



UNIVERSIDAD TÉCNICA PARTICULAR DE LOJA
La Universidad Católica de Loja

ÁREA TÉCNICA

TITULACIÓN DE INGENIERO EN SISTEMAS INFORMATICOS Y
COMPUTACION

Análisis de patrones de navegación para determinar el comportamiento en los
estudiantes de la UTPL que hacen uso del EVA, mediante la aplicación de
minería de datos.

TRABAJO DE FIN DE TITULACIÓN.

AUTOR: Ruiz Chávez, Fanny Elizabeth

DIRECTOR: Valdiviezo Díaz, Priscila Marisela, Mg.

LOJA – ECUADOR
2014

APROBACIÓN DEL DIRECTOR DEL TRABAJO DE FIN DE TITULACIÓN

Magíster.

Priscila Maricela Valdiviezo Díaz

DOCENTE DE LA TITULACIÓN

De mi consideración:

El presente trabajo de fin de titulación: **Análisis de patrones de navegación para determinar el comportamiento en los estudiantes de la utpl que hacen uso del EVA, mediante la aplicación de técnicas de minería de datos** realizado por: FANNY ELIZABETH RUIZ CHÁVEZ ha sido orientado y revisado durante su ejecución, por lo que se aprueba la presentación del mismo.

Loja, Abril de 2014

f)

DECLARACIÓN DE AUTORÍA Y CESIÓN DE DERECHOS

“ Yo, Fanny Elizabeth Ruiz Chávez, declaro ser autora del presente trabajo de fin de titulación: **Análisis de patrones de navegación para determinar el comportamiento en los estudiantes de la UTPL que hacen uso del EVA, mediante la aplicación de técnicas de minería de datos**, de la Titulación de Sistemas Informáticos Computación, siendo Mg. Priscila Valdiviezo directora del presente trabajo; y eximo expresamente a la Universidad Técnica Particular de Loja y a sus representantes legales de posibles reclamos o acciones legales. Además certifico que las ideas, conceptos, procedimientos resultados vertidos en el presente trabajo investigativo, son de mi exclusiva responsabilidad.

Adicionalmente declaro conocer y aceptar la disposición del Art. 67 del Estatuto Orgánico de la Universidad Técnica Particular de Loja, que en su parte pertinente textualmente dice: “Forman parte del patrimonio de la Universidad la propiedad intelectual de investigaciones, trabajos científicos o técnicos y tesis de grado que se realicen a través o con el apoyo financiero, académico o institucional (operativo) de la Universidad”.

f

Autor: Fanny Elizabeth Ruiz Chavez

Cédula: 1103927362

DEDICATORIA

La presente tesis se la dedico con mucho cariño y gratitud a todos los que de una u otra manera colaboraron para su desarrollo, tanto con su apoyo moral, intelectual y espiritual.

A Dios, por permitirme no darme por vencida y saber enfrentar los obstáculos en mi diario vivir, por su infinita sabiduría, amor y fortaleza puesta en mi persona para conseguir las metas propuestas en el transcurso de mi vida.

A mis padres por la formación que me dieron y los valores que me enseñaron en mi niñez, porque con sus consejos, palabras y advertencias construyeron las bases para mi formación personal y lograr las metas propuestas en mi vida. Por enseñarme que las cosas se las consigue con perseverancia, paciencia y mucho amor, que la Fe es parte fundamental en nuestras vidas y que una caída no es más que una razón para seguir adelante, ya que de los errores se aprende a ser mejores.

A mi madre, Fanny Chávez por estar siempre a mi lado apoyándome en todo, por sus sabios consejos, su paciencia y amor puro y desinteresado. Gracias madrecita por ser la mejor madre del mundo a ti va dedicado este logro.

A mi esposo, Joffre Betancourt por apoyarme en todo, por ser un pilar muy importante en el transcurso de estudio y en el desarrollo de mi tesis. Gracias mi amor, Te amo. A mis hermanos, Vanessa, Galo y Manuel por su apoyo constante y consejos y a todas mis amigos y amigas.

A mi directora de tesis, Ing. Priscila Valdiviezo, a quien estoy infinitamente agradecida, ya que con su apoyo, y exigencias me ayudaron a culminar la tesis, por compartir sus conocimientos conmigo, por sus consejos y ayuda en el desarrollo de mi tesis.

Tita

AGRADECIMIENTO

Expreso mis más sinceros agradecimientos a todas las personas que creyeron y confiaron en mí, a Dios, mi madre, amigos y los profesores que aportaron con sus conocimientos impartidos durante mi transcurso de formación estudiantil, a las personas que de una u otra manera colaboraron con aportaciones y me ayudaron a culminar mi tesis.

Principalmente le agradezco a mi directora de tesis Ing. Priscila Valdiviezo Díaz por su ayuda y apoyo, y sobre todo por confiar en mi persona y asignarme este tema, ya que de no ser así no hubiese sido posible el desarrollo de esta tesis. Le gratifico también que gracias a sus constantes exigencias, consejos y cambios me ayudo a adquirir más conocimientos y habilidades relacionados al tema.

Le agradezco a mi madre por ser mi base en el transcurso de mi camino y por el apoyo constante que me brindo en el todo el transcurso de mi vida, y por brindarme la oportunidad de ser una profesional y la persona que soy ahora.

Les agradezco a mis amigos que me apoyaron constantemente y me brindaron toda su confianza y ánimos para continuar, de manera especial a la Lcda. Mónica Cajamarca que es una excelente amiga, me ha sabido brindar su confianza y apoyo en todo momento de mi formación en la Universidad, con sus consejos y palabras de aliento a cada momento en el transcurso de mi tesis. Al Ing. Greyson Alberca y Héctor Gómez los cuales me brindaron consejos, sugerencias y pautas que me sirvieron de guía para el desarrollo de la tesis y por todos sus conocimientos brindados en el transcurso de la misma.

Con mucho cariño y aprecio.

Tita

ÍNDICE DE CONTENIDOS

CARATULA.....	i
APROBACION DEL DIRECTOR DEL TRABAJO DE FIN DE TITULACIÓN.....	ii
DECLARACIÓN DE AUTORÍA Y SECIÓN DE DERECHOS.....	iii
DEDICATORIA.....	iv
AGRADECIMIENTO.....	v
ÍNDICE DE CONTENIDOS.....	vi
RESUMEN.....	1
ABSTRACT.....	2
INTRODUCCIÓN.....	3
CAPITULO I: ESTADO DEL ARTE.....	6
1. Minería de Datos	7
1.1. Clases de minería Web	7
1.1.1. Minería de utilización Web.....	8
1.1.2. Minería de contenidos Web.....	8
1.1.3. Minería de estructura Web.....	8
1.2. Minería de Uso Web.....	9
1.3. Objetivos de la Minería de Uso Web.....	10
1.4. Fases de la Minería de Uso Web.....	10
1.4.1. Recolección de datos.....	11
1.4.2. Pre-procesamiento de datos.....	11
1.4.3. Inferencia de patrones.....	12
1.4.4. Aplicar la minería de datos.....	12
1.4.5. Extracción de conocimiento.....	12
1.4.6. Análisis de patrones.....	12
1.4.7. Interpretación y evaluación de resultados.....	13
1.5. Técnicas de Minería de Uso Web.....	13
1.5.1. Puntos de vista de orientación de la aplicación de técnicas de minería de datos.....	13
1.5.1.1. Orientado hacia el tutor.....	13
1.5.1.2. Orientado hacia los estudiantes.....	13
1.6. Clasificación de las técnicas de minería.....	14
1.7. Algoritmos de minería de datos.....	15
1.7.1. Reglas de asociación.....	16
1.7.2. Reglas de asociación secuenciales.....	16
1.7.3. Análisis de secuencias.....	16

1.7.4.	Clasificación y Clustering.....	17
1.7.5.	Árboles de decisión.....	18
1.8.	Tipos de algoritmos de minería de datos.....	18
1.8.1.	Algoritmos de clasificación.....	18
1.8.2.	Algoritmos de regresión.....	18
1.8.3.	Algoritmos de segmentación.....	18
1.8.4.	Algoritmos de asociación.....	19
1.8.5.	Algoritmos de análisis de secuencias.....	19
1.9.	Descripción de los algoritmos para la minería.....	19
1.9.1.	GSP (GeneralizedSequentialPatterns).....	19
1.9.2.	A priori.....	20
1.9.3.	Simple KMeans.....	21
1.9.4.	Algoritmo EM.....	21
1.9.5.	Algoritmo J48.....	21
2.	Herramientas para el análisis de Logs	22
2.1.	Logs.....	22
2.2.	Herramientas incorporadas al servidor.....	24
2.3.	Herramientas en máquinas personales.....	24
3.	Sistema hipermedia adaptativo.....	24
3.1.	Técnicas de adaptación a la navegación.....	25
3.2.	Soporte a la navegación adaptativa.....	25
3.2.1.	Métodos.....	25
3.2.2.	Técnicas.....	26
3.3.	Identificación de patrones secuenciales.....	27
3.4.	Patrones de navegación.....	28
4.	Soporte adaptativo al aprendizaje.....	30
4.1.	Componentes del soporte adaptativo.....	30
4.1.1.	Modelo del usuario.....	30
4.1.2.	Modelo del dominio.....	30
4.1.3.	Modelo del profesor.....	30
4.2.	Objetivos del soporte adaptativo.....	31
5.	Modelo de estudiante abierto.....	32
5.1.	Soporte adaptativo a la navegación con OLM.....	32

5.1.1.	Actividades de adaptación.....	33
5.1.2.	Actividades de apoyo a la navegación.....	34
5.1.3.	Colaboración adaptativa en las actividades de apoyo a la navegación.....	34
5.1.4.	Adaptación en las actividades de apoyo a la colaboración.....	34
5.1.4.1.	Predicción del nivel de actividad.....	34
6.	Trabajos relacionados.....	35
6.1.	Reglas de asociación.....	35
6.2.	Clustering.....	37
6.3.	Análisis de secuencias.....	37
Capítulo II: ANÁLISIS DE LA PLATAFORMA DEL EVA DE LA UTPL.....		39
2.	Análisis de la plataforma del EVA de la UTPL.....	40
2.1.	Introducción.....	40
2.2.	Moodle como sistema de enseñanza.....	40
2.3.	Tablas de información relacionadas al estudiante.....	41
2.3.1.	Diagrama Entidad – Relación.....	42
2.3.2.	Tablas del EVA para la minería de datos.....	42
2.3.3.	Atributos de las entidades en el (EVA) que se trabajaran para la minería.....	43
2.3.4.	Descripción de los datos de usuario en el EVA.....	44
2.3.5.	Descripción de los atributos de interacción en el EVA.....	45
2.4.	Descripción de las herramientas dentro de la plataforma.....	47
2.4.1.	Foros.....	47
2.4.2.	Recursos.....	47
2.4.3.	Tareas.....	48
2.4.4.	Chat.....	49
2.4.5.	Consulta.....	49
2.4.6.	Cuestionario.....	49
2.4.7.	Lección.....	50
2.4.8.	Blog.....	51
2.4.9.	Recursos Educativos Abiertos.....	51
2.5.	Datos para la obtención de rutas de navegación de los estudiantes en el EVA.....	51
2.6.	Herramientas de interacción en el EVA.....	53
Capítulo III: ANÁLISIS Y SELECCIÓN DE HERRAMIENTAS.....		55
3.	Análisis y selección de herramientas y técnicas.....	56
3.1.	Introducción.....	56

3.2.	Selección de herramientas.....	56
3.2.1.	Herramientas para la recolección de datos.....	56
3.2.2.	Herramientas para la extracción y procesamiento de datos del EVA.....	57
3.2.2.1.	XAMPP Control.....	57
3.2.2.2.	Navicat for MySQL.....	57
3.2.2.3.	MySQL--connector--java--5.1.17.....	58
3.3.	Herramientas para la minería de datos.....	58
3.3.1.	Webalizer.....	59
3.3.2.	Sawmill.....	60
3.3.3.	Weka.....	61
3.4.	Características de Weka.....	61
3.5.	Filtros de atributos en Weka.....	62
3.5.1.	Filtro de discretización.....	62
3.5.2.	Filtro de selección.....	62
3.5.3.	Filtro de instancias.....	63
3.6.	Tabla comparativa de las herramientas para el pre-procesamiento de datos.....	63
3.7.	Herramientas para el descubrimiento de patrones.....	64
3.7.1.	Técnicas para la minería de datos.....	64
3.7.2.	Herramienta para el análisis grafico.....	65
3.7.2.1.	Herramienta Gephi para la visualización de grafos.....	66
Capítulo IV: pre-proceso: selección, limpieza y transformación de datos.....		67
4.	pre-proceso: selección, limpieza y transformación de datos.....	68
4.1.	Introducción.....	68
4.2.	Información de la base de datos del EVA.....	68
4.3.	Proceso de análisis de los Logs.....	70
4.3.1.	Pre-proceso de los datos del EVA.....	70
4.3.2.	Selección, limpieza y transformación de datos.....	71
4.3.2.1.	Reconocimiento y Limpieza de Datos.....	71
4.3.2.2.	Resúmenes de Estadísticas.....	72
4.3.2.3.	Diagramas.....	72
4.4.	Selección de datos para el análisis.....	72
4.4.1.	Datos de usuario.....	72
4.4.2.	Datos de interacción.....	72
4.5.	Descripción del proceso de extracción de datos.....	73
4.5.1.	Datos de extracción.....	74

4.5.2.	Limpieza de datos.....	75
4.5.3.	Transformación de datos.....	75
4.6.	Tablas de la base de datos del EVA.....	76
4.6.1.	Tabla creada para almacenar los datos de interacción de los estudiantes...	77
4.7.	Resultados obtenidos de la interacción con las herramientas.....	78
4.8.	Promedio de interacción de los estudiantes con las herramientas del EVA.....	81
4.9.	Datos de navegación de los estudiantes en el EVA.....	82
Capítulo V: Minería de datos: Experimentación y resultados.....		83
5.	Minería de datos: Experimentación y resultados.....	84
5.1.	Introducción.....	84
5.2.	Clasificación de páginas del EVA.....	84
5.3.	Descripción del escenario.....	86
5.4.	Descripción del proceso de experimentación.....	88
5.5.	Obtención de datos de navegación a partir de los datos del usuario.....	90
5.6.	Aplicación de los algoritmos seleccionados.....	92
5.6.1.	Experimentación con la técnica de asociación.....	93
5.6.1.1.	Aplicación del algoritmo GSP.....	93
	Resultados obtenidos de la experiencia.....	94
5.6.1.2.	Aplicación del algoritmo A-priori.....	95
	Resultados obtenidos de la experiencia.....	98
5.6.2.	Experimentación con la técnica de clustering.....	99
5.6.2.1.	Aplicación del algoritmo Simple-KMeans.....	99
	Resultados obtenidos de la experiencia.....	102
5.6.2.2.	Aplicación del algoritmo EM.....	102
	Resultados obtenidos de la experiencia.....	107
5.6.3.	Experimentación con la técnica de clasificación con Árboles de decisión..	107
5.6.3.1.	Aplicación del algoritmo J-48.....	108
	Análisis e interpretación de resultados de la experimentación.....	111
5.6.4.	Resumen de los resultados obtenidos en la minería de los datos.....	112
5.7.	Gephi para la visualización de las interacciones de los usuarios en el EVA.....	114
5.7.1.	Interpretación de los datos del grafo.....	117
5.7.2.	Obtención de resultados de interacción en los cursos de experimentación.....	118
5.7.3.	Conclusiones.....	121

5.8.	Información de la navegación de los estudiantes en las páginas del EVA...	122
5.8.1.	Obtención de resultados de interacción en las páginas del EVA.....	122
5.8.2.	Obtención de resultados de interacción en las tareas.....	123
5.8.3.	Obtención de resultados de interacción en los foros.....	125
5.8.4.	Obtención de resultados de interacción en los recursos.....	126
5.8.5.	Obtención de resultados de interacción en los cuestionarios.....	128
5.8.6.	Obtención de resultados de interacción en el perfil de usuario.....	129
5.8.7.	Obtención de resultados de interacción en el blog.....	131
5.8.8.	Obtención de resultados de interacción en la página UTPL.....	132
5.8.9.	Obtención de resultados de interacción en la página de subir archivos.....	133
5.9.	Resultados de interacción de los estudiantes en el EVA.....	134
5.10.	Validación de resultados obtenidos con las técnicas de minería frente a Gephi.....	135
Capítulo VI: Representación de indicadores y atributos en modelo de datos.....		138
6.	Representación de indicadores y atributos en modelo de datos.....	139
6.1.	Introducción.....	139
6.2.	Resultados del comportamiento de navegación de los estudiantes.....	139
6.3.	Interacción de los estudiantes en el EVA.....	140
6.4.	Indicadores obtenidos en las actividades de aprendizaje.....	141
6.4.1.	Nivel de interés en los cursos.....	142
6.4.2.	Nivel de interés en páginas de Actividades.....	143
6.4.3.	Nivel de interés en páginas de Apoyo.....	143
6.4.4.	Nivel de interacción en los cursos.....	143
6.4.5.	Nivel de colaboración en los cursos.....	144
6.5.	Metodología Entidad – Relación.....	144
6.5.1.	Modelo Entidad – Relación.....	145
6.5.2.	Elementos del esquema E–R.....	145
6.6.	Modelo de datos del usuario con los indicadores obtenidos.....	146
6.6.1.	Modelo de usuario.....	147
DISCUSIÓN.....		149
CONCLUSIONES.....		152
RECOMENDACIONES.....		155
TRABAJOS FUTUROS.....		157
BIBLIOGRAFIA.....		158
ANEXOS.....		161

RESUMEN:

La inclusión de Entornos Virtuales de Aprendizaje, en todas las áreas como métodos de aprendizaje tanto para estudiantes y docentes es total y ayudan a la mejora constante en el aprendizaje del estudiante y comunicación entre docente-estudiante. En el presente trabajo se pretende analizar y aplicar varias técnicas y métodos de minería de datos con el fin de determinar patrones de comportamiento en la navegación de los estudiantes durante su interacción con el EVA. Actualmente existen varias investigaciones sobre la aplicación específica de técnicas tanto de minería de datos a los sistemas de enseñanza a distancia, como a los sistemas basados en Web o sistemas de e-Learning; las mismas que tienen un gran crecimiento, por lo que su unión está despertando interés entre los investigadores y empresas de ambas áreas.

