

Proceso de Identificación de Comportamiento de Estudiantes Recursantes Utilizando EVEAs

Pablo Cigliuti^{1,2}, Sebastian Martins², Ramón García-Martínez²

¹ Programa de Maestría en Tecnología Informática Aplicada en la Educación. Universidad Nacional de La Plata (UNLP), Argentina

² Grupo de Investigación en Sistemas de Información. Universidad Nacional de Lanús (UNLa), Argentina
pabcig@gmail.com, martinsebastian@yahoo.com.ar, rgm1960@yahoo.com

Resumen

Los cursos universitarios dados a distancia se apoyan en Entornos Virtuales de Enseñanza y Aprendizaje (EVEAs). Estas plataformas generan información que se puede interpretar como las conductas de uso que los participantes exhiben en relación a las herramientas que dan soporte al proceso de enseñanza-aprendizaje mediado y la comunicación entre los sujetos que participan en él. Se asume que estas conductas se correlacionan con el aprovechamiento que cada estudiante hace de los contenidos del curso. En este artículo se propone un proceso de explotación de información orientado a encontrar patrones sobre el comportamiento de estudiantes recursantes que en oportunidad de volver a cursar la asignatura lo hacen en la versión mediada por EVEA.

Palabras clave: Entornos Virtuales de Enseñanza y Aprendizaje. Proceso de enseñanza aprendizaje mediado. Procesos de explotación de información. Patrones sobre el comportamiento.

Abstract

Distance college courses rely on Virtual Environments for Teaching and Learning (VETLs). These platforms generate information that can be interpreted as behaviors of use that participants show using the tools that support the communication process of teaching-learning mediated by technology. It is assumed that these behaviors are correlated with use that each student makes of course contents. In this paper is proposed a process of information mining aimed at finding behavior patterns of students taking the course by second chance and they do in version mediated by VETL.

Keywords: Virtual environments for teaching learning. Teaching learning process mediated by technology. Information mining processes. Behavior patterns.

1. Introducción

La Explotación de Información es una subdisciplina de los Sistemas de Información [1] 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 [2].

Un Proceso de Explotación de Información, puede ser definido como un conjunto de tareas relacionadas lógicamente [3] 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 [4-5].

Britos [6] 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 cuales 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 cuales 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 cuales 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.

Cada proceso de explotación de información tiene asociado una familia de técnicas de minería de datos [7] de utilidad para descubrir los patrones de conocimiento que están siendo buscados. Varias de estas técnicas vienen del campo del Aprendizaje Automático [8].

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.

La educación a distancia es un sistema de enseñanza mediado por tecnología, donde el profesor y alumno no comparten un espacio físico en común para llevar a cabo este proceso. Como principales características de la enseñanza a distancia podemos mencionar [9]: (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 [10], 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. Con la aparición de la computadora personal, y muchos más adelante de Internet, la tecnología utilizada como soporte para la educación a distancia ha ido evolucionando.

Actualmente los cursos universitarios dados a distancia se apoyan, en su mayoría, en Entornos Virtuales de Enseñanza y Aprendizaje (EVEAs). Estas plataformas contienen datos de los participantes, además de herramientas que dan soporte a la comunicación entre alumnos (foros, mensajes, chats, repositorios de materiales de lectura, entre otros), las cuales generan gran cantidad de información [11].

El EVEA registra todas las instancias de comunicación interpersonales. Tales registros podrían utilizarse como base para descubrir patrones de conocimiento sobre el comportamiento de alumnos en cursos en modalidad a distancia.

En este contexto, este artículo presenta el estado del arte sobre las experiencias internacionales y nacionales en el uso educativo de la explotación de información (sección 2), se describe el problema abordado en el trabajo (sección 3), se introduce el proceso de explotación propuesto (sección 4), se presenta un caso de estudio

para ilustrar el proceso y se formulan conclusiones preliminares (sección 5).

2. Estado del Arte

En esta sección se presenta una revisión de experiencias paradigmáticas en el uso educativo de la explotación de información en los ámbitos internacional (sección 2.1) y nacional (sección 2.2).

