Pirttimäki, V. (2006). *The Measurement of Business Intelligence*. Information Systems Management, 23(1): 32-40.
- Luan, J. (2002). *Data mining and its applications in higher education*. New directions for institutional research, 2002(113), 17-36.
- Ma, Y., Liu, B., Wong, C., Yu, P., Lee, S. (2000). *Targeting the right students using data mining*. In KDD '00: Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 457–464).
- Martins, S. (2014). *Derivación del Proceso de Explotación de Información Desde el Modelado del Negocio*. Revista Latinoamericana de Ingeniería de Software, 2(1): 53-76, ISSN 2314-2642
- Michalski, R. (1983). *A Theory and Methodology of Inductive Learning*. Artificial Intelligence, 20: 111-161.
- Michalski, R. Bratko, I. Kubat, M. (1998). *Machine Learning and Data Mining, Methods and Applications* (Editores) John Wiley & Sons.
- Minaei-Bidgoli, B., Kashy, D. A., Kortmeyer, G., Punch, W. F. (2003). *Predicting student performance: an application of data mining methods with an educational web-based system*. In Frontiers in Education, 2003. FIE 2003 33rd Annual (Vol. 1, pp. T2A-13). IEEE.
- Mobasher, B, R Cooley and J Srivastava (1999). *Creating adaptive web sites through usage-based clustering of URLs*. Proceedings Workshop on Knowledge and Data Engineering Exchange, Pág. 19-25.
- Moore, M. G. y Kearsley, G. (1996), *Distance Education. A Systems View*, Belmont, CA, Wadsworth Publishing Company.
- Mostow, J., Beck., J. (2006). Some useful tactics to modify, map and mine data from intelligent tutors. In Journal Natural Language Engineering, 12 ,2, 195-208.

- Mostow, J., Beck, J., Cen, H., Cuneo, A., Gouvea, E., Heiner, C., (2005). An educational data mining tool to browse tutor-student interactions: Time will tell! In Proceedings of the Workshop on Educational Data Mining, Pittsburgh, USA (pp. 15–22).
- MSU (2014). *Computer Science and Engineering – Michigan State University*. <http://www.cse.msu.edu>. Último acceso octubre 2014.
- Musen, M., Neumann, B., Studer, R. (2003). *Guest Editors' Introduction: IFIP Conference on Intelligent Information Processing*. IEEE Intelligent Systems 18(2): 16-17.
- Negash, S. Gray, P. (2008). *Business Intelligence*. In Handbook on Decision Support Systems 2, eds. F. Burstein y C. Holsapple (Heidelberg, Springer), Pp. 175-193.
- Novak, J.D., Cañas, A.J. (2006). *The theory underlying concept maps and how to construct and use them*. Technical Report IHMC CmapTools 2006-01.
- Ord1129. (2014). *Ordenanza de la Universidad Tecnológica Nacional N° 1129 “Cursado Intensivo”*. Secretaría Académica UTN-FRBA http://academica.frba.utn.edu.ar/ordenanzas/1129_cursado_intensivo.pdf. Último acceso octubre 2014.
- Ortigosa, A., Carro, R.M. (2003). *The Continuous Empirical Evaluation Approach: Evaluating Adaptive Web-based Courses*. In Int. Conf. User Modeling, Canada (pp. 163-167).
- Pahl, C., Donnellan, C. (2003). *Data mining technology for the evaluation of web-based teaching and learning systems*. In Proceedings of the congress e-learning, Montreal, Canada.
- Paulsen, M. (2003). Online education and learning
- Pedró, F. (1997), *Didactic design, production and distribution of study materials in open higher education. A case-study on the transition from printed materials to digital media*. Paper submitted to the ICDE Conference, Pennsylvania State University
- Perkins, D. (1995), *La escuela inteligente*. Gedisa.
- PROMEI. (2014). *Programa de Mejoramiento para la Enseñanza de la Ingeniería*. <http://portales.educacion.gov.ar/spu/calidad-universitaria/proyectos-de-mejoramiento/promei-i-y-ii/>. Último acceso octubre 2014.
- Quesada Catillo, Rocío. (2006). *Evaluación del aprendizaje en la educación a distancia “en línea”*. RED Revista de Educación a Distancia
- Quinlan J. (1990). Learning Logic Definitions from Relations. Machine Learning, 5:239-266.