Para la realización del presente proyecto se pretende hacer un estudio de diferentes algoritmos que demuestran el uso de técnicas como: **Reglas de Asociación** (*A-priori, GSP*), **Clasificación** (*Árboles de decisión, J-48*) **Clustering** (*SimpleKMeans, EM*), y con los resultados obtenidos determinar el comportamiento de navegación de los estudiantes con el EVA.

PALABRAS CLAVES: Minería de datos, e-Learning, Reglas de Asociación, Clustering, Clasificación, rutas de navegación, patrones.

ABSTRACT:

The inclusion of Virtual Learning Environments, in all areas as methods of learning for both students and teachers is help full and continuous improvement in student learning and teacher - student communication. The present work aims to analyze and apply various techniques and methods of data mining in order to determine patterns in the navigation of students during their interaction with the EVA. Currently there are many studies on the specific application of techniques from both data mining systems distance learning, as Web -based systems or e - learning systems; thereof having a great growth, so that their union is arousing interest among researchers and companies in both areas.

For the realization of this project is to make a study of different algorithms that demonstrate the use of techniques such as **Association Rules** (*A- priori, GSP*), **classification** (*Decision trees, J-48*) **Clustering** (*SimpleKMeans, EM*) and the results determine the browsing behavior of students with the EVA.

KEYWORDS: Data mining, e_Learning, association rules, clustering, classification, navigation paths, patterns.

INTRODUCCIÓN:

La presente investigación pretende buscar y consolidar los principales patrones de navegación interesantes en la interacción de los estudiantes con el EVA; para lo cual se pretende realizar un análisis en los modelos de acceso de los estudiantes a la información que se obtiene de los *Logs*. Esta entidad, almacena las actividades que se producen dentro de la base de datos como consecuencia directa de la navegación de los estudiantes en el EVA, y a esta información se le aplicarán técnicas de minería de datos, para encontrar el conocimiento sobre la forma en la cual los estudiantes navegan a través de la misma. El proceso con las técnicas de minería de datos, será centrado principalmente en el uso de herramientas de e-learning; ya que su principal objetivo es guiar a los estudiantes durante su aprendizaje para maximizarlo; es por eso que este estudio se enfoca en la investigación del descubrimiento de reglas de asociación, clasificación y clustering ya que son las técnicas más utilizadas para determinar ciertos comportamientos en la interacción de los estudiantes con el sistema, y de esta manera determinar la herramienta o técnica con la que se va a trabajar para conseguir estos patrones en la interacción con las herramientas del EVA.

Entre los algoritmos mejor conocidos dentro de las técnicas de minería están las reglas de asociación, las cuales permiten encontrar dependencias entre elementos, en muchas ocasiones estas reglas se usan como punto de partida cuando no se sabe con exactitud qué tipos de patrones se está buscando, también se utilizan estas técnicas de minería de reglas de asociación y filtrado colaborativo para descubrir patrones de navegación útiles, con la finalidad de proponer un modelo de usuario según la navegación del usuario en la Web. También existen otros algoritmos dentro de las técnicas de agrupamiento (clustering), como son los vecinos más cercanos o *k-means* que encuentran relaciones entre grupos de objetos con ciertas afinidades o características similares. Estos algoritmos junto con otros algoritmos más, dentro de las técnicas de minería de datos, permiten clasificar a los estudiantes para predecir su clasificación final basándose en las características extraídas de las actividades y eventos almacenados en la tabla de *Log* del EVA.

El desarrollo de la presente tesis, está principalmente centrada en el proceso de extracción de datos de interacción de los estudiantes en la plataforma de aprendizaje de la UTPL (EVA), y a partir de estos datos generados, extraer los patrones de navegación en el análisis de interacción de los estudiantes con las herramientas del EVA, con la finalidad de determinar los atributos e indicadores que tienen los estudiantes y representarlo en un modelo de datos del estudiante con los indicadores obtenidos de una metodología de estudiante abierto.

FASES Y ACTIVIDADES DEL PROYECTO DE TESIS

FASE 1: ESTADO DEL ARTE

Estado del arte sobre la aplicación de técnicas de minería en entornos educativos.

FASE 2: ANÁLISIS DE LA PLATAFORMA DEL EVA DE LA UTPL

Breve descripción con las utilidades del EVA para analizar módulos y recursos con los cuales el estudiante cuenta para interactuar con el sistema, para el proceso de la minería.

Las actividades a realizar en esta fase son:

- Diagrama Entidad–Relación con entidades relacionadas a la navegación del estudiante.
- Realizar una extracción de las principales entidades para la minería de datos.
- Identificación de los atributos de interacción de los estudiantes con el EVA.
- Extracción de herramientas del EVA con los cuales interactúa el estudiante en el sistema.

FASE 3: ANÁLISIS Y SELECCIÓN DE HERRAMIENTAS Y TÉCNICAS

Selección de herramientas para el proceso de extracción de la información del EVA, y la selección técnicas para cada una de las fases del minado.

Las actividades a realizar en esta fase son:

- Selección de herramientas para el proceso de recolección y extracción de datos.
- Selección de las herramientas para el pre-proceso y proceso de minería de datos.
- Selección de las herramientas para el análisis de la tabla de Logs.
- Selección de las técnicas y algoritmos para el proceso de minería de los datos
- Selección de la herramienta para la visualización de los datos de interacción (grafos).

FASE 4: PRE-PROCESO: SELECCIÓN, LIMPIEZA Y TRANSFORMACIÓN DE DATOS

Extracción de datos mediante consultas en lenguaje SQL de la base de datos, luego se le realizará una limpieza, seleccionando sólo aquellos datos necesarios para el estudio; y se los transformará a un formato entendible para la herramienta del proceso de minería.

Las actividades a realizar en esta fase son:

- Selección de la información del EVA para el pre-proceso de minería.
- Limpieza de los datos seleccionados mediante consultas en lenguaje SQL
- Transformación de los datos extraídos a un formato legible para Weka.

FASE 5: MINERÍA DE DATOS: EXPERIMENTACIÓN Y RESULTADOS

En esta fase se realizarán varias experimentaciones con cada uno de las técnicas de minería seleccionados en la fase anterior y con cada uno de los algoritmos propuestos.

Las actividades a realizar en esta fase son:

- Aplicar varias técnicas de minería para extraer patrones de navegación en los estudiantes, y clasificarlos según sus intereses y accesos a las páginas para determinar las herramientas de interés del estudiante.
- Determinar el nivel de interacción de los estudiantes con cada una de las herramientas del EVA para su aprendizaje.
- Según los resultados obtenidos, seleccionar la técnica y algoritmos más aptos para este tipo de estudio.

FASE 6: REPRESENTACIÓN DE INDICADORES Y ATRIBUTOS DE NAVEGACIÓN DEL USUARIO EN UN MODELO DE DATOS

En esta fase se realizará el análisis de una metodología de modelado de usuario, en este caso se lo realizará con el modelo Entidad - Relación para representar los datos del usuario.

Las actividades a realizar en esta fase son:

- Extraer atributos e indicadores de navegación según el nivel de interacción en las páginas de las diferentes herramientas y servicios de la plataforma.
- Representarlos en un modelo de datos del estudiante los datos extraídos de la actividad anterior.

CAPITULO I:

ESTADO DEL ARTE

1. Minería de datos

Para poder conocer lo que es una minería de datos es importante la citación de varias definiciones de algunos autores del significado de minería de datos, para poder entender y realizar un análisis según ellas:

"La minería de Datos es el proceso de ex0traer conocimiento útil y comprensible, previamente desconocido, desde grandes cantidades de datos almacenados en distintos formatos" [Witten & Frank, 2000].

"Es el uso de datos históricos para descubrir regularidades generales y mejorar las decisiones futuras" [Mitchell, 1999].

"Es el proceso que tiene como objetivo convertir datos en conocimiento" [Hernández, 2004].

"La minería de datos es un paso particular en el proceso de KDD que consiste en la aplicación de algoritmos específicos para extraer patrones (o modelos) desde los datos" [Fayyad, 1996].

La **minería de datos** (o **Data Mining**) según las referencias mencionadas anteriormente, se puede decir que, es un proceso de extracción de información y búsqueda de patrones de comportamiento y el descubrimiento de la información más esencial y útil de los mismos que permanecen ocultos entre grandes cantidades de información, para convertirlo en conocimiento mejorado y útil, mediante la aplicación de varias técnicas.

1.1. Clases de minería Web.

Un caso particular de la minería de datos es la **Minería Web** o **Web Mining**, la cual tiene como objetivo fundamental, la detección de patrones desconocidos y potencialmente útiles en los datos del Web, con el propósito de reformar la toma de decisiones con relación al diseño, contenido y estructura de los sitios Web, para lo cual, aplica técnicas de minería de datos para extraer conocimiento a partir de los datos de la Web.

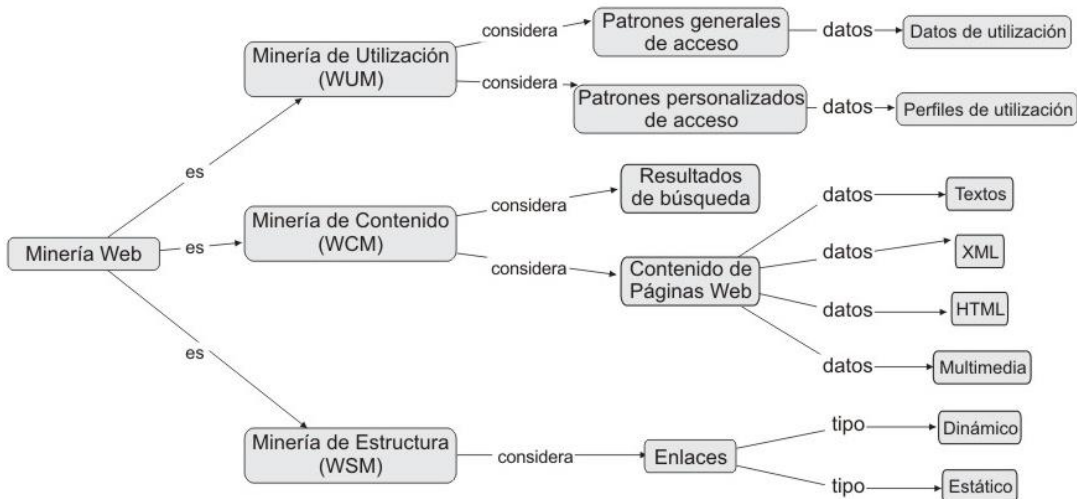


Fig. 1 Mapa Conceptual de la Clasificación de la Minería Web

Fuentes: S. & Ruiz M., 2007.

En materia de minería Web, existen tres clases fundamentales:

1.1.1. Minería de utilización Web.

Es el proceso de descubrir información a partir de los datos de utilización de la Web. Tiene como principal objetivo extraer patrones de uso del Web por parte de los usuarios. Para lo cual, se utilizan los Logs (registros de sucesos y eventos) de los servidores Web.

Este tipo de minería tiene dos objetivos principales:

- Extraer patrones generales de uso de un sitio Web de manera que pueda reestructurarse para que sea más fácil de utilizar y mejore el acceso por parte de los usuarios.
- Obtener perfiles de los distintos tipos de usuarios a partir de su comportamiento y navegación, para ofrecer una atención más personalizada.

1.1.2. Minería de contenidos Web.

Es el proceso de extraer información a partir de los contenidos de los documentos Web.

1.1.3. Minería de estructura Web.

Es el proceso de descubrir información dentro de la Web y permite conocer cómo se organiza a partir de la estructura, de cómo ocurre y la navegación en la Web [Fuentes S. & Ruiz M., 2007].

La **Fig. 1** muestra las tres clases de minería Web que se pueden utilizar para el proceso de la información, de la cual, la clase de minería que más se ha utilizado para el descubrimiento de información en los sistemas de enseñanza basada en Web es la minería de utilización Web o Web usage mining.

El desarrollo de la presente tesis se enfoca principalmente en los conceptos relacionados a la minería del uso de la Web, ya que esta minería sirve específicamente en el descubrimiento de información a partir del aprendizaje y la extracción de patrones de navegación que se le realizará en el EVA.

1.2. Minería de uso Web.

La minería de uso Web tiene dos enfoques importantes dentro de las aplicaciones de descubrimiento; uno es la búsqueda de patrones de acceso general, el cual permite analizar el tráfico para entender los patrones de acceso y comportamiento que tienen los usuarios durante la navegación por la Web y sus tendencias, con el fin de mejorar la estructura del EVA, según las necesidades del usuario; clasificando los contenidos de forma ordenada y más accesible o para ubicar y dirigir a los usuarios hacia lugares de interés, ya sean los más relevantes e importantes para ellos; la segunda tendencia es la búsqueda para personalizar el uso, en la que se analizan las tendencias individuales de cada uno de los usuarios que navegan e ingresan a la Web para personalizar o adaptar dinámicamente la información del sitio Web, su estructura o recursos a cada visitante según el patrón de acceso que exhiba

El proceso de la minería de uso de la Web parte de un proceso previo de los datos, las cuales pasan por fases de pre-proceso como son filtrado y limpieza de datos, así como transformación y agregación de datos. En esta parte del proceso se utilizan diferentes técnicas para analizar y descubrir patrones interesantes en el comportamiento de los usuarios en la navegación dentro de la Web. La información obtenida de este proceso será de gran utilidad para mejorar el rendimiento de los servidores Web y para una mejor usabilidad en el sitio Web por parte del usuario. **[Castaño A., 2009]**.

La minería de utilización Web se centra principalmente en el aprendizaje de patrones de navegación de los usuarios cuando interactúan con la plataforma virtual, y el aprendizaje de perfiles de usuario que sirven para modelar interfaces adaptativos mediante personalización de las mismas, por lo cual, esta clase de Minería de Uso Web, será el centro de estudio para la línea de investigación en la presente tesis; la necesidad de aplicar estas técnicas de minería de datos en el descubrimiento de patrones de uso de la información Web es con el

objetivo de entender y servir de mejor manera los requerimientos del estudiante y de esta manera poder satisfacer las necesidades del mismo.