2.1. Ámbito Internacional

En esta sección se describen resultados en el ámbito internacional que se consideran antecedentes del problema abordado y la solución propuesta.

Desde una visión de marco teórico, en [12-13] se revisa la historia y las tendencias en el campo de la Explotación de Información Educativa [EIE]. 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, el surgimiento de trabajo utilizando los modelos existentes para hacer descubrimiento de conocimiento en el ámbito educativo. Por otra parte, Romero y Ventura [14] 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 [15] formulan una revisión bibliográfica sobre estudios relevantes realizados en EIE. Describe los diferentes grupos de usuarios, tipos de entornos educativos y los datos que proporcionan. Enumeran las tareas típicas en el ámbito educativo que se han resuelto a través de explotación de información y mencionando las técnicas de minería de datos utilizadas en dichos procesos. En colaboración con García [16] se describen 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 el [17] se busca clasificar 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. Bravo y Ortigosa [18] 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 [19] se presenta un caso de estudio en el cual, utilizando procesos de explotación de información, se busca predecir el abandono de estudiantes de una carrera de Ingeniería luego del primer semestre de sus estudios, o

antes de entrar a la carrera, y trata de identificar los factores de éxito para esta carrera.

2.2. Ámbito Nacional

En esta sección se describen resultados en el ámbito nacional que se consideran antecedentes del problema abordado y la solución propuesta.

En [20-21] 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 de programación con base en un proceso de descubrimiento del conocimiento de tres pasos. En [22] 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 relevados sobre cada individuo. Estos patrones de conocimiento descubiertos son utilizados para determinar el protocolo pedagógico, conforme a los descriptos por la Teoría Uno, que mejor aplica a cada estudiante. En [23-24] se aborda el estudio del abandono de los estudios universitarios en el nivel de pregrado buscando posibles causas de deserción. Se utilizan 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. En [25] se define una metodología basada en procesos de explotación de información: descubrimiento de reglas y descubrimiento de atributos significativos, que permite 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.

3. Descripción del Problema

En [26] se plantea la línea de investigación en la cual se encuadra los resultados parcialmente descriptos en este artículo y 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 el contexto investigativo planteado, el problema abordado en este trabajo es el de la definición de un proceso de explotación de información orientado a encontrar patrones sobre el comportamiento de estudiantes que cursan asignaturas mediadas por EVEA.

4. Proceso Propuesto

Para abordar la solución del problema identificado, se propone un proceso que permite estudiar el comportamiento de los estudiantes a lo largo 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 en el curso e intervenir de manera temprana pudiendo corregir factores que impacten en el resultado de la cursada.

El proceso propuesto hace uso de los procesos de explotación de información definidos por Britos [6] y del proceso de derivación del problema de explotación de información definido por Martins [27-28].

A continuación se presenta el proceso propuesto el cual se compone de 5 pasos:

Paso 1: Conceptualización del Dominio

Subpaso 1.1: Identificación de los elementos relevantes del dominio

Subpaso 1.2: Identificación de las relaciones entre elementos del dominio

Subpaso 1.3: Representación conceptual del dominio.

Paso 2: Conceptualización del Problema de Explotación de Información

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

Subpaso 2.2: Identificación de los conceptos pertenecientes al problema de explotación de información

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

Subpaso 2.4: Identificación de dependencias entre los elementos conceptuales del problema

Subpaso 2.5: 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

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

Subpaso 4.2 Conversión de los datos

Subpaso 4.3 Generación de la base de datos

Paso 5: Implementación

Subpaso 5.1 Implementación

Subpaso 5.2 Interpretación de los resultados

El primero paso se orienta a la comprensión y conceptualización del negocio a analizar (identificando 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, el cual 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, consiste en implementar e interpretar los resultados obtenidos, aportando información relevante para la solución del problema de negocio previamente identificado.

5. Caso de Estudio

En esta sección se presenta el enunciado que resume al caso de estudio (sección 5.1), y los resultados obtenidos de aplicar cada uno de los 5 pasos que conforman al proceso propuesto: conceptualización del dominio (sección 5.2), conceptualización del problema de explotación de información (sección 5.3), identificación del proceso de explotación de información (sección 5.4), preparación de los datos (sección 5.5) e implementación y reporte (sección 5.6).