- Raus, N. A., Lujan, F. N., Deroche, A., Vegega, C., Pytel, P., & Pollo-Cattaneo, M. F. (2013) *Aplicación del Proceso de Ponderación de Reglas de Pertenencia a Grupos en Evaluaciones Finales en Carreras de Grado*. Memorias del 1er Congreso Nacional de Ingeniería Informática / Sistemas de Información. ISSN 2347-0372.
- Reinschmidt, J., Allison F. (2000). *Business Intelligence Certification Guide*. IBM International Technical Support Organization. SG24-5747-00.
- Rice, W.H. (2006) *Moodle E-learning Course Development*. A complete guide to successful learning using Moodle. Packt publishing.
- Romero, C., Ventura, S. (2006). *Data mining in e-learning*, Southampton, UK: Wit Press.
- Romero, C., Ventura, S. (2007). *Educational data mining: A survey from 1995 to 2005*. Expert systems with applications, 33(1), 135-146.
- Romero, C., Ventura, S. (2010). *Educational data mining: a review of the state of the art*. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 40(6), 601-618.
- Romero, C., Porras, A.R., Ventura, S., Hervás, C., Zafra, A. (2006). *Using sequential pattern mining for links recommendation in adaptive hipermedia educational systems*. In International Conference Current Developments in Technology-Assisted Education. Sevilla, Spain (pp.1016-1020).
- Romero, C., Ventura, S., Bra, P. D. (2004). *Knowledge discovery with genetic programming for providing feedback to courseware author*. User Modeling and User-Adapted Interaction: The Journal of Personalization Research, 14(5), 425–464.
- Romero, C., Ventura, S., García, E. (2008). *Data mining in course management systems: Moodle case study and tutorial*. Computers & Education, 51(1), 368-384.
- Saavedra Martínez, P., Pollo Cattaneo, M. F., Pytel, P., Rodríguez, D., García Martínez, R. (2012a). *Detección de problemas de aprendizaje basado en explotación de información*.
- Saavedra-Martínez, P., Pollo-Cattaneo, F., Rodríguez, D., Britos, P., & García-Martínez, R. (2012b). *Proceso de identificación de errores de apropiación de conceptos basado en explotación de información*. In VII Congreso de Tecnología en Educación y Educación en Tecnología.
- Schiefer, J., Jeng, J., Kapoor, S. Chowdhary, P. (2004). *Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence*. Proceedings IEEE International Conference on E-Commerce Technology. Pp. 162- 169.

- Shen, R., Yang, F., & Han, P. (2002). *Data analysis center based on elearning platform*. In Proceedings of the 5th international workshop on the internet challenge: Technology and applications (pp. 19–28).
- Sherman, B. Y Judkins, P. (1992), *Glimpses of Heaven, Visions of Hell*. Londres, Hodder & Stoughton.
- SIU-Guarani. (2014). ¿Qué es SIU Guaraní y para qué sirve? <http://www.perio.unlp.edu.ar/node/1014>. Último acceso septiembre 2014.
- Sowa, J. (1992). *Semantic Networks*. En Encyclopedia of Artificial intelligence (Editor S. Shapiro, 2da edición). Wiley & Sons. ISBN 978 – 0471503071.
- Spotts, T y Bowman, M. (1995). *Faculty Use of Instructional Technologies in Higher Education*, en Educational Technology, 35(2): 56-64.
- Srivastava, J., Cooley, R., Deshpande, M., Tan, P. (2000). *Web usage mining: Discovery and applications of usage patterns from web data*. SIGKDD Explorations, 1(2), 12–23.
- Tan, P., Steinbach, M., Kumar, V. (2006). *Introduction to Data Mining*. Pearson-Addison Wesley Publishers, Boston, USA.
- Tiffin, J., & Rajasingham, L. (1997). *En busca de la clase virtual: la educación en la sociedad de la información*. (Vol. 43). Editorial Paidós.
- Thomsen, E. (2003). *BI's Promised Land*. Intelligent Enterprise, 6(4): 21-25.
- Tsantis, L., Castellani, J., (2001). *Enhancing learning environments through solution-based knowledge discovery tools*. Journal of Special Education Technology, 16 (4), 1-35.
- Umaphathy, K. (2007). *Towards Co-Design of Business Processes and Information Systems Using Web Services*. Proceedings 40th Annual Hawaii International Conference on System Sciences. Pág. 72-181.
- Yu, P., Own, C., Lin, L. (2001). *On learning behavior analysis of web based interactive environment*. In Proceedings of ICCEE, Oslo/Bergen, Norway
- Zaïane, O., Luo, J. (2001). *Web usage mining for a better web-based learning environment*. In Proceedings of conference on advanced technology for education, Banff, Alberta (pp. 60–64).
- Zorrilla, M. E., Menasalvas, E., Marin, D., Mora, E., Segovia, J. (2005). *Web usage mining project for improving web-based learning sites*. In Web mining workshop, Cataluna.