1.3. Objetivos de la minería de uso Web.

El principal objetivo de la minería de uso Web, es entender de mejor manera y servir mejor las necesidades de los usuarios cuando estos navegan en aplicaciones basadas en la Web.

Las aplicaciones de este tipo de minería pueden clasificarse en:

- Aprendizaje de patrones de navegación.
- Aprendizaje de perfiles de usuario para modelar interfaces adaptativas, (personalizado).

1.4. Fases de la minería de uso Web.

La minería del uso de la Web consiste en tres fases, las cuales son: *pre-procesamiento*, *descubrimiento de patrones* y *análisis de patrones*.

La razón por la que es conveniente realizar un proceso de fases previas a la minería de datos, es debido a que los datos reales pueden ser impuros y esto nos puede conducir a la extracción de reglas poco útiles; La preparación de los datos, ayuda a generar “datos de calidad”, las cuales nos pueden conducir a reglas de calidad.

El pre-procesamiento de datos puede generar un conjunto de datos más pequeño que el original, lo cual mejora la eficiencia de los algoritmos de extracción de reglas.

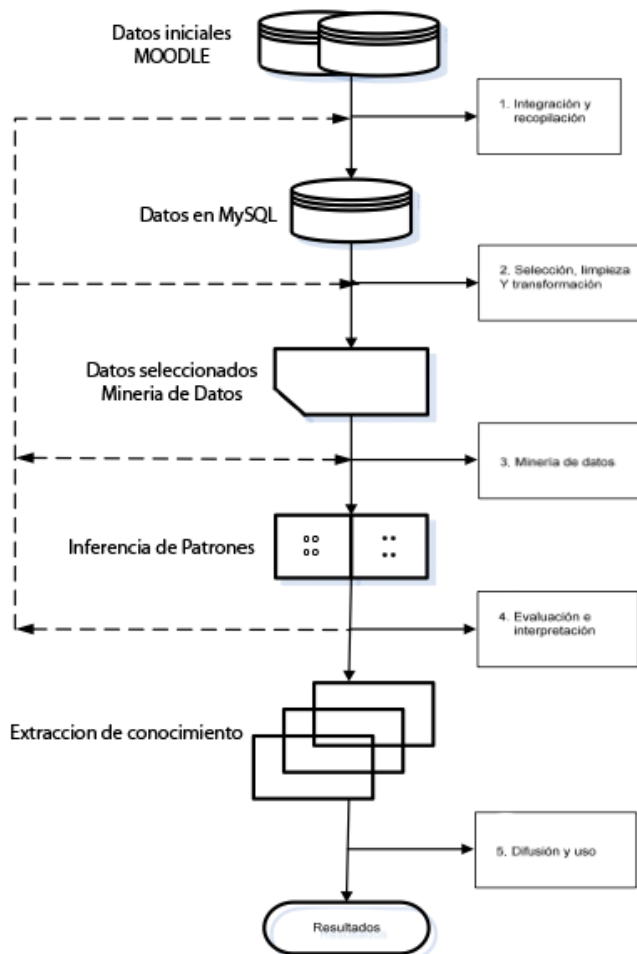


Fig. 2. Fases de la minería de uso Web

Fuente: Adaptado de Alcívar P; Idrovo F; Macas V, 2007.

1.4.1. Recolección de datos.

El sistema LMS es utilizado por los estudiantes y el uso de la información, y la interacción se almacena en la base de datos. En esta actividad se van a utilizar los datos de los estudiantes en la interacción del sistema Moodle.

1.4.2. Pre-procesamiento de datos.

Se deben eliminar el mayor número posible de datos erróneos o inconsistentes e irrelevantes para la limpieza de los datos, transformándolo en un formato adecuado y con información necesaria para ser explotada. Para realizar el pre-proceso de los datos de Moodle, podemos usar una herramienta de administración de bases de datos o alguna herramienta de pre-procesamiento específicas.

1.4.3. Inferencia de patrones.

El descubrimiento de patrones utiliza los métodos y algoritmos desarrollados de varios campos como la estadística, minería de datos, aprendizaje de máquina y reconocimiento de patrones.

Métodos desarrollados desde otros campos deben tomar en consideración los diferentes tipos de abstracciones de datos y el conocimiento previo disponible para la Minería Web. En minería del uso de la Web, una sesión de servidor es una secuencia ordenada de páginas pedidas por el usuario. Además, debido a la dificultad en identificar sesiones únicas, previo conocimiento adicional es requerido (tal como seleccionar un periodo de tiempo por defecto).

1.4.4. Aplicar la minería de datos.

Los algoritmos de minería de datos son aplicados para crear y ejecutar el modelo que se descubre y se resumen los conocimientos de interés para el usuario (profesor, estudiante, administrador, etc.). Para aplicar los algoritmos de minería de datos podemos utilizar una herramienta general o específica de minería de datos, y se puede usar una herramienta de minería de datos comerciales o libres.

1.4.5. Extracción de conocimiento.

Con la implementación de técnicas de minería de datos, lo que se busca obtener de ello, es un modelo de conocimiento, el mismo que representa los patrones de comportamiento observados dentro de los valores de las variables del problema o las diferentes relaciones de asociación que existen entre dichas variables. Además se pueden realizar una selección múltiple de varias técnicas al mismo tiempo, para generar distintos modelos, pero a esto generalmente se le debe realizar obligadamente un pre-procesado diferente de los datos.

1.4.6. Análisis de patrones.

El análisis de los patrones es el último paso en el proceso total de Minería del Uso de la Web. La motivación detrás del análisis de patrones es filtrar reglas no interesantes o patrones de un conjunto encontrado en la fase de descubrimiento de patrones.

La metodología exacta de análisis es usualmente gobernada por la aplicación por la cual la minería Web es hecha. La forma más común de análisis de patrones consiste de un conocimiento de un mecanismo de consulta como SQL. Otro método es cargar los datos de uso en un cubo de datos para realizar operaciones OLAP. Técnicas de visualización tal como gráficos

de patrones o asignación de colores a diferentes valores, pueden frecuentemente señalar el patrón total o tendencias en los datos. El contenido y la información de la estructura pueden ser usadas para filtrar patrones que contienen páginas de un cierto tipo de uso, tipo de contenido, o páginas que coinciden con una cierta estructura de enlaces.

1.4.7. Interpretación y evaluación de resultados.

Los resultados obtenidos son interpretados y utilizados por el docente para realizar nuevas medidas. El profesor puede utilizar la información descubierta para tomar decisiones e los estudiantes y herramientas del curso, con el fin de mejorar el aprendizaje del estudiante.

1.5. Técnicas de minería de uso Web.

Lo primero que se debe realizar antes del proceso de descubrimiento de conocimiento y de aplicar cualquier técnica de minería de datos, es realizar una transformación de los datos para que estos puedan ser manejados de una manera más eficiente en su totalidad; filtrando datos de poco o nulo interés y transformando los datos en un formato más manipulable; a continuación se detallan las técnicas a estudiar y los principales trabajos de investigación agrupados dentro de las dos reglas que estudiaremos en esta investigación; la aplicación automatizada de algoritmos de minería de datos permite detectar fácilmente patrones en los datos, aunque algunos de los investigadores no sólo utilizan una única técnica sino varias.

1.5.1. Puntos de orientación de la aplicación de técnicas de minería de datos.

- **Orientado hacia el tutor.** Con el objetivo de ayudar a los profesores y/o autores de los sistemas de e-learning para que puedan mejorar el funcionamiento o rendimiento de estos sistemas a partir de la información de utilización de los alumnos. Sus principales aplicaciones son: conocer más sobre como los estudiantes aprenden en el Web, obtener una mayor realimentación de la enseñanza, evaluar a los estudiantes por sus patrones de navegación, reestructurar los contenidos el sitio Web para personalizar los cursos, clasificar a los estudiantes en grupos, etc.
- **Orientado a los estudiantes.** Con el objetivo de ayudar o realizar recomendaciones a los alumnos durante su interacción con el sistema para poder mejorar su aprendizaje. Sus principales aplicaciones son: sugerir buenas experiencias de aprendizaje al estudiante, adaptación del curso según el progreso del aprendiz, ayudar al estudiante dando sugerencias y atajos, recomendar caminos más cortos y personalizados, etc.

Las técnicas que se utilizan según el estudio realizado por [Romero, Ventura, de Castro & García, 2005] en la minería de datos aplicada a los sistemas de e-learning son: *clasificación y agrupamiento*, *descubrimiento de reglas de asociación*, y *análisis de secuencias*, de las cuales las dos últimas son de mi interés en el presente estudio, ya que se centran en la búsqueda de patrones de navegación que es el centro de enfoque en el presente análisis.

1.6. Clasificación de las técnicas de minería.

Los algoritmos de minería de datos se clasifican en dos grandes categorías: ***supervisados o predictivos*** y ***no supervisados o de descubrimiento del conocimiento***.

La diferencia entre los algoritmos ***supervisados o predictivos*** con los no supervisados es que como su nombre lo indica son algoritmos que predicen el valor de un atributo, llamado (*etiqueta*) dentro de un conjunto de datos, a partir de otros atributos (*Atributos descriptivos*). El *Aprendizaje supervisado* es una técnica que permite deducir una función a partir de un conjunto de datos de entrenamiento. Los datos de entrenamiento consisten de pares de objetos (*vectores*): una de la componente del par son los datos de entrada y el otro, son los resultados deseados. La salida de la función puede ser un valor numérico (*problemas de regresión*) o una etiqueta de clase (*problemas de clasificación*). Esas relaciones sirven para realizar la predicción en datos cuya etiqueta es desconocida. El Aprendizaje supervisado se desarrolla en dos fases: la primera fase es el ***Entrenamiento*** (construcción de un modelo usando un subconjunto de datos con etiqueta conocida) y la segunda fase que es la ***prueba*** (modelo sobre el resto de los datos).

A diferencia de los *algoritmos supervisados*, es que cuando una aplicación no es lo suficientemente madura no tiene el potencial necesario para una solución predictiva, por lo tanto en ese caso hay que acudir a los métodos ***no supervisados o de descubrimiento del conocimiento***, ya que estos descubren patrones y tendencias en los datos actuales (*no utilizan datos históricos*). El descubrimiento de esa información sirve para llevar a cabo acciones y obtener un beneficio (*científico o de negocio*) de ellas. La **Tabla 1** muestra algunas de las técnicas de minería de ambas categorías [Moreno M.; Miguel L.; García F.; Polo M., s.a].

Según lo mencionado anteriormente, para el análisis de la presente tesis y de acuerdo a la definición de cada una de las categorías de la minería de datos, la minería se la realizará con Algoritmos ***no supervisados o de descubrimiento del conocimiento***, ya que estos algoritmos permiten el descubrimiento de patrones y tendencias dentro de los datos, que será en la tabla de los Log de la base de datos del EVA y además se realizará un experimento con uno de los

algoritmos supervisados (*Árboles de decisión*), ya que este permitirá determinar las rutas de navegación según la secuencia de navegación del estudiante para encontrar los indicadores.

Dentro de los algoritmos supervisados y no supervisados o de descubrimiento del conocimiento que se aplicarán para el proceso de minería de los datos, se los detalla a continuación:

Tabla 1. Clasificación de las técnicas de minería de datos.

SUPERVISADOS	NO SUPERVISADOS
<i>Árboles de decisión</i>	<i>Clustering</i>
<i>Inducción neuronal</i>	<i>Reglas de asociación</i>
<i>Regresión</i>	<i>Patrones secuenciales</i>
<i>Series temporales</i>	<i>Segmentación</i>

Fuente: Autora de la tesis

De los técnicas mencionados en la **Tabla 1**, tres de ellas serán tomados para la minería de los datos del EVA, Clustering, y Reglas de asociación (algoritmos de la categoría de los *Algoritmos no supervisados o de descubrimiento del conocimiento*), y Árboles de decisión (*categoría de los algoritmos supervisados*), se implementara estos algoritmos con la finalidad de descubrir los patrones de navegación en la interacción de los estudiantes con la plataforma del EVA.

1.7. Algoritmos de minería de datos.

Uno de los principales objetivos dentro de la minería de datos es establecer el proceso de descubrimiento o extracción de conocimiento, conocido como (*Knowledge Discovery in Databases, KDD*) en inglés, el mismo que trata de obtener la extracción no trivial de información potencialmente necesaria, útil, válida, novedosa y perceptible a partir de un gran volumen de datos.

Según un estudio realizado por [TechNet, 2011] Un algoritmo analiza primero un conjunto de datos y luego busca patrones y tendencias específicos. El algoritmo utiliza los resultados de este análisis para definir los parámetros del modelo de minería de datos. A continuación, estos parámetros se aplican en todo el conjunto de datos para extraer patrones procesables y estadísticas detalladas.

El modelo de minería de datos que crea un algoritmo puede tomar diversas formas, incluyendo:

- Un conjunto de reglas que describen cómo se agrupan los productos en una transacción.
- Un árbol de decisión que predice si un cliente determinado comprará un producto.

- Un modelo matemático que predice las ventas.
- Un conjunto de clústeres que describe cómo se relacionan los casos de un conjunto de datos.

1.7.1. Reglas de asociación.

Las reglas de asociación permiten determinar patrones en los conjuntos de datos en los que ocurren transacciones de datos. Con esta técnica, pueden encontrarse relaciones sin que exista intervención alguna por parte de algún operador. El aprendizaje de reglas de asociación se divide normalmente en dos fases: **[Fuentes S. & Ruiz M., 2007]**.

1. Extracción de los conjuntos de ítems que cumplen con la cobertura requerida a partir de los datos.
2. Generación de las reglas a partir de estos documentos.

Las reglas de asociación descubren relaciones entre atributos de un conjunto de datos que superan unos determinados umbrales. Su aplicación a más de e-learning permite descubrir relaciones o asociaciones entre distintas páginas Web visitadas. Más adelante se describen algunos trabajos de aplicación de minería de datos en e-learning que utilizan esta técnica **[Romero, Ventura, de Castro & García, 2005]**.

1.7.2. Reglas de asociación secuenciales.

Esta técnica se utiliza con gran frecuencia para determinar patrones secuenciales en los datos, son un caso especial dentro de las reglas de asociación. Esta técnica es de gran ayuda para encontrar patrones de navegación, ya que estos patrones se basan principalmente en secuencias temporales de las acciones y difieren de las reglas de asociación en que las relaciones entre los datos se basan en el tiempo.

1.7.3. Análisis de secuencias.

El análisis de secuencias o secuencia de patrones es una técnica de minería de datos que descubre secuencias dentro de un conjunto de datos. Su aplicación a sistemas de e-learning permite analizar secuencias de páginas visitadas durante una sesión o en distintas sesiones de un mismo usuario. **[Romero, Ventura, de Castro & García, 2005]**.

La minería de secuencias permite descubrir el tiempo de las secuencias ordenadas de Urls que han seguido los usuarios y predecir los futuros. En general, en las bases de datos de transacciones están disponibles los datos en un período de tiempo y se dispone de la fecha en

que se realizó la transacción. El descubrimiento de patrones de secuencia (sequential patterns) en el Log puede utilizarse para predecir las futuras visitas y así poder organizar mejor los accesos y publicidades para determinados períodos de tiempo [Fuentes, 2007].

La aplicación de técnicas de e-learning, dentro del ámbito de la educación es de mucha importancia, específicamente las que tienen que ver con la navegación, para la mejora constante en el aprendizaje del estudiante; ya que nos permite mediante la aplicación de estas técnicas, a descubrir patrones de comportamiento en la navegación que estén relacionados especialmente con la planificación temporal y la gestión de contenidos, a las necesidades del estudiante, así como a cada una de sus preferencias particulares en el momento en que los estudiantes interactúan con el entorno virtual, y a partir de ello, se pueda determinar cómo aprende o capta el estudiante para su aprendizaje. Por lo tanto estos métodos y técnicas nos sirven para determinar y establecer las diversas formas de navegación para a partir de ello, poder después analizar los resultados obtenidos y de esta forma poder guiarlos y orientarlos de una mejor manera por el entorno virtual, para que la adaptación del aprendizaje sea adecuada y ofrezca a los estudiantes la posibilidad de mejorar su rendimiento y nivel de satisfacción, y además facilitarles la posibilidad de interactuar de una mejor manera y con resultados más satisfactorios para su aprendizaje personalizado según sus intereses y necesidades.