5.1. Descripción

El caso de estudio se describe a partir de la siguiente pieza de texto:

“...La UTN-FRBA implementó hace cuatro años la posibilidad de cursar a distancia la asignatura Matemática Discreta. La materia corresponde al primer año de la carrera de Ingeniería en Sistemas de Información. La cursada en esta modalidad está orientada a los alumnos que ya cursaron dicha materia en forma presencial y no cumplieron satisfactoriamente con las instancias evaluativas. La materia se encuentra dividida en diferentes 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 el 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 dentro 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. Esto 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 1.

Tabla 1. Listado de campos registrados

Campo	Descripción
ld_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 entro 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 creo algún tema de discusión en un foro correspondiente a una unidad
vio_foro_alumno	Indica si el alumno entro 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 creo algún tema de discusión en el foro de alumnos.
vio_foro_novedades	Indica si el alumno entro 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.

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 cuales son de tipo numérico.

El problema de negocio presentado consiste en estudiar a los alumnos con respecto a su participación en el curso con el objetivo de disminuir la cantidad de alumnos que desaprobaban la cursada de la materia.

5.2. 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 [27]. Los formalismos resultantes de ejecutar dichas técnicas en el caso de estudio propuesto son la tabla término-categoría-definición del dominio (tabla 2) y la tabla concepto-atributo-relación-valor del dominio (tabla 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 [27], 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 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 propuesta en [27]. El resultado de aplicar dicha técnica al caso de estudio propuesto, se presenta en la figura 1.

5.3. 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:

“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 [27]. El resultado de aplicar cada una de ellas al caso de estudio propuesto, se presenta en las tablas 5 y 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 [27]: 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 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 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 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 [27]. El resultado de aplicar dicha técnica al caso de estudio propuesto, se presenta en la figura 2.

Tabla 2. Tabla Término-Categoría-Definición del dominio aplicada al caso de estudio

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 3. Tabla Concepto-Atributo-Relación-Valor 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 4. Tabla Concepto-Relación del dominio aplicada al caso de estudio

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

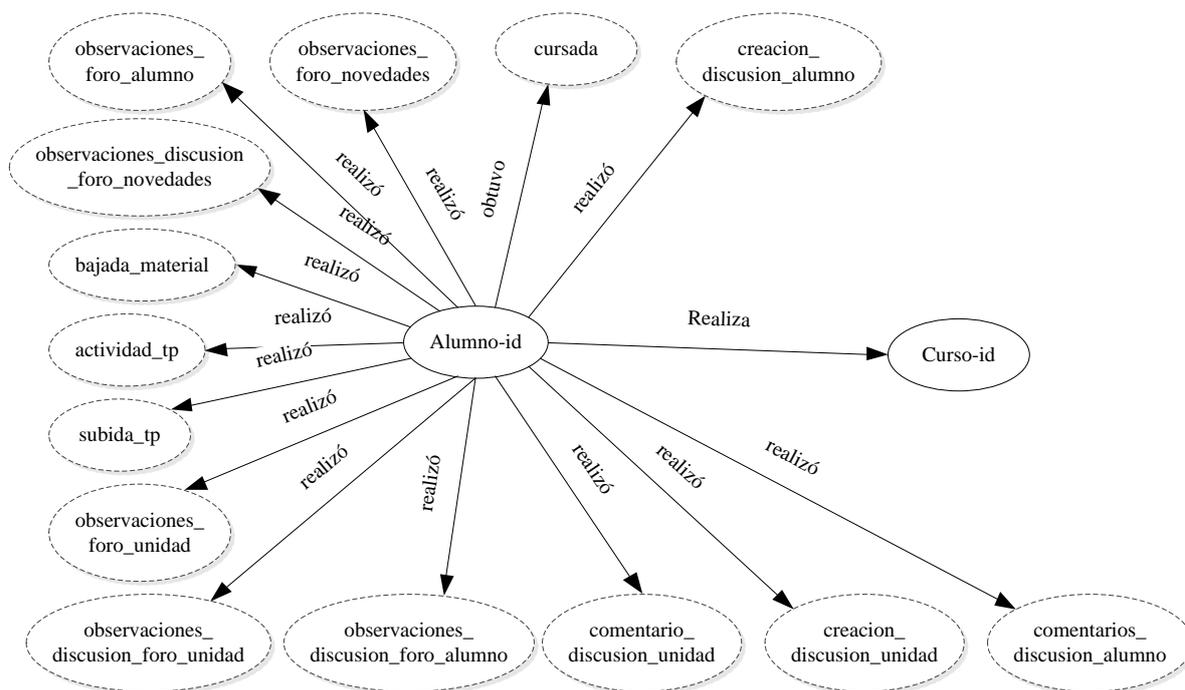


Figura 1. Red Semántica del Modelo de Negocio aplicada al caso de estudio

Tabla 5. Tabla Término-Categoría-Definición del Problema de Explotación de Información aplicada al caso de estudio

Termino	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_caracteristicas_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 6. Tabla Concepto-Atributo-Relación-Valor Extendida del Problema de Explotación de Información aplicada al caso de estudio

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

Tabla 7. Tabla Concepto-Relación 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 8. Tabla de Dependencias aplicada al caso de estudio.

Elemento	Elemento dependiente	Eliminar
actividad_tp	subida_tp	NO

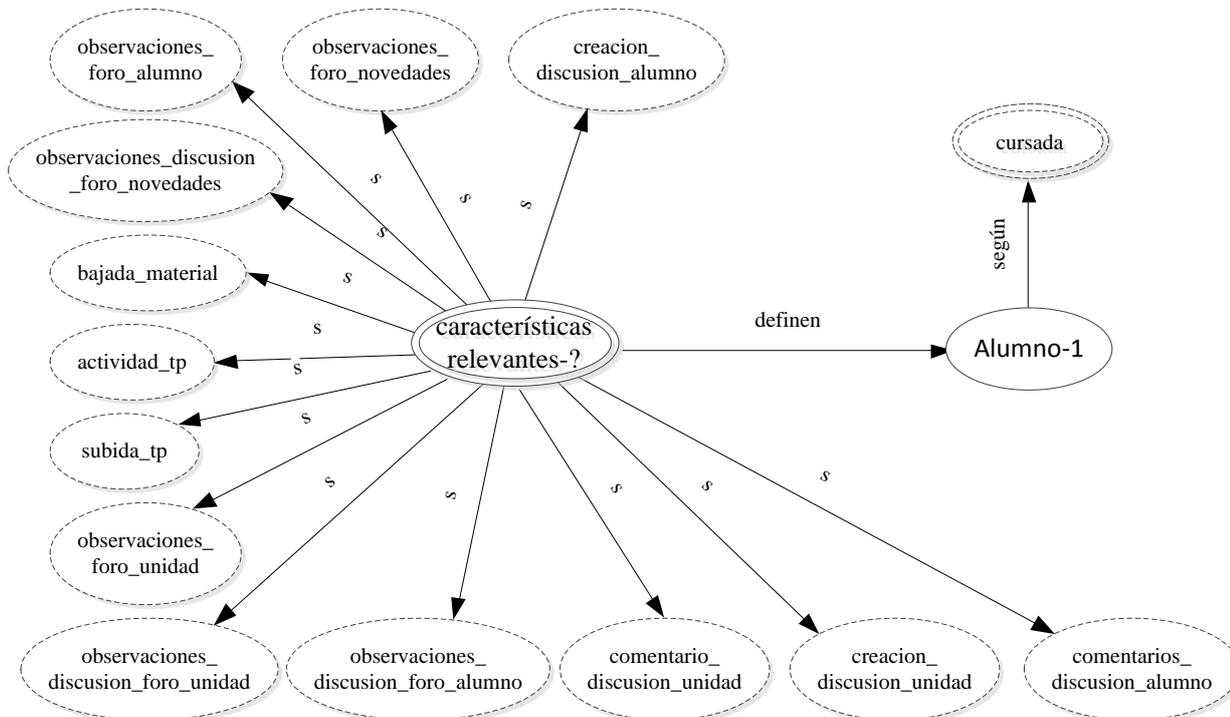


Figura 2. Red Semántica del Problema de Explotación de Información aplicada al caso de estudio

5.4. 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 [27]. El proceso de explotación de información obtenido es: *el proceso de Ponderación de Interdependencia de atributos*.