El objetivo del presente estudio está enfocado hacia los estudiantes, ya que se pretende realizar un nuevo modelo de datos del estudiante con el tratado previo del comportamiento del estudiante con patrones de navegación, para la mejora continua del aprendizaje mediante técnicas de e-Learning; por lo tanto para la realización del análisis nos enfocaremos sólo en las dos técnicas que corresponden a patrones de navegación: descubrimiento de reglas de asociación, y análisis de secuencias, que es a la cual está dirigida la presente investigación.

1.7.4. Clasificación y Clustering.

Los algoritmos de agrupamiento buscan grupos de instancias con características similares, según un criterio de comparación entre valores de atributos de las instancias definidas en los algoritmos. Estas técnicas de clasificación permiten desarrollar un perfil para los ítems pertenecientes a un grupo particular de acuerdo con sus atributos comunes. Este perfil luego puede ser utilizado para nuevos ítems que se agreguen en la base de datos.

En el contexto de la minería Web, las técnicas de clasificación permiten desarrollar un perfil para usuarios que acceden a páginas o archivos particulares. Esta información puede ser obtenida analizando los requerimientos de los usuarios y la información transmitida de los navegadores

incluyendo el URL. La utilización de la técnica de agrupamiento sobre los Log del EVA, puede ser utilizada para estrategias de marketing dirigido según las clases obtenidas.

1.7.5. Árboles de decisión.

Los árboles de decisión está incluido dentro de la metodología del aprendizaje supervisado, los cuales se asemejan a la teoría de *divide y vencerás* para partir el espacio del problema de subconjuntos. Los árboles de clasificación, son árboles con un conjunto o serie de condiciones organizadas de forma ordenada y con un orden jerárquico en forma de árbol. Los árboles son uno de los métodos de aprendizaje inductivo supervisado no paramétrico bastante utilizados, son sencillos y útiles para encontrar estructuras en espacios de alta dimensionalidad y además pueden ser utilizados en diferentes áreas. El análisis de clasificación basados en árboles de decisión que realizan los sistemas son técnicas de explotación de datos (***data mining***) que consisten en estudiar grandes masas de datos con el fin de descubrir patrones no triviales.

Los árboles de decisión se dividen en dos grupos: ***Árboles de clasificación*** que permiten predecir variables categóricas en un conjunto de datos, y los ***Árboles de regresión*** que permiten predecir variables continuas dentro de un conjunto de datos.

1.8. Tipos de algoritmos de minería de datos.

En base a las tareas de DM antes identificadas se presentan los siguientes tipos de algoritmos según [TechNet, 2011]:

- *Algoritmos de clasificación*, que predicen una o más variables discretas, basándose en otros atributos del conjunto de datos. Un ejemplo de algoritmo de clasificación es el Algoritmo de árboles de decisión de Microsoft.
- *Algoritmos de regresión*, que predicen una o más variables continuas, como las pérdidas o los beneficios, basándose en otros atributos del conjunto de datos. Un ejemplo de algoritmo de regresión es el Algoritmo de serie temporal de Microsoft.
- *Algoritmos de segmentación*, que dividen los datos en grupos, o clústeres, de elementos que tienen propiedades similares. Un ejemplo de algoritmo de segmentación es el Algoritmo de clústeres de Microsoft.

- *Algoritmos de asociación*, que buscan correlaciones entre diferentes atributos de un conjunto de datos. La aplicación más común de esta clase de algoritmo es la creación de reglas de asociación, que pueden utilizarse en un análisis de la cesta de compra. Un ejemplo de algoritmo de asociación es el Algoritmo de asociación de Microsoft.
- *Algoritmos de análisis de secuencias*, que resumen secuencias o episodios frecuentes en los datos, como un flujo de rutas Web. Un ejemplo de algoritmo de análisis de secuencias es el Algoritmo de clústeres de secuencia de Microsoft.

1.9. Descripción de los algoritmos para la minería.

Los algoritmos que se consideraron convenientes aplicar para el minado de los datos en la herramienta de Weka se detallan a continuación en una tabla con los datos descriptivos de cada una de las herramientas de minado junto con el tipo de técnica y descripción de las mismas:

Associate: Algoritmos de búsqueda de reglas de asociación.

1.9.1. GSP (Generalized Sequential Patterns).

El algoritmo GSP (**Generalized Sequential Patterns**) consiste en la resolución de problemas de secuencia, ya que se basan principalmente en el a priori (nivel-wise) algoritmo. Una forma de utilizar el paradigma de nivel prudente es descubrir primero todos los elementos frecuentes de una manera inteligente de nivel. Significa, simplemente, contar las ocurrencias de todos los elementos Singleton en la base de datos. Entonces, las transacciones son filtradas por la eliminación de las partidas no frecuentes. Al final de este paso, cada transacción se compone de sólo los elementos frecuentes que originalmente contenía. Esta base de datos modificados se convierte en una entrada para el algoritmo del SGP. Este proceso requiere de un paso sobre el conjunto de bases de datos .

Una clase que implementa un algoritmo del **SGP** para el descubrimiento de patrones en un conjunto de datos secuenciales. El atributo de identificación de las secuencias de datos distintas en el *set* puede ser determinada por la opción correspondiente. Por otra parte, el conjunto de resultados de salida puede ser restringido mediante la especificación de uno o más atributos que tienen que estar contenidos en cada conjunto de elementos de una secuencia.

1.9.2. Algoritmo A-priori.

A priori es un algoritmo de aprendizaje de reglas de asociación muy simple y popular, que permite identificar las posibles correlaciones o interdependencias entre distintas acciones o sucesos; pudiendo reconocer cómo la ocurrencia de un suceso o acción puede inducir o generar la aparición de otros. Las reglas de asociación son una manera de expresar patrones de datos de una base de datos. Estos patrones pueden servir para conocer el comportamiento general del problema que genera la base de datos, y de esta manera, disponer de información que pueda asistir en la toma de decisiones. Una regla de asociación es una proposición probabilística sobre la ocurrencia de ciertos estados en una base de datos. A diferencia de las reglas de clasificación que puede aparecer cualquier atributo, y además puede aparecer más de un atributo.

Dentro de las reglas de asociación existen parámetros para evaluar la calidad de las reglas, las cuales son: el soporte (cobertura) y la confianza. El soporte o cobertura de una regla se entiende como el número de las instancias que predice correctamente la regla. Y la confianza de una regla es la que mide el porcentaje de veces que dicha regla se cumple cuando se puede aplicar. “El algoritmo más simple y popular dentro del aprendizaje de las reglas de asociación es el algoritmo *A priori*, ya que el funcionamiento de este algoritmo se centra en la búsqueda dentro de los datos de los conjuntos de ítems con una cobertura determinada, y sobre esos datos lo que busca es reducir el número de conjuntos considerados, filtrando sólo aquellos ítems con mayor recurrencia. Al momento de definir la cobertura mínima, se generan todos los conjuntos que cumplen con la condición de tener un soporte menor o igual a la cobertura mínima, construyendo los conjuntos formados por sólo un ítem que superan la cobertura mínima; este conjunto de conjuntos se utiliza para construir el conjunto de conjuntos de dos ítems, hasta llegar a un tamaño en el cual no existan conjuntos de ítems con la cobertura requerida. Para la búsqueda de los conjuntos de ítems se emplea un conjunto de ítems formado por X ítems frecuentes; si y sólo si cada uno de los X ítems es frecuente por sí sólo, permitiendo generar los conjuntos de ítems frecuentes de una manera incremental”. [Castaño A., 2009].

En **WEKA** el algoritmo obtiene de diferente manera el conjunto de ítems, ya que este no genera de una sola vez todos los conjuntos de ítems frecuentes dentro de los cuales se cumpla con el valor de umbral establecido, sino que el algoritmo A priori primero va iterando los datos y cada vez se obtienen los conjuntos de ítems de un tamaño determinado, y con estos conjuntos va generando las reglas. Además, para mejorar la eficiencia del algoritmo en la búsqueda de los conjuntos de ítems frecuentes, elimina los atributos que tengan sus valores desconocidos en todos los ejemplos. Por otra parte, el algoritmo A priori de Weka permite seleccionar las reglas atendiendo a diferentes métricas no únicamente al valor de la confianza.

Cluster: Algoritmos de agrupamiento.

1.9.3. Algoritmo Simple KMeans.

Los datos del clúster utilizados en el algoritmo de k medias puede utilizar la distancia euclídea (por defecto) o la distancia de Manhattan. Si la distancia Manhattan se usa, entonces los centroides se calculan como la media de componentes más racional que encuentre.

1.9.4. Algoritmo EM.

EM pertenece a una familia de modelos que se conocen como *Finite Mixture Models*¹, los cuales se pueden utilizar para segmentar conjuntos de datos. Es un método de clustering probabilístico que trata de obtener la FDP (Función de Densidad de Probabilidad) desconocida a la que pertenecen el conjunto completo de datos.

El algoritmo EM asigna a cada instancia una distribución de probabilidad de pertenencia a cada clúster. El algoritmo puede decidir cuántos clústeres crear basado en validación cruzada o se le puede especificar A-priori cuantos debe generar. Utiliza el modelo Gaussiano finito de mezclas, asumiendo que todos los atributos son variables aleatorias independientes. Este algoritmo es bastante más elaborado que el *K-Medias*, ya que requiere muchas más operaciones. Este algoritmo permite buscar el número de grupo más apropiado. **[García M. & Álvarez A., s. a]**

Classify: Facilidades para aplicar esquemas de clasificación, entrenar modelos y evaluar su precisión

1.9.5. Algoritmo J-48.

El algoritmo J48 implementado en Weka es una versión del clásico algoritmo de árboles de decisión C4.5 propuesto por Quilan². Los árboles de decisión entran dentro de los métodos de clasificación supervisada, es decir, se tiene una variable dependiente o clase, y el objetivo del clasificador es determinar el valor de dicha clase para casos nuevos.

El proceso de construcción del árbol comienza por el nodo raíz, el que tiene asociados todos los ejemplos o casos de entrenamiento. Lo primero es seleccionar la variable o atributo a partir de la cual se va a dividir la muestra de entrenamiento original (nodo raíz), buscando que en los subconjuntos generados haya una mínima variabilidad respecto a la clase. Este proceso es

¹ Modelos de mezcla finitos.

² Quinlan J. C4.5: Programs for machine learning, Morgan Kaufmann Pub., 1993 (ISBN: 1558602380)

recursivo, es decir, una vez que se haya determinado la variable con la que se obtiene la mayor homogeneidad respecto a la clase en los nodos hijos, se vuelve a realizar el análisis para cada uno de los nodos hijos. Aunque en el límite este proceso se detendría cuando todos los nodos hojas contuvieran casos de una misma clase, no siempre se desea llegar a este extremo, para lo cual se implementan métodos de pre-poda y post-poda de los árboles.

El algoritmo *J-48* amplía las funcionalidades del C4.5, tales como permitir la realización del proceso de post-poda del árbol mediante un método basado en la reducción del error (reducedErrorPruning) o que las divisiones sobre las variables discretas sean siempre binarias (binarySplits) 4,5. Algunas propiedades concretas de la implementación son las siguientes:

- Admite atributos simbólicos y numéricos, aunque la clase debe ser simbólica
- Se permiten ejemplos con valores desconocidos
- El criterio de división está basado en la entropía y la ganancia de información

2. Herramientas para el análisis de Logs.

Con el crecimiento explosivo de las fuentes de información disponibles en Internet, se hace cada vez más necesario que los investigadores utilicen herramientas automatizadas para el hallazgo de los recursos deseados de la información, y así poder rastrear y analizar sus patrones de uso.

Para realizar el proceso de extraer conocimiento del contenido de documentos y de sus descripciones o lo que también se conoce como explotación minera para extraer patrones interesantes en registros de acceso de Internet existen muchas herramientas.

Según el estudio realizado por [Sánchez, 2008] Estas herramientas son sistemas inteligentes que trabajan tanto del lado del servidor, como del lado del cliente, para poder minar toda la información que se genera con el uso de Internet y su análisis se realiza a partir de la información que hay en los Log del servidor de Internet y el servidor de correo:

2.1. Logs.

La entidad que contiene los Logs son una grabación de las acciones actividad y eventos dentro del servidor Web a lo largo de un periodo de tiempo determinado. Los Logs van guardando información acerca de cada una de las visitas de los usuarios al sitio [Sánchez, 2008]:

- A qué hora un usuario navega por el sitio,
- La dirección IP desde la máquina que ingreso al sistema,

- Qué acciones realiza y qué archivos entra a visitar dentro del sitio,
- cuánto dura la visita y cuánto tiempo pasa en cada uno de los eventos dentro del sitio,
- de qué carrera es, qué navegador usa, y qué sistema operativo,
- El número de veces que accede al sistema,
- El número de veces que realiza determinada acción, etc.

La información de acciones y eventos realizados por el usuario presentada por las herramientas de la Web Mining, y en especial el sistema inteligente de los ficheros Log, analizan y procesan estos Logs para producir información significativa, clara y detallada; y es por tal motivo, que se decidió trabajar con esta herramienta para el descubrimiento de patrones. Parte de la información presentada es; la navegación del estudiante durante la interacción con el EVA, y los patrones que tiene cuando realiza una petición en el sistema. Debido a que los contenidos de Internet consisten en varios tipos de datos, como texto, imagen, video, metadatos o hiperligas, investigaciones recientes usan el término multimedia data mining (minería de datos multimedia) como una instancia del Web mining para tratar ese tipo de datos.

El detalle de las acciones que realiza el usuario cuando interactúa con el EVA, que se encuentran dentro de los ficheros Log; son necesarios para la inferencia de patrones de navegación, ya que estas acciones del estudiante nos ayudan a determinar la manera de navegación del estudiante y de cómo los estudiantes aprenden y realizan sus tareas dentro del EVA, cuando navegan por el mismo; y de esta manera poder descubrir y predecir los patrones de navegación que tiene el estudiante cuando interactúa con el EVA de la manera más acertada.

La información almacenada dentro de los archivos Logs, también puede ser procesada por programas de estadísticas como son: awstats, webtrends o clicktracks, que, a pesar de ser limitados en cuanto al proceso de la información, nos ofrecen pistas para la mejora del entorno, ya que devuelven información estructurada y significativa acerca de la navegación, por ejemplo:

- Cantidad de visitas por hora, por día, por mes, etc.,
- horas pico y horas de baja audiencia,
- acciones hechas con mayor frecuencia,
- acciones de entrada y salida más frecuentes dentro del entorno virtual,
- uso del buscador, ranking de palabras clave usadas para llegar,
- herramientas utilizadas durante la navega del usuario,
- servicios visitados por el usuario,
- hora de entrada y salida dentro del sistema, etc.

2.2. Herramientas incorporadas al servidor.

Estas son aplicaciones del lado del servidor, que corresponden a programas que procesan en tiempo real los datos que van almacenando en los archivos Log. Corren en el servidor, y el acceso a la información del tráfico, tanto estadística, como gráfica es mediante una interface en línea. Generalmente, este tipo de soluciones vienen incluidas en las ofertas de alojamiento Web, ya sea un servidor dedicado o compartido. **[Sánchez, 2008]**.

2.3. Herramientas en máquinas personales.

Son un software que se instala de manera independiente en máquinas de escritorio y su objetivo es hacer de igual manera análisis de Logs, pero no en tiempo real. Esta opción consiste en la descarga de los Log, para el posterior procesamiento de los mismos, por lo tanto, es necesario tener acceso a estos registros, cuestión que debe consultarse con el proveedor de alojamiento Web. Luego, mediante uno de este software especializado que se utiliza en una típica PC de escritorio, y sin requerir acceso a Internet, se desarrollan informes sobre estadísticas a fondo, en poco tiempo. Este es uno de los modos más atractivos y productivos de análisis para investigadores **[Sánchez, 2008]**.

Para el análisis de los datos dentro de la presente tesis se consideraran dos de las herramientas de trabajo, las cuales son los Logs que son las acciones que realizan los estudiantes dentro de la plataforma, y además la herramienta en máquinas personales que será la herramienta de Weka, la cual permitirá realizar la minería en los archivos Log; los cuales serán extraídos de la base de datos del EVA mediante consultas en lenguaje SQL, y de esta manera poder posteriormente realizar un análisis de las técnicas aplicados para descubrir con ellas los patrones de navegación en los estudiantes, cuando interactúan con las herramientas del EVA.

3. Sistema hipermedia adaptativos

*“Los sistemas Web adaptativos presentan información al usuario en documentos hipermedia generados dinámicamente de acuerdo con un modelo automáticamente actualizado del usuario” **[Castells, 2002]**.*

*“Los sistemas adaptativos hipermediales elaboran un modelo de metas, preferencias y conocimiento de cada sujeto individual y utilizan este modelo en la interacción con el usuario con la finalidad de adaptar el hipertexto a las necesidades del sujeto” **[Brusilovsky, 1996]**.*

Son considerados tres criterios los cuales son fundamentales que deben satisfacer un sistema hipermedia adaptativo: **[Brusilovsky, 1996]**

1. Ser un sistema hipertexto o hipermedia.
2. Tener un modelo de usuario.
3. Ser capaz de adaptar el hipermedia usando este modelo.

3.1. Técnicas de adaptación de la navegación.

Las técnicas de adaptación del soporte de la navegación consisten en la modificación de las posibilidades de navegación, con lo que se intenta reducir las posibilidades de pérdida del estudiante en el hiperespacio: a) recomendación de enlaces a seguir; b) etiquetado de enlaces. (Ej., el sistema ELM-ART, etiqueta pero permite atravesarlos, dejándole esa responsabilidad al usuario); c) ocultamiento de enlaces; d) utilización de herramientas adicionales para la navegación [Brusilovsky, 1998] y [De Bra, 1997].

- a. **Orientación directa:** el enlace más adecuado será destacado visualmente en la Pág. o un enlace tipo "next" es generado vinculando la página más adecuada al perfil del usuario.
- b. **Clasificación de enlaces:** crea una lista de enlaces del más relevante al menos relevante para el usuario en aquel momento. [Hohl, 1996].
- c. **Ocultamiento de enlaces:** la idea aquí es restringir el espacio de navegación del usuario.
- d. **Anotación adaptativa de enlaces:** incrementa los enlaces con algún comentario para informar al usuario sobre el actual estado de los nodos.

3.2. Soporte a la navegación adaptativa.

La importancia del soporte a la navegación en un sistema, es que permiten ayudar a los usuarios cuando navegan, adecuando el contenido y enlaces a las necesidades específicas de cada uno de los usuarios al momento de la interacción con el sistema, el soporte a la navegación emplea un modelo del usuario para proveer adaptación automática. Además permite realizar acciones como: agregar, cambiar, elimina, ordenar o anotar enlaces y/o destinos a los que están dirigidos que permiten realizar cambios en el formato y en los estilos de presentación.

3.2.1. Métodos

Uno de los principales objetivos que tienen los métodos de soporte a la navegación adaptativa es que se encargan de ayudar a los usuarios cuando navegan en el sistema, guiándolos, o

proporcionándoles una orientación para facilitarles la posibilidad de crear vistas personalizadas según sus intereses o necesidades. Se dividen en **[Berlanga, 2006]**:

- a. **Guía global.** Asistir a los usuarios cuando tienen un objetivo de información que se encuentra en uno o varios nodos de la Web y necesitan navegar para encontrar información.
- b. **Guía local.** Sugerir los enlaces más relevantes a seguir desde el nodo actual.
- c. **Orientación global.** Mostrar al usuario su localización en la estructura hipermedial y su posición absoluta en él.
- d. **Orientación local.** Indicar al usuario qué está alrededor y cuál es su posición relativa en la estructura hipermedial, y ayudarlo a entender cuáles son los enlaces próximos para que decida correctamente qué camino seguir.
- e. **Gestión de vistas personalizadas.** Organizar un espacio de trabajo electrónico para los usuarios que necesitan acceso a una parte más pequeña de la estructura hipermedial.

3.2.2. Técnicas

La información generada por las técnicas empleadas para la implementación de los métodos del soporte a la navegación adaptativa, presentan datos importante y apropiados para cada sujeto que manipula los enlaces que contienen cada una de las páginas. Y para ello se lo puede realizar de diferentes maneras. Y a continuación se detallan los propuestos por **[Brusilovsky, 1996; 2001 & Koch, 2000]**:

- a. **Guía directa.** El sistema decide cuál es el mejor enlace que el usuario puede seguir, partiendo de las características objetivos o necesidades del usuario para luego presentarlo como una única opción de navegación en forma de enlace o botón. Al ser la tecnología más simple restringe por completo el paso a la navegación libre, por lo cual se vuelve incapaz de orientar a aquellos que no siguen los enlaces propuestos.
- b. **Ordenación adaptativa de enlaces.** Esta técnica primeramente ordena los enlaces que contiene una página según la relevancia que tienen para cada usuario en particular, considerando para ello el modelo del usuario y algunas características de valor, Pero no asegura una navegación precisa y con éxito ya que puede desorientar al usuario, debido

a que las características adaptativas del sistema, cada vez que el usuario acceda a una misma página los enlaces que contiene el sistema podrían aparecer ordenados de diferente manera.

- c. **Ocultamiento adaptativo de enlaces.** Esconde, deshabilita o elimina los enlaces que no son relevantes para el usuario. La diferencia entre esconder, deshabilitar o eliminar un enlace consiste en que en el primer caso se mantiene el texto del enlace, en el segundo se elimina su funcionalidad pero se presenta como tal [Carro, 2001], y en el tercero se remueve para que no esté disponible y se le reemplaza.
- d. **Anotación adaptativa de enlaces.** Marca o señala los enlaces para darle al usuario claves sobre su contenido, estado actual o relevancia. Estas claves pueden mostrarse en forma de texto, iconos, colores, viñetas, etc. Este tipo de adaptación puede ser dinámica, es decir, tomar en cuenta el estado del modelo del usuario, o ser independiente del usuario y, por tanto, estática. Algunos ejemplos de esta técnica son la anotación adaptativa booleana (visitado/no visitado; recomendado/no recomendado) o la metáfora del semáforo en donde cada color (rojo, amarillo, verde) representa el grado de conveniencia o pertinencia que tiene el enlace para el usuario.
- e. **Generación adaptativa de enlaces.** Los objetivos de esta tecnología de adaptación de alto nivel, son: el descubrimiento de nuevos enlaces que sean útiles entre documentos y agregarlos al conjunto de enlaces ya existentes, generar enlaces basados en la similitud entre elementos, y recomendar enlaces que sean relevantes.
- f. **Mapas adaptativos.** La estructura de mapas hipermediales como son las estructuras gráficas de navegación que se puedan modificar según el modelo del usuario.
- g. **Navegación pasiva.** Para ofrecer ayuda al usuario, se agregan enlaces los mismo que no disponen de ningún texto de enlace, cuando se detecta un patrón de conducta, ya sea la inactividad del usuario durante un determinado periodo de tiempo, o cuando éste navega para adelante y para atrás repetidas veces [Koch, 2000].

3.3. Identificación de patrones secuenciales.

Las actividades realizadas por los usuarios que acceden a las diferentes páginas del EVA son almacenados en una tabla dentro de la base de datos del sistema, la tabla de Log (*mdl_log*), esta tabla contiene información en orden cronológico de todos los accesos y actividades/eventos

realizados por los usuarios dentro del EVA, y la interacción con cada una de las herramientas que presta la plataforma, los datos almacenados dentro de esta tabla son de principal ayuda para cumplir con los objetivos propuestos en este estudio, datos como el curso (id, nombre), el usuario (id), las herramientas (nombre), y la URL a la que accedió el estudiante.

Como ya se mencionó en la sección inicial del Estado del arte, existen varias técnicas con algoritmos no supervisados que permiten la minería para el descubrimiento de los datos en grandes repositorios, Algoritmos que permitan el de descubrimiento de patrones secuenciales y a partir de estas técnicas y de los intereses y preferencias de los usuarios el descubrimiento de los patrones de navegación, y de clasificación de páginas según sus características a partir de los datos de entrada que permitan descubrir relaciones entre ellos.

Por esta razón entre otras, es por la cual se considera conveniente la implementación de minería de datos de patrones secuenciales y la aplicación de varios tipos de técnicas de minería en el presente estudio, conseguir a partir de los datos de entrada de los Log, las rutas de navegación de los estudiantes a través de la plataforma, y con ello tratar de inferir los patrones de navegación de los estudiantes y a partir de ello saber cuál es la información necesaria y de utilidad para el estudiante en la plataforma y poder con ello mejorar los servicios prestados por la plataforma para el aprendizaje del estudiante y la comunicación entre profesor y estudiantes.

Para conseguir los objetivos propuestos en la presente tesis, se pretende mediante las técnicas de minería de datos agrupar páginas Web de la plataforma a las que el estudiante interactúa con mayor frecuencia y que tienen entre ellas características similares, asociarlas para que a partir de estos datos descubrir los patrones de navegación existentes dentro de los datos obtenidos y con ello, determinar los recursos de mayor utilidad para el usuario.

3.4. Patrones de navegación

La navegación no tiene una definición específica y clara de lo que es en sí, pero con el paso del tiempo y a partir de la aparición de las aplicaciones hipertexto hasta los sistemas de software e ingeniería Web, esta ha ido adquiriendo un interés especial, sobre todo en los temas de enseñanza virtual.

A continuación se describe algunas de las definiciones de navegación, para a partir de ello poder adquirir una idea principal de lo que es la navegación:

La navegación se define como un subconjunto de relaciones semánticas establecidas entre clases. Permite ir de un objeto a otro relacionado semánticamente [Molina, 2003].

La navegación es la sensación que el usuario tiene cuando navega hacia un objeto dentro del dominio de la aplicación [Schwabe & Rossi 1998].

La forma en que se alcanza un nodo mediante un enlace es navegación. La navegación es la más importante característica de la hipermmedia [Suh & Lee 2001].

La navegación es el proceso cognitivo de adquirir conocimiento sobre un espacio, estrategias para moverse a través del espacio, y cambiar el meta-conocimiento del espacio [Schwabe & Rossi, 1998].

La navegación expresa cómo las páginas y las unidades con contenido son linkados para formar hiper-espacios [Ceri, 2000].

A partir de estas definiciones detalladas, podemos decir que la navegación es la forma en la que los sistemas manejan y gestionan la información con una infinidad de rutas a diferentes destinos, los cuales son representados de una manera apropiada durante la iteración del usuario por el sistema, además de que debe ser la característica del software que permita estructurar la manera en cómo se desea publicar la información al usuario y poner a su disposición la funcionalidad de una forma adecuada a sus necesidades, para que el usuario a su vez sea capaz de navegar y conseguir la información que necesita y adquirir el conocimiento sobre el espacio dentro del entorno en el que navega.

La importancia de estudiar los patrones de navegación del estudiante, cuando éste interactúa con el EVA, es para determinar y analizar las diferentes técnicas de aprendizaje del estudiante, y a partir de ello, tal vez en trabajos futuros se pueda mejorar varios aspectos del diseño y la usabilidad de las herramientas de aprendizaje y los recursos al momento de la iteración del estudiante con el EVA, con el fin de que el estudiante tenga una experiencia de uso más satisfactoria para su aprendizaje virtual. Es por eso que se desea conseguir, mediante el análisis de comportamiento de los usuarios con el entorno virtual, patrones que nos ayuden a identificar como es el proceso de aprendizaje cuando el estudiante navega por dicho entorno.

Para el descubrimiento de los patrones de navegación mediante el análisis de los mismos, se tiene los ficheros Log, como datos de entrada, los cuales se encuentran almacenados en una tabla de la base de datos del EVA (*mdl_log*), se lo realizará con la ayuda de la técnica de identificación de patrones secuenciales y la implementación de tres de las técnicas de minería de datos (Algoritmos supervisados) y una técnica de minería de Algoritmos no supervisados; todas estos algoritmos de minería se encuentran dentro del conjunto de técnicas de la herramienta de

Weka, los mismos que son las Reglas de asociación con los algoritmos (**GSP y A-priori**); la técnica de Clustering 'Segmentación' con los algoritmos (**SimpleKMeans (clustering numérico) y EM (clustering probabilístico)**) y la técnica de Árboles de decisión con el algoritmo (**J-48**); con la finalidad de descubrir patrones o grupos de estudiantes con similares características, intereses en común o conducta de interacción durante su navegación en el EVA.

Las diferentes técnicas empleadas nos permitirán determinar los algoritmos más apropiados para este análisis, tanto para la búsqueda de patrones de navegación como la identificación de indicadores en el comportamiento del estudiante cuando interactúa con el sistema.

4. Soporte adaptativo al Aprendizaje.

Un Sistema Hipermedia Adaptativo (SHA) se define como:

Un sistema basado en hipertexto e hipermedia que refleja algunas características del usuario en el modelo de usuario y aplica este modelo para adaptar varios aspectos visibles del sistema al usuario [Brusilovsky, 1996].

4.1. Componentes del soporte adaptativo.

Los sistemas Hipermedia Adaptativos utilizan tres componentes básicos para lograr la adaptabilidad. Esta división proporciona claridad al desarrollo de aplicaciones adaptativa y permite asignar responsabilidades específicas a cada modelo.

Modelo del Usuario.- Este modelo describe la información del usuario que el **SHA** guarda en un registro permanente. Esta información incluye una representación del estado del conocimiento adquirido por el estudiante y un registro de los nodos que ha visitado.

Modelo del Dominio.- Este modelo describe cómo se enlaza y estructura la información.

Modelo del Profesor.- Este modelo está compuesto por reglas pedagógicas que definen cómo se combinan el Modelo del Dominio y el Modelo del Estudiante para proveer de adaptación al sistema [Arteaga C.; Fabregat R.; Mérida D., s. a].

Se utiliza con gran frecuencia el soporte adaptativo al aprendizaje colaborativo, ya que este es más nuevo que el soporte adaptativo al aprendizaje individual, y su objetivo principal es fomentar los procesos colaborativos.

4.2. Objetivos del soporte adaptativo.

Los objetivos de la adaptación del soporte adaptativo a la colaboración son: **[Gaudioso, 2002]**

- Formar grupos de trabajo cuyo objetivo es facilitar la cooperación entre alumnos con perfiles similares.
- Dar soporte en el uso de servicios de comunicación para fomentar su uso en el grupo.
- Dar soporte en el uso de servicios para compartir información y para aconsejar al usuario que documentos pueden ser de su interés, bien porque el usuario que los mandó era afín o bien porque la temática puede ser de su interés.
- Encontrar un usuario afín para colaborar. Seleccionar entre todos los usuarios del grupo, aquellos con los que el sistema piensa que la colaboración puede ser efectiva.
- Detectar perfiles de colaboración. Identificar estereotipos que ayuden a gestionar los grupos de trabajo y las relaciones personales.

Por otra parte, se enumeran las siguientes tareas de soporte adaptativo a la colaboración:

- *Aconsejar a un usuario que interactúa más con un determinado servicio.* Para esto se necesita conocer el nivel de actividad de cada estudiante. Es especialmente útil para aquellos sistemas en los que las actividades se gestionan a través de algún servicio de la plataforma que da soporte al curso.
- *Avisar al tutor de la posibilidad de que un alumno tenga dificultades en el uso de algún servicio.* Para ello necesitamos conocer el nivel de fracaso de un usuario en una determinada actividad.
- *Agrupación automática de usuarios dentro de una misma comunidad.* Esta actividad tiene que ver con la formación de grupos de acuerdo a ciertas condiciones pre-establecidas. El sistema sugiere u organiza los grupos de acuerdo a similitudes y/o diferencias entre los estudiantes.
- *Aconsejar al tutor una clasificación de mensajes en el foro.* Esta tarea puede ser de gran utilidad para el tutor a la hora de administrar un determinado grupo, identificando temas de interés que sirvan para añadir categorías al foro. Consideran que ésta es una tarea de soporte adaptativo a la colaboración, puesto que una mejor gestión de los servicios de comunicación mejora sin duda, el acceso a la información y por tanto la colaboración.