5.5. 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:

1. bajo_material
2. actividad_tp
3. subio_tp
4. vio_foro_unidad
5. vio_discusion_unidad
6. comentario_discusion_unidad
7. creo_discusion_unidad
8. vio_foro_alumno
9. vio_discusion_alumno
10. comentario_discusion_alumno
11. creo_discusion_alumno
12. vio_foro_novedades
13. vio_discusion_novedades
14. 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.

5.6. 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:

- comentario_discusion_unidad = NO, con un 91% de incidencia.
- vio_discusion_novedades = NO, con un 86% de incidencia.
- comentario_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 tienen mayor influencia para determinar que un alumno no apruebe la cursada son: si no comenta en algún foro de alguna unidad, si no observa las discusiones en los foros de novedades, si no comenta en las discusiones en los foros de alumnos y si no realiza las actividades. Mientras que para determinar que un alumno aprueba la cursada, son: que realice las actividades, que vea las discusiones y los foros de las unidades.

Conclusiones

En este trabajo se ha presentado 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 cursan en la modalidad virtual una materia del área de ciencias básicas cuya primera cursada fue presencial. Este caso difiere de los descritos en [16-17] en que a diferencia de los alumnos considerados en aquellos casos, en el que se presenta aquí los alumnos

tuvieron un primer acercamiento presencial a los contenidos.

La población estudiada estuvo integrada por 363 estudiantes, que al término del curso virtual aprobaron el curso 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 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 mayor en un orden de magnitud, puede hacer emerger atributos específicos que permitan refinar en distintos grupos las causales de desaprobación.

El proceso propuesto constituye una nueva herramienta de diagnóstico en el área; y si bien los resultados obtenidos en este estadio son consistentes, solo sirven para ilustrar la aplicación de la solución propuesta.

Los próximos pasos de esta investigación son: replicar la aplicación del proceso propuesto a una población mayor; y contrastar el proceso en otras asignaturas mediadas por EVEA a efectos de evaluar si los comportamientos descubiertos en las poblaciones son semejantes o hay especificidades propias dependientes de los contenidos.

Agradecimientos

Las investigaciones que se reportan en este capítulo han sido financiadas parcialmente por los Proyectos de Investigación 33A166 y 33A167 de la Secretaría de Ciencia y Técnica de la Universidad Nacional de Lanús (Argentina); y por Beca BENTR13 de la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (Argentina).

Referencias

- [1] 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.
- [2] Schiefer, J., Jeng, J., Kapoor, S. & Chowdhary, P. (2004). Process Information Factory: A Data Management Approach for Enhancing Business Process Intelligence. *Proceedings 2004 IEEE International Conference on E-Commerce Technology*. Pp. 162- 169.
- [3] Curtis, B., Kellner, M., Over, J. (1992). Process Modelling. *Communications of the ACM*, 35(9): 75-90.
- [4] 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 E-Commerce Technology*. Pp. 471-474.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] García Martínez, R., Servente, M. y Pasquini, D. (2003). *Sistemas Inteligentes*. Editorial Nueva Librería. ISBN 987-1104-05-7
- [9] Aretio, L. G. (1987). Rendimiento académico y abandono en la educación superior a distancia. *Universidad Nacional de Educación a Distancia, UNED*.
- [10] Quesada Catillo, R. (2006). Evaluación del aprendizaje en la educación a distancia “en línea”. *RED Revista de Educación a Distancia*.
- [11] Charum, V, 2007. *Modelo de Evaluación de Plataformas Tecnológicas*. Tesis de Magíster en Telecomunicaciones. Escuela de Posgrado, Instituto Tecnológico de Buenos Aires.
- [12] Luan, J. (2002). Data mining and its applications in higher education. *New directions for institutional research*, 2002(113), 17-36.
- [13] 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.
- [14] Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1): 135-146.
- [15] 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.