5. Modelo de Estudiante Abierto.

Un modelo de estudiantes abierto (*Open Learner Model, OLM*) es un modelo de estudiante al que el usuario puede acceder. Algunos de los propósitos que se pueden tener para abrir el modelo de estudiante son entre otros: mejorar la exactitud del modelo del estudiante; promover la reflexión del estudiante, ayudar a los estudiantes a planear y hacer seguimiento de su aprendizaje; y permitir a los estudiantes un mayor control sobre su aprendizaje.

Los modelos de usuario no son conocidos por los usuarios de un sistema adaptativo. Esto es válido en escenarios en los cuales el único interés del usuario es satisfacer una necesidad u obtener una información de forma ágil y oportuna en escenarios donde una búsqueda manual puede ser impensable. Estas tareas son generalmente realizadas con objetivos específicos a corto plazo que generan poco o ningún esfuerzo mental por parte de quien interactúa con el sistema adaptativo.

Otro escenario con elementos comunes, como lo es la sobrecarga de información y de fuentes de consulta, pero con características bastante distintas son los entornos de aprendizaje virtual adaptativos. Estos últimos requieren que quien los utilice desarrollen un gran esfuerzo mental en pro de lograr resultados realmente significativos que vayan más allá de la simple obtención de una certificación o de cualquier otro tipo de recompensa diferente al aprendizaje. Los entornos educativos virtuales requieren de un gran compromiso, disciplina y dedicación por parte de quien los usa, aún más si el entorno carece en su totalidad de tutores que guíen dicho proceso [Vélez Ramos J., 2009].

Los propósitos para abrir un modelo de estudiante, puede llevar a la reflexión de los estudiantes sobre su nivel de competencia como elemento importante para alcanzar estados meta-cognitivos. Los estudiantes que logran activar un nivel meta-cognitivo en su proceso de aprendizaje logran resultados más significativos que aquellos que no lo hacen. Hay evidencias que sugiere que la efectividad en los procesos de aprendizaje puede ser mejorada cuando los estudiantes reflexionan acerca de su propio conocimiento. Adicionalmente, algunos trabajos sugieren el potencial que tiene un OLM para fomentar la reflexión.

5.1. Soporte adaptativo a la navegación con OLM.

El soporte adaptativo a la navegación dentro de entornos virtuales permiten encontrar las preferencias que tiene usuario y el conocimiento previo de los usuarios en las herramientas del entorno durante la interacción con la plataforma. Con esta información es posible realizar algunas

actividades de adaptación que facilitan el uso de las herramientas y los servicios por parte del estudiante y la navegación dentro de los contenidos disponibles. Como, por ejemplo, desde el nivel de actividad del estudiante en el uso de un servicio específico que le puede recomendar el para interactuar de forma más o menos precisa con este servicio, o desde la última página visitada se puede recomendar la página adecuada a visitar o acceder.

5.1.1. Actividades de adaptación.

Los *AHA* son uno de los sistemas de adaptación orientada a los sistemas educativos centrados en la adaptación de los servicios dentro del curso. Además, en entornos de colaboración de las actividades fundamentales que se deben ofrecer son las tareas que proporcionan apoyo a la colaboración de adaptación. Por el contrario, el apoyo a la navegación de adaptación se debe aplicar para ofrecer al usuario una navegación personalizada de todos los servicios disponibles en el sistema.

Los modelos de estudiante abierto con soporte adaptativo a la navegación ayudan a la mejora en el acceso de los usuarios a los servicios y herramientas que la plataforma del EVA ofrece (foros, cuestionarios, chat, fuentes de información, mensajería, recursos, exámenes, tareas, calendario, perfil del usuario, etc.,...), Así como facilitar la colaboración entre los estudiantes y tutores del mismo grupo.

Según el estudio realizado por [**Gaudioso E. & Boticario J., 1997**] las siguientes actividades de apoyo a la colaboración de adaptación son necesarias para la mejora de un modelo:

- *Predecir el nivel de actividad de un usuario específico*: esta tarea es útil para aquellos sistemas en los que todas las actividades que se realizan a través de la plataforma del curso.
- *Agrupación automática de usuarios en una misma comunidad*: aunque los grupos de la comunidad de usuarios con un objetivo común, cuando se hace el trabajo de grupo que será útil para formar algunos subgrupos teniendo en cuenta los aspectos comunes y diferenciar los estudiantes.
- *El análisis de los mensajes enviados a los foros*: esta tarea puede ser muy útil para el tutor en la administración de un grupo específico, la identificación de temas de interés que se puede utilizar para añadir categorías al foro.

5.1.2. Actividades de apoyo a la navegación.

Las actividades de adaptación de apoyo a la navegación se presentan a continuación:

- Proporcionar algunas páginas que describen la información del curso. El recorrido de estas páginas se pueden personalizar para cada usuario.
- Proporcionar enlaces a las fuentes de información, (direcciones Web, archivos de área de almacenamiento, mensaje del foro), que puede ser interesante (de acuerdo a los intereses del usuario, el perfil de la persona que ha proporcionado esta fuente de información, etc.).

Estas son algunas de las actividades de adaptación que permite poner de relieve el modelo del estudiante. Dentro de las actividades de adaptación propuestas por [Gaudioso E. & Boticario J., 1997], las actividades de colaboración adaptativa de apoyo se resuelven por medio de técnicas de aprendizaje automático. Además, las actividades de navegación adaptativo de apoyo se resuelven mediante técnicas basadas en el conocimiento (reglas). Por lo tanto el modelo de usuario propuesto permite combinar fácilmente los dos tipos de mecanismos de inferencia.

5.1.3. Colaboración adaptativa en las actividades de apoyo a la navegación.

La colaboración adaptativa y las actividades de apoyo a la navegación como ya se lo mencionó anteriormente, se basan en técnicas de aprendizaje automático, y por otra parte hay las actividades de navegación adaptativo de apoyo las cuales se basan en reglas, ambas permiten describir la navegación del estudiante con las diferentes herramientas y servicios disponibles dentro de la plataforma para predecir el nivel de interacción que tienen los estudiantes EVA. Para ello se describen las actividades de apoyo propuestas para predecir el nivel de interacción y colaboración que tienen los estudiantes con los servicios del sistema.

5.1.4. Adaptación en las actividades de apoyo a la colaboración.

5.1.4.1. Predicción del nivel de actividad

Para predecir los niveles de actividad del usuario con las diferentes herramientas de la plataforma del EVA y sus interacciones con cada una de ellas ayudan a determinar el objetivo de las actividades de aprendizaje, es decir, predecir el nivel interacción de cada usuario con las herramientas de la plataforma y los servicios generales e individuales.

Para aprender este atributo y como cualquier tarea de aprendizaje, es necesario determinar cuáles son serán los atributos de entrada que van a formar parte de cada entrenamiento. Mediante la extracción y preparación de datos de la base de datos del EVA, que corresponde a la tabla de los eventos y actividades realizadas por los estudiantes durante su navegación por la plataforma, la cual está compuesta por un conjunto de instancias de interacción como datos de entrada para los algoritmos de aprendizaje que se aplicaran. De este proceso se pretende obtener un conjunto de datos con diferentes casos, del cual cada uno contiene atributos con un valor de clase que indica el nivel de interacción con la actividad dentro de las herramientas o servicios del EVA dentro de los cursos.

Según [**Gaudioso E. & Boticario J., 1997**], en un entorno Web hay una gran variedad de datos de interacción y esto hace que los clasificadores individuales que no estén en algunas regiones de los datos de entrenamiento o se precisa en los demás (región de la competencia). Si tenemos clasificadores con las regiones de competencias diferentes, podemos combinarlos con el fin de mejorar el rendimiento. Esta combinación se suele denominar conjunto de clasificadores.

6. Trabajos relacionados.

La minería de datos aplicada a sistemas tradicionales de e-learning suele utilizar sólo la información proporcionada por los ficheros Log con las preferencias personales de los estudiantes. Existe una relación con los sistemas de comercio electrónico al utilizar la misma fuente de datos y técnicas de minería de datos pero adaptadas a los entornos educativos.

Según (**Romero, Ventura, de Castro & García, 2005**) la minería de datos aplicada a SHA y STI para educación basados en Web, suele disponer de mayor información e información necesaria y útil para el análisis de navegación de usuarios dentro del sistema en los ficheros Logs, la cual contiene la interacción entre el estudiante y el sistema sobre actividades realizadas, aciertos y fallos, notas y niveles de conocimiento, etc. Además la capacidad de estos sistemas para la enseñanza adaptada, hacen que la aplicación de minería de datos se diferencie de los sistemas aplicados en comercio electrónico, al estar mucho más orientados al problema concreto del proceso de aprendizaje de los estudiantes.

6.1. Reglas de asociación.

Uno de los pioneros de la utilización de técnicas de minería Web en sistemas de e-learning es [**Osmar Zaïne, s.a**] y actualmente propone utilizar agentes recomendadores para recomendar actividades de aprendizaje en línea o atajos en un curso Web basándose en los historiales de

acceso y mejorar el proceso de aprendizaje en línea. Concretamente utiliza minería de reglas de asociación para entrenar al agente recomendador y construir un modelo que representa el comportamiento de acceso o asociaciones entre actividades de aprendizaje en línea. Otro trabajo que analiza los ficheros Log de entornos de aprendizaje Web utilizando técnicas de minería de reglas de asociación y filtrado colaborativo, es el realizado por Feng-Hsu Wan, para descubrir patrones de navegación útiles y proponer un modelo de navegación. El modelo de navegación consiste en dos tipos de relaciones: relaciones de asociación y relaciones de secuencia entre documentos.

La utilización de métodos heurístico como regresión lineal en combinación con reglas de asociación es propuesta por **Freyberger** y otros para buscar buenos modelos de transferencia de aprendizaje de estudiantes a partir de los ficheros Logs de interacción de sistemas tutores inteligentes. El modelo de transferencia relaciona preguntas con los componentes del conocimiento necesarios para responderlas correctamente. Las reglas de asociación guían el proceso de búsqueda para encontrar modelos de transferencia que predicen el éxito de los estudiantes.

También se están utilizando técnicas de softcomputing, por ejemplo **[Pao-Ta Yu, s. a]** y otros proponen la utilización de reglas de asociación difusas para descubrir relaciones entre patrones de comportamiento de los estudiantes, incluyendo el tiempo de acceso, números de páginas leídas, preguntas contestadas, mensajes leídos y enviados, etc. Mediante monitorización y análisis el sistema indicará mediante umbrales de alarma los resultados a los estudiantes y al instructor. Un trabajo más orientado a los sistemas hipermedia adaptativos es el de Romero y Ventura que utilizan minería de reglas como técnica de descubrimiento de información útil para los autores de este tipo de cursos con el objetivo de poder realizar mejoras tanto del contenido, como de la estructura de los cursos y de su adaptación.

Concretamente proponen la utilización de algoritmos evolutivos multiobjetivo para el descubrimiento de relaciones importantes a partir de los datos de utilización (tiempos de acceso a páginas, aciertos y fallos, y niveles de conocimiento) de los estudiantes. Otro trabajo que también emplea algoritmos evolutivos es el realizado por **Behrouz Minaei-Bidgoli y William F. Punch** para realizan un análisis de asociación para predecir el rendimiento de los estudiantes. Utilizan clustering de recursos Web valorados y descubrimiento de reglas de asociación interesantes mediante algoritmos genéticos para optimización de minería de datos con el objetivo es clasificar a los estudiantes para predecir su clasificación final basándose en las características extraídas de los ficheros Logs.

6.2. Clustering.

Otra propuesta similar a los trabajos presentados anteriormente es la del análisis de caminos realizado por **Sung Ho Ha y otros** que utilizan el análisis de caminos transversales de páginas Web para la personalización de la enseñanza y la asociación de páginas Web para la estructuración de conocimiento. El descubrimiento de caminos se realiza desde dos puntos de vista: caminos individuales y caminos agregados que incluye el proceso de clustering utilizando mapas auto organizativo de Kohonen.

6.3. Análisis de secuencias.

El análisis de patrones de navegación en entornos de aprendizaje basado en Web es utilizado por **Karin Becker y otros** dentro de una herramienta de minería de utilización Web para el análisis de patrones y pre procesamiento de datos de utilización de entornos de aprendizaje basados en Web. Las técnicas de descubrimiento de patrones utilizadas son asociación y secuencia, y la secuencia de patrones describen accesos a páginas relaciones en un orden específico. Otra propuesta de análisis de caminos es el realizado por **Sung Ho Ha** y otros que utilizan el análisis de caminos transversales de páginas Web para la personalización de la enseñanza y la asociación de páginas Web para la estructuración de conocimiento. El descubrimiento de caminos se realiza desde dos puntos de vista: caminos individuales y caminos agregados que incluye el proceso de clustering utilizando mapas auto organizativo de Kohonen.

El análisis de los patrones de comportamiento es realizado por **[Claus Pahl, s. a]** para el análisis de la interacción en sistemas e-learning. Para ello utiliza técnicas de minería Web como clasificación, patrones secuenciales y series temporales para el análisis y la evaluación de las interacciones de los estudiantes con los contenidos de los más de e-learning. El análisis de los patrones de comportamiento permite describir los caminos de aprendizaje ideales dentro de la topología de un curso. Otro trabajo sobre el análisis y la minería de la información sobre el comportamiento del aprendizaje en entornos basados en SCORM es el realizado por Wei Wang y otros para ayudar a los profesores a comprender los motivos del buen aprendizaje o no de los alumnos. Utilizan primero técnicas de minería de secuencias de patrones, posteriormente agrupan a los estudiantes mediante clustering y finalmente generan un árbol de actividades personalizadas.

Los trabajos mencionados anteriormente aplican técnicas de minería de datos para medir el nivel de aprendizaje del estudiante en entornos virtuales mediante la aplicación de sistemas e-Learning para determinar las preferencias de los estudiantes en cuanto a las diferentes

herramientas y servicios disponibles para el uso del estudiante dentro de la plataforma de enseñanza de la UTPL (EVA).

CAPITULO II:

ANÁLISIS DE LA PLATAFORMA DEL EVA DE LA UTPL.

2. Análisis de la plataforma del EVA de la UTPL.

2.1. Introducción.

La WEB ha sido desarrollada como una herramienta para los navegadores (*browsing*) y para la recuperación de información. Sin embargo, se trata de una las herramientas educativas más utilizadas por muchas empresas e instituciones de educación.

En los sistemas basados en la Web, dedicados a la educación como lo es el Entorno Virtual de Aprendizaje (EVA), son plataformas diseñadas para la educación en el ámbito del aprendizaje social y colaborativo entre profesor y estudiantes. El EVA está diseñado bajo la plataforma Moodle implementada en la Universidad Técnica Particular de Loja, tanto para la modalidad Presencial como a Distancia, este análisis está enfocado en la modalidad a Distancia, y será el principal factor de estudio para el análisis y descubrimiento de patrones. El EVA cuenta con varias herramientas de Apoyo, Colaboración y Comunicación de las que puede disponer el profesor y estudiantes para su aprendizaje.

2.2. Moodle como sistema de enseñanza.

“Moodle es un paquete de software para la creación de cursos y sitios Web basados en Internet. Es un proyecto en desarrollo diseñado para dar soporte a un marco de educación social constructivista. Moodle se distribuye gratuitamente como Software libre (Open Source) (bajo la Licencia Pública GNU). Básicamente esto significa que Moodle tiene derechos de autor (copyright), pero que usted tiene algunas libertades. Puede copiar, usar y modificar Moodle siempre que acepte: proporcionar el código fuente a otros, no modificar o eliminar la licencia original y los derechos de autor, y aplicar esta misma licencia a cualquier trabajo derivado de él. Moodle puede funcionar en cualquier ordenador en el que pueda correr PHP, y soporta varios tipos de bases de datos (en especial MySQL).

La palabra Moodle era al principio un acrónimo de *Modular Object-Oriented Dynamic Learning Environment* (Entorno de Aprendizaje Dinámico Orientado a Objetos y Modular), lo que resulta fundamentalmente útil para programadores y teóricos de la educación. También es un verbo que describe el proceso de deambular perezosamente a través de algo, y hacer las cosas cuando se te ocurre hacerlas, una placentera chapuza que a menudo te lleva a la visión y la creatividad. Las dos acepciones se aplican a la manera en que se desarrolló Moodle y a la manera en que un estudiante o profesor podría aproximarse al estudio o enseñanza de un curso en línea.”³

³ http://docs.moodle.org/19/es/Acerca_de_Moodle

Moodle es un sistema de gestión de cursos de libre distribución, lo cual permite realizar cambios y/o modificaciones en el código de acuerdo a las necesidades que se requiera para la construcción del sistema de evaluación de aprendizaje. Al sistema de evaluación los docentes lo señalan como una plataforma que les permite mostrar al alumnado, y determinar cómo va a ser evaluadas las actividades, número de alumnos, tanto por aula como matriculados en las asignaturas; y otros elementos más que se relacionan con el sistema de evaluación y en el curso.

El diseño de expresión de Moodle hace que la creación de cursos, el complemento de contenido que se desea impartir a los estudiantes sea más fácil. Moodle está diseñado para apoyar el estilo de aprendizaje de los estudiantes, llamada (*social constructionist pedagogy*) por sus siglas en inglés; pedagogía constructivista social (**Rice, 2006**). Estos estilos de aprendizaje consideran que los estudiantes aprenden de mejor manera cuando interactúan directamente con el material de aprendizaje, la construcción de nuevos materiales para los demás, y de la interacción con otros estudiantes y tutor sobre la materia. Moodle no requiere el uso de este estilo en los cursos, pero este estilo es lo que mejor soporta.

Moodle tiene una gran variedad de actividades dentro de los módulos y recursos para crear cinco tipos de material del curso estático (*una página de texto, una página Web, un enlace a cualquier cosa en la Web, una visión en uno de los directorios del curso y una etiqueta que muestra el texto o la imagen*), seis tipos de materiales dentro del curso interactivo (*tareas, elegir, revista, lección, cuestionarios y encuestas*) y cinco tipos de actividades donde los estudiantes interactúan entre sí (*chat, foro, glosario, wiki y talleres*).

2.3. Tablas de información relacionadas al estudiante.

En esta sección se detallan las tablas que se relacionan entre sí para el diseño del escenario con las rutas de navegación del estudiante dentro de las páginas que se encuentran enlazadas entre objetos o recursos relacionados cuando este interactúa con el sistema, y a partir de ello poder descubrir la secuencia de navegación que tiene un determinado modelo de estudiante en las acciones que realiza para su aprendizaje. Estas relaciones permiten deducir los eventos o peticiones secuenciales de recursos que el usuario realizará cuando navega por la plataforma del EVA, y esta información nos expresa de manera lineal el comportamiento del usuario desde los registros de cada uno de los eventos generados en el momento que el usuario realiza una petición o solicitud al sistema.

2.3.1. Diagrama Entidad – Relación.

En el proceso de análisis se representarán las actividades que el estudiante lleva a cabo cuando interactúa con la plataforma, para determinar las tablas necesarias de Moodle para acoplarlas en el escenario, para ello, se empezará con la realización del **modelo Entidad – Relación**, el cual detalla un subconjunto de tablas y atributos de la base de datos, que se co-relacionan entre sí cuando el usuario realiza peticiones al sistema.

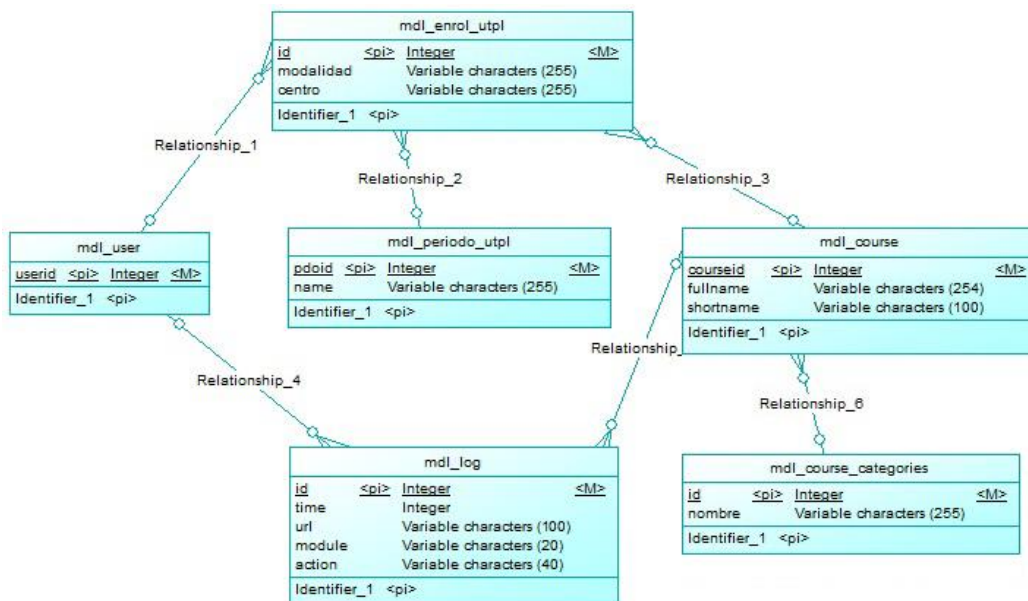


Fig. 3 Diagrama E – R de las tablas del estudiante en Moodle (EVA).

Fuente: Autora del trabajo de Tesis.

La razón por la que es necesaria la extracción de datos, es debido a la gran cantidad de información no estructurada, y la cantidad de tablas existentes en el EVA, de las cuales, no todas son útiles para el estudio de los patrones de navegación, y es por esto, que se ha elaborado el *diagrama Entidad-Relación* para determinar las tablas más importantes para el estudio y proceso de minería de datos, ver **Fig. 3**, y de ello extraer sólo la información útil, que existe en la base de datos, para un posterior análisis de los resultados.

2.3.2. Tablas del EVA para la minería de datos.

La base de datos del EVA, tiene una gran cantidad de entidades, de las cuales se realizará una breve descripción sólo de aquellas entidades de las cuales se necesitan para el desarrollo del presente trabajo. Estas entidades se las obtendrán mediante consultas en lenguaje SQL mediante la conexión directa a la base de datos del EVA, para luego realizar el pre-proceso de abstracción previo a la minería de los datos.

Tabla 2. Tablas del EVA previas al proceso de minería de datos.

Tabla	Descripción
mdl_user	Información de todos los estudiantes.
mdl_role	Información acerca de las diferentes entidades que tienen acceso al EVA.
mdl_log	Registros de acciones de cada usuario.
mdl_course	La información de cada uno de los cursos en el sistema.
mdl_course_categories	La información de las categorías de los diferentes cursos en el sistema.
mdl_centro	Información sobre los centros universitarios en las sucursales.
mdl_enrol_utpl	Información acerca de matrícula de los estudiantes en el sistema.
mdl_periodo_utpl	Información acerca del periodo académico que cursaron los estudiantes.
mdl_modules	Información acerca de todas las herramientas disponibles.

En el **Diagrama Entidad–Relación** se muestran las tablas que sirven para el proceso de minería de datos, la **Tabla 2**, muestra las entidades que permitirán obtener un seguimiento de las peticiones y acciones realizados por el estudiante, tales como: accesos a los cursos, herramientas y actividades del EVA, para identificar patrones de navegación en el usuario.

2.3.3. Atributos de las entidades en el (EVA) que se trabajaran para la minería

Las entidades y atributos que se relacionan entre sí, y que serán analizadas más adelante.

Tabla 3. Atributos de las Entidades del EVA para el proceso de minería de datos.

TABLAS DE MOODLE		DESCRIPCIÓN
TABLA	ATRIBUTOS	
mdl_log	<ul style="list-style-type: none"> ✓ userid ✓ course ✓ module ✓ action ✓ url 	Datos de los usuarios dentro del EVA, objetos en los que el usuario puede navegar durante la interacción.
mdl_enrol_utpl	<ul style="list-style-type: none"> ✓ courseid ✓ userid ✓ pdoid ✓ rol ✓ modalidad ✓ centro 	Datos de matrícula del estudiante dentro del sistema. (Modalidad, Categoría, cursos en los que está enrolado).
mdl_user	<ul style="list-style-type: none"> ✓ userid 	Datos del usuario dentro de Moodle (EVA).
mdl_centro	<ul style="list-style-type: none"> ✓ Id ✓ nombre 	Datos del centro universitario al que pertenecen los estudiantes.
mdl_course	<ul style="list-style-type: none"> ✓ courseid ✓ fullname 	Datos de los cursos que el usuario está cursando en el sistema.
mdl_role	<ul style="list-style-type: none"> ✓ id ✓ name 	Acciones que el estudiante realiza dentro de los cursos en las categorías que están matriculados.

Fuente: Ruiz, F (2014).

En la **Tabla 3**, se pueden apreciar junto a las entidades, los atributos, y las principales características de cada una de las entidades para el proceso de minería de los datos en el EVA. Estas entidades permitirán obtener mediante consultas en lenguaje SQL, toda la información necesaria de las actividades y eventos realizados por los estudiantes durante su interacción con cada una de las herramientas y servicios de los cuales dispone la plataforma del EVA.

2.3.4. Descripción de los datos de usuario en el EVA

Los datos del usuario permiten identificar al usuario en su totalidad dentro del sistema, tanto en la interacción con el sistema como en la navegación a través de cada una de las páginas que tiene el curso y herramientas por las cuales navega el estudiante y tutor en el EVA, que son los usuarios a los cuales hacemos referencia en la presente investigación.

Los datos de usuario en el EVA se presentan en la siguiente tabla.

Tabla 4. Atributos de los estudiantes de la UTPL en el EVA.

ATRIBUTO	DESCRIPCIÓN	TABLA MOODLE	
		Atributo	Tabla
Id de Usuario	Código que identifica a cada uno de los usuarios.	userid	mdl_user
Id de Curso	Código de las materias ingresadas en el sistema.	courseid	mdl_course
Nombre Curso	Nombre completo de la materia.	Fullname	mdl_course
Id de Categoría	Código de categoría al que pertenece el usuario	id	mdl_course_categories
Nombre Categoría	Nombre de la categoría a la cual pertenece el curso.	name	mdl_course_categories
Id de Periodo	Id del periodo académico que cursa el usuario.	pdoid	mdl_periodo_utpl
Nombre Periodo	Nombre del periodo académico que cursa el usuario	name	mdl_periodo_utpl
Modalidad	Modalidad de estudio que pertenece el usuario	modalidad	mdl_enrol_utpl
Id del centro	Código de identificación del centro al cual pertenece el estudiante.	id	mdl_enrol_utpl
Nombre de Centro	Nombre del centro de estudio al cual pertenece el usuario dentro del sistema.	centro	mdl_enrol_utpl
Id curso	Código único de cada una de las materias.	courseid	mdl_enrol_utpl
Nombre del curso	Nombres del curso en el que se encuentra enrolado el estudiante dentro del sistema.	name	mdl_course
rol	Rol del usuario en el sistema (estudiantes)	rol	mdl_enrol_utpl
Id log	Id del evento/actividad o requerimiento que realizó el estudiante dentro del sistema.	id	mdl_log
Fecha	Fecha de acceso del usuario al curso.	time	mdl_log
Acciones	Peticiones hechas por el usuario en el sistema.	action	mdl_log
Herramientas	Servicios a los cuales accede el usuario cuando realiza una acción dentro de un objeto	info	mdl_log
Url	Dirección de la página a la que ingresa el usuario.	url	mdl_log

Autor: Ruiz, F (2014).

Los datos de los estudiantes dentro del sistema, ver **Tabla 4**. Es la información que se tomará en cuenta para la extracción de los datos más relevantes dentro del sistema, ya que esta información permitirá identificar datos como el id de los estudiantes que se encuentran enrolados en los cursos a ser evaluados, el periodo de estudio, la carrera, la modalidad y el ciclo al cual pertenecen, para realizar el proceso de minado.

Una vez identificados los datos de los estudiantes que serán analizados, a partir de la entidad *mdl_log*, la cual reduce las acciones realizadas por el usuario cuando navega por el sistema, y de esta manera poder obtener los datos de interacción de los estudiantes con el EVA, se buscará extraer los datos a los que se les desea realizar la minería, es decir filtrar sólo aquella información que sirve dentro de la base de datos, para poder trabajar sólo con los datos que necesitamos realizar la minería.

2.3.5. Descripción de los atributos de interacción en el EVA

Para realizar la obtención de la información de la interacción de los estudiantes en las diferentes herramientas y servicios de la plataforma del EVA, primeramente se extraerá los datos de interés de los estudiantes durante su navegación en el sistema, es decir, las páginas a las que accede con mayor frecuencia para realizar actividades dentro de las herramientas con las que cuenta en los cursos en los que está matriculado, ya sean:

- Herramientas de *Colaboración y comunicación* (*chat, mensajería, foros, twitter, blog, etc.*).
- Herramientas de *Actividades y contenidos* de los objetos con los que el estudiante interactúa con sus compañeros y tutor (*foros, cuestionarios, tareas, etc.*).
- Herramientas de *Información general* (*perfil del usuario, eventos, calendario, etc.*).
- Herramientas que sirven de *Apoyo a su aprendizaje* (*recursos, REA, subir archivos, utpl, blogs, etc.*).
- a los diversos objetos y contenidos del EVA, y en si todas las acciones de interacción que el estudiante realiza en los cursos dentro del sistema.

A continuación se presenta una tabla con los datos de interacción que los estudiantes tienen con cada una de las herramientas de las cuales dispone la plataforma del EVA:

Tabla 5. Datos de las interacciones de los estudiantes en el EVA.

MOODLE		DESCRIPCIÓN
TABLAS	ATRIBUTOS	
mdl_user	firstaccess	Fecha del primer acceso del usuario a los recursos del sistema
mdl_user	lastaccess	Fecha del último acceso del usuario al curso dentro del sistema.
mdl_user	lastlogin	Fecha del último acceso del usuario al sistema
mdl_user	currentlogin	Fecha de inicio de sesión actual del usuario al sistema
mdl_log	userid, course, module, acción	Acciones que realiza el usuario en las herramientas, tales como foros, tareas, blogs, cursos, exámenes, etc., dentro del EVA.
mdl_log	userid, firstaccess, lastaccess, lastlogin, currentlogin.	Atributos que detallan el tiempo promedio que un usuario accede a los cursos que está matriculado dentro del sistema.
mdl_log	time, userid, course, action, module, count	Detalla el total de peticiones hechas a las herramientas (foros, exámenes, tareas, etc.) por cada día de ingreso al sistema.
mdl_log	time, userid, course, module, action, count	Detalle de las veces que un estudiante realiza peticiones al sistema por cada curso que está tomando en el sistema.
mdl_log	action, userid, time	Detalle del tiempo que toma el estudiante en realizar una acción, ya sea: ingresar, revisar un foro, salir, etc.
mdl_log	userid, courseid, module, action, info	Detalla las actividades de interacción del estudiante con el tutor como: foros, tareas, cuestionarios y recursos en el sistema
mdl_user	userid, firstaccess, currentlogin	Detalle de las veces que cada estudiante accede al sistema.
mdl_event	nombre, descripción, courseid, userid, modulename.	Detalle del total de veces que un estudiante realiza las tareas propuestas en el sistema, ya sea: tarea, examen, lección, etc.

Fuente: Ruiz, F (2014)

La interacción que los estudiantes tienen con la plataforma ayuda a medir el nivel de aporte que brindan las herramientas y actividades y los contenidos que tienen cada una de ellas dentro del sistema; estos datos permiten determinar patrones de navegación que los estudiantes tienen durante la interacción con el EVA y su tutor, y de esta manera, poder determinar indicadores que permitan descubrir cuáles son las rutas para satisfacer de mejor manera el aprendizaje del estudiante con el uso de las herramientas del EVA.

Las iteraciones que el estudiante realiza al momento de navegar en el EVA, y las acciones que realiza sobre los objetos del curso se detallan en la **Tabla 5.**, estos datos sirven para realizar las consultas a la base de datos y determinar la ruta de navegación que tiene el estudiante cuando este interactúa con el EVA, desde el momento que envía la solicitud de acceso al EVA, las acciones que realiza dentro del sistema, hasta que sale del mismo.