- [16] Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, 51(1), 368-384.
- [17] 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. *33rd Annual IEEE Frontiers in Education*, Vol. 1, pág. T2A-13.
- [18] Bravo, J., & Ortigosa, A. (2009). Detecting Symptoms of Low Performance Using Production Rules. *International Working Group on Educational Data Mining*.
- [19] Dekker, G., Pechenizkiy, M., & Vleeshouwers, J. (2009). Predicting Students Drop Out: A Case Study. *EDM*, 9, 41-50.
- [20] Jiménez Rey, E., Rodríguez, D., Britos, P., García-Martínez, R. (2008). Identificación de Problemas de Aprendizaje de Programación con Explotación de Información. *Proceedings XIV Congreso Argentino de Ciencias de la Computación*, Artículo 1881. ISBN 978-987-24611-0-2.
- [21] Britos, P., Jiménez Rey, E., García-Martínez, E. (2008). Work in Progress: Programming Misunderstandings Discovering Process Based On Intelligent Data Mining Tools. *Proceedings 38th IEEE Frontiers in Education Conference*. Session F4H: Assessing and Understanding Student Learning. ISBN 978-1-4244-1970-8.
- [22] P Britos, Z Cataldi, E Sierra, R García-Martínez (2008). Pedagogical Protocols Selection Automatic Assistance. *New Frontiers in Applied Artificial Intelligence*. Pag. 331-336. Springer Berlin.
- [23] 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.
- [24] 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.
- [25] Saavedra-Martinez, P., Pollo-Cattaneo, F., Britos, P., Rodríguez, D., García-Martínez, R. 2012. Explotación de Información Aplicada a Identificación de Fallas de Apropiación de Conceptos. *Proceedings IX Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento*. Pág. 103-110. Sello Editorial de la PUCP. ISBN 978-612-4057-85-4.
- [26] Cigliuti, P., Pollo Cattaneo, M., García Martínez, R. (2012). Procesos de identificación de comportamiento de comunidades educativas centradas en EVEAs. *Proceedings XIV WICC*. Pág. 954-956. ISBN 978-950-766-082-5.
- [27] 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.
- [28] Martins, S., Rodríguez, D., García-Martínez, R. 2014. Derivación del Proceso de Explotación de Información desde el Dominio de Negocio. Capítulo X en "Ingeniería del Software e Ingeniería del Conocimiento: Dos Disciplinas Interrelacionadas". Pág. 159-177. Sello Editorial de la Universidad de Medellín. ISBN 978-958-8815-31-2.

Dirección de Contacto de los Autores:

Pablo Cigliuti
Medrano 951 3er Piso,
Departamento de Ingeniería de Sistemas de Información
Universidad Tecnológica Nacional - FRBA
Ciudad Autónoma de Buenos Aires
Argentina
e-mail: pabcig@gmail.com
sitio web: <http://www.sistemas.frba.utn.edu.ar/>

Sebastian Martins y Ramón García-Martínez
29 de Septiembre 3901 (1826)
Laboratorio de Ingeniería de Explotación de Información
Licenciatura en Sistemas
Universidad Nacional de Lanús
Remedios de Escalada, Lanús
e-mail: martinsebastian@yahoo.com.ar,
rgm1960@yahoo.com
sitio web: <http://www.unla.edu.ar/sistemas/gisi/LIDIEI.htm>

Ing. Pablo Cigliuti. Ingeniero en Sistemas por UTN. Cándido del Programa de Maestría en Tecnología Informática Aplicada en la Educación. UNLP. Tesista del Laboratorio de Ingeniería de Explotación de Información, UNLa.

Lic. Sebastian Martins. Licenciado en Sistemas por UNLa. Investigador Asistente del Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información de la Licenciatura en Sistemas, UNLa.

Dr. Ramón García Martínez. Doctor en Informática por Universidad Politécnica de Madrid. Director del Laboratorio de Investigación y Desarrollo en Ingeniería de Explotación de Información de la Licenciatura en Sistemas, UNLa.