2.4. Descripción de las herramientas dentro de la plataforma

Las herramientas disponibles en el EVA con las cuales los estudiantes pueden interactuar durante su navegación en el sistema, sirven de ayuda al estudiante para su comunicación con el profesor, para ello tiene herramientas como (chat, twitter, mensajería, etc.), además cuenta con herramientas que le servirán de ayuda como recursos adicionales, ya sean consultas de otros trabajos relacionados a las tareas o trabajos que esté realizando, herramientas de apoyo como son (recursos, blog, links) y también tiene a la disposición herramientas de actividades para realizar aportes a su aprendizaje como son (foros, tareas, cuestionarios, recursos, chat, twitter), entre otras, las cuales se detallan a continuación:

Foros.- la herramienta de los foros permite al estudiante realizar actividades para el debate entre usuarios dentro de un mismo curso.

- Existen dos tipos de foros disponibles en la plataforma del EVA: foros exclusivos para los profesores, de noticias del curso y foros abiertos para todos.
- Todos los mensajes que se crean dentro del curso llevan adjunta la foto del autor.
- Las discusiones pueden verse anidadas, por rama, o presentar primero según el orden que fueron realizados, el más antiguo o el más nuevo.
- Si el profesor desea puede obligar a que el usuario se suscriba a todos a un determinado foro dentro del curso o dejar a disposición de cada usuario elegir a qué foros suscribirse, con la finalidad de recibir una copia de los mensajes realizados del foro al cual se suscribió por correo electrónico.
- El profesor tiene la opción de permitir o no permitir aceptar respuestas dentro de un foro a los estudiantes.
- El profesor puede mover fácilmente los temas de discusión realizados dentro del curso entre distintos foros.
- Las imágenes adjuntas se muestran dentro de los mensajes.
- Si se usan las calificaciones de los foros, pueden restringirse a un rango de fechas.

Recursos.- Los recursos son medios o archivos de apoyo que el profesor pone a disposición de los estudiantes para su uso, la plataforma permite la presentación de cualquier contenido digital, Word, PowerPoint, Flash, video, sonidos, PDFs, etc.

- Los archivos pueden ser elaborados de diferentes maneras, pueden subirse y manejarse dentro del servidor, o pueden ser creados directamente usando formularios Web (de texto o HTML).
- Los archivos se pueden enlazar con contenidos externos en la Web o incluirlos perfectamente en la interfaz dentro del curso.
- Pueden enlazarse aplicaciones Web, transfiriéndoles datos.

La herramienta de los recursos permite al profesor dar pautas o recomendaciones para el desarrollo de las tareas y actividades propuestas dentro del curso, y a los estudiante les sirve como material de apoyo dependiendo del tipo de recurso que se les entregue.

Tareas.- Las tareas son actividades que los profesores envían a los estudiantes a través de la plataforma y permiten al profesor calificar estos trabajos enviados por los alumnos.

- Dentro de las tareas existen opciones como especificar la fecha final de entrega de la tarea, además del valor máximo de calificación que se le puede asignar a las tareas.
- Los estudiantes pueden subir sus tareas (en cualquier formato de archivo) a través del EVA. Y con ello se registra automáticamente la fecha en que se han subido.
- También se puede enviar tareas fuera de tiempo, pero el profesor puede ver claramente el tiempo de retraso que hay desde la fecha establecida.
- Para cada tarea en particular, puede evaluarse a la clase entera (calificaciones y comentarios) en una única página con un único formulario.
- En cada tarea del estudiante se le puede agregar observaciones que el profesor considere, y estas se adjuntan a la página de cada una de las tareas de cada estudiante y se le envía un mensaje de notificación.
- El profesor tiene la posibilidad de permitir el reenvío de una tarea tras su calificación (para volver a calificarla).

La herramienta de las tareas es una actividad que permite al profesor evaluar o medir el nivel aporte dentro del curso en el que se encuentra enrolado el estudiante, además permite al estudiante realizar aportes dentro del curso.

Chat.- El chat permite conversaciones entre usuarios en tiempo real

- El chat permite una interacción fluida mediante texto síncrono.
- Esta herramienta incluye las fotos y nombre del usuario en la ventana de chat.
- Soporta direcciones URL, emoticones, integración de HTML, imágenes, etc.
- Todas las sesiones quedan se almacenan dentro del perfil del usuario, las cuales ser vistas en cualquier momento, y pueden ponerse a disposición de los estudiantes.

El chat es una herramienta que permite al usuario comunicarse mediante conversaciones en tiempo real, además permite la comunicación con su profesor/tutor para realizar preguntas o dudas acerca de los contenidos del curso.

Consulta.- La herramienta de consulta es similar a una encuesta, el profesor formula una única pregunta y ofrece a los usuarios distintas elecciones.

- Esta herramienta puede usarse para votar o dar una valoración sobre algo o para recibir alguna opinión o respuesta de cada estudiante.
- El profesor puede ver una tabla que presenta de forma intuitiva la información sobre quién ha elegido qué o que decisión ha tomado.
- Se puede permitir que los estudiantes vean un gráfico actualizado de los resultados.

La herramienta de la consulta permite al estudiante dar la opinión o criterio de algo, eligiendo mediante una sola pregunta diferentes opciones, esta herramienta sirve para realizar votaciones entre estudiantes cuando se desee realizar alguna actividad y existan varias opciones y permite que el estudiante elija una de las opciones.

Cuestionario.- Los cuestionarios son una actividad que permite en los estudiantes medir el nivel de conocimientos de los contenidos impartidos en el curso, existen exámenes de diferente tipo: de respuesta múltiple, verdadero/falso y de respuestas cortas.

- El profesor puede definir una base de datos de preguntas que podrán ser reutilizadas aleatoriamente en diferentes cuestionarios.
- Las preguntas pueden ser almacenadas en categorías de fácil acceso, y también pueden ser "publicadas" para hacerlas accesibles desde cualquier curso del sitio.

- Los cuestionarios se califican automáticamente, y pueden ser recalificados si se modifican las preguntas.
- Los cuestionarios pueden tener un límite de tiempo y luego no estarán disponibles.
- El profesor puede determinar si los cuestionarios pueden ser resueltos varias veces y si se mostrarán o no las respuestas correctas y los comentarios.
- Las preguntas y las respuestas de los cuestionarios pueden ser mezcladas (aleatoriamente) para disminuir las copias entre los alumnos.
- Las preguntas pueden crearse en HTML y con imágenes.
- Las preguntas pueden importarse desde archivos de texto externos.
- Los intentos pueden ser acumulativos, y acabados tras varias sesiones.
- Las preguntas de opción múltiple pueden definirse con una única respuesta o múltiples respuestas que sean correctas.
- Pueden crearse preguntas de respuesta corta (palabras o frases).
- Pueden crearse preguntas tipo verdadero/falso.
- Pueden crearse preguntas de emparejamiento.
- Pueden crearse preguntas aleatorias.
- Pueden crearse respuestas incrustadas.
- Pueden crearse preguntas numéricas (con rangos permitidos).
- Pueden crearse textos descriptivos y gráficos.

La herramienta de los cuestionarios permite a los estudiantes medir su nivel de conocimiento de los contenidos recibidos, además permite a los profesores saber según los resultados que contenidos necesita profundizar más.

Lección.- La herramienta de la lección consiste en una serie de páginas flexibles que ofrecen contenido y preguntas para el estudiante.

- Al final de cada página se plantea una pregunta con varias posibles respuestas.
- Según la opción que escoja el estudiante para esa respuesta se le mostrarán unas u otras de las páginas restantes.

Blog.- La herramienta del blog es un recurso de consulta que permite al estudiante realizar búsquedas de temas disponibles dentro del blog.

- Variedad de temas subidos por estudiantes dentro del blog.
- Permite realizar búsquedas de temas referentes a las materias dictadas dentro de los diferentes cursos.

Recursos Educativos Abiertos.- Los recursos son medios o archivos de apoyo que el profesor pone a disposición de los estudiantes para su uso, los cuales pueden ser cualquier archivos, artículos o contenidos de internet que estén bajo la licencia Creative Commons.

- Los archivos se pueden enlazar con contenidos externos en la Web o incluirlos perfectamente en la interfaz dentro del curso.
- Pueden enlazarse aplicaciones Web, transfiriéndoles datos.
- Puede también enlazarse con las tutorías académicas que se ofertan por el sistema de videoconferencias al canal de youtube.
- Se pueden usar diferentes recursos de apoyo, como láminas de power point desde el canal de Slideshare, donde se depositan presentaciones de trabajos de la UTPL.

2.5. Datos para la obtención de las rutas de navegación de los estudiantes en el EVA

El análisis de los datos de navegación de los estudiantes en el EVA, se los puede encontrar a partir de la obtención de ciertos parámetros de interacción, su método de aprendizaje en la plataforma y las herramientas que más frecuenta y navega, ya que representan la secuencia que el estudiante tiene con los objetos y contenidos del curso dentro del sistema, estos patrones se los obtiene con el análisis previo de interacción con las herramientas disponibles en el EVA, para la comunicación del estudiante con el profesor.

Para obtener los datos de navegación, se lo realizará mediante consultas en lenguaje SQL, y para ello, se deben extraer las entidades y atributos con los que interactúa el estudiante.

Tabla 6. Atributos relacionados a la navegación del usuario con el EVA.

MOODLE		DESCRIPCIÓN
TABLAS	ATRIBUTOS	
mdl_log	action	Peticiones realizadas por el usuario en el curso dentro del sistema
mdl_log	module	Herramientas en las que navega el usuario dentro de un curso.
mdl_log	url	Dirección del recurso al cual accede el usuario dentro del sistema.
mdl_log	info	Objeto o lugar al cual accede el usuario luego de una petición.
mdl_message	message	Herramienta que permite escribir, leer, enviar mensajes.
mdl_message	userid	id del usuario que envía o remite el mensaje.
mdl_assignment	course	Tareas a las que accede en los cursos.
mdl_assignment	view	Accesos del estudiante a revisar las tareas dentro del sistema.
mdl_assignment	view all	Ingreso del estudiante a la herramienta tareas dentro del EVA
mdl_assignment	upload	Tareas que el estudiante sube al sistema dentro de un curso.
mdl_chat	course	Chat al que accede el estudiante en el curso que esta matriculado.
mdl_chat	chattime	Fecha que el estudiante ingreso al chat.
mdl_chat_messages	message	Mensajes a los que el estudiante accede dentro de los cursos.
mdl_chat_messages	timestamp	Fecha de ingreso de los mensajes en el sistema.
mdl_course_categories	coursecount	Número de veces que el usuario ha ingresado a un curso.
mdl_forum	course	Curso al cual pertenece el foro al que accede el estudiante.
mdl_forum	assesstimestart	Fecha y hora de acceso del estudiante al foro dentro del sistema.
mdl_forum	assesstimefinish	Fecha y hora de finalización del estudiante al foro en el sistema
mdl_forum_discussions	course	Curso en el que el estudiante realiza una discusión del foro.
mdl_forum_discussions	forum	Foro en el que el estudiante realiza una discusión
mdl_forum_discussions	firstpost	Primer mensaje ingresado en el foro dentro del sistema.
mdl_forum_discussions	timestart	Fecha y hora de inicio de ingreso a la discusión del foro
mdl_forum_discussions	timeend	Fecha y hora de finalización del ingreso a la discusión del foro
mdl_forum_posts	discussion	Número de discusiones realizadas por el estudiante en los foros.
mdl_forum_posts	created	Fecha y hora de creación de la discusión del estudiante en el foro.
mdl_forum_posts	modified	Fecha y hora de modificación de la discusión del usuario en el foro
mdl_forum_posts	message	Mensaje que ingresa el estudiante dentro del foro.
mdl_forum_read	forumid	Id del foro al que ingresa el estudiante dentro del sistema.
mdl_forum_read	discussionid	id de la discusión dentro del foro a la que accede el estudiante.
mdl_forum_read	firstread	Fecha y hora que el estudiante leyó por primera vez el foro.
mdl_forum_read	lastread	Fecha y hora que el estudiante leyó por última vez el foro.
mdl_forum_subscriptions	forum	Número de veces que el estudiante se suscribe en un foro.
mdl_modules	name	Nombre de herramienta a las que accede el estudiante.
mdl_quiz	course	Curso al que el estudiante ingresa para rendir el exámen.
mdl_quiz	name	Nombre del examen que el estudiante ingresa dentro del sistema.
mdl_quiz	timeopen	Fecha y hora que el estudiante abre el exámen dentro del EVA.
mdl_quiz	timeclose	Fecha y hora que el estudiante cierra el exámen dentro del EVA
mdl_twt	body	Mensajes ingresados por el usuario dentro del curso.

Fuente: Ruiz, F., (2014)

Una vez realizado el análisis de entidades y atributos en los que el estudiante interactúa dentro del sistema, se realizó la extracción de los principales datos relacionados a la navegación de los estudiantes en el EVA, como se detallan en la **Tabla 6**, estos datos son los atributos de las tablas seleccionadas de la base de datos del EVA, con una breve descripción de la función que tienen que cada una de ellas dentro del sistema, estos atributos permitirán identificar el modelo de usuario y su navegación por el EVA de la UTPL, de la carrera de Informática, que es en la que se encuentran enrolados los estudiantes; esta permitirá seguir la ruta de navegación mediante las acciones realizadas en las herramientas dentro del curso, y partir de ello, determinar indicadores de comportamiento del estudiante cuando este navega por el EVA.

2.6. Herramientas de interacción en el EVA

Las herramientas disponibles en los cursos dentro de la plataforma del EVA.

Tabla 7. Herramientas de interacciones de los estudiantes en el EVA.

Herramientas	Descripción
Curso	Cursos en los que se encuentran enrolados los estudiantes.
Assignment	Tareas que realizan los estudiantes a través del EVA
Chat	Conversaciones que realiza el estudiante con los demás usuarios del curso, además del tutor; creando un entorno de colaboración entre estudiantes y tutor.
Forum	Medio que permite mantener discusiones sobre algún tema específico dentro del contenido de la asignatura.
Twitter	Son mensajes que el estudiante realiza a través del EVA, dentro del curso con otros estudiantes que se encuentran enrolados en el mismo curso y tutor.
Quiz	Son exámenes que el estudiante rinde a través del EVA, dentro del curso al cual pertenece dicho examen.
Resource	Recursos que el tutor pone a disposición del estudiante, con información que puede ser leída, vista, bajada o usada de alguna forma para extraer información que apoye el proceso de aprendizaje.
Wiki	Medio que permite buscar tareas o trabajos almacenadas dentro de las wikies
Cuestionario	Permite realizar preguntas, obtener respuestas y generar una calificación. El tipo de ítems que se pueden plantear es variado: de opción múltiple, verdadero/falso, emparejamiento, respuesta corta, numérico, calculadas y de completar.
Blog	Permite publicar, comunicar y exponer e intercambiar ideas; puede organizar y almacenar toda esa información.
Message	Medio de mensajería del cual disponen los estudiantes para la comunicación con los estudiantes de un curso y su tutor.
User	Perfil de usuario del estudiante dentro de la plataforma con la información de matrícula.
Upload	Permite subir archivos (Word, Excel, jpg, swf etc.) dentro de la plataforma del EVA.

Fuente: Ruiz, F., (2014)

Las herramientas de las cuales dispone el estudiante dentro de la plataforma del EVA, permiten determinar las rutas de navegación e interacción que los estudiantes tienen con cada una de

ellas. En la **Tabla 7**, se describen las herramientas con la descripción de cada una de las mismas, las cuales fueron consideradas en esta sección, con la finalidad de seleccionar sólo aquella información que en verdad sea útil e importante para el proceso de minería de datos en la cual se desea determinar a través de la interacción de los estudiantes con el EVA, las rutas de navegación de los estudiantes, y con ello descubrir los patrones que estos tienen cuando navegan por el sistema.

Para la extracción de los datos de interacción de los estudiantes con los servicios o herramientas de la plataforma del EVA, se tomara en cuenta sólo aquellas tablas que están directamente relacionadas con su navegación como fue detallado en el diagrama Entidad-Relación mencionado en el análisis de la plataforma del EVA.

En el (**ANEXO A**), se pueden apreciar una descripción de estas tablas las cuales fueron detalladas cada una de ellas y la relación que tienen con las demás tablas. Estas tablas son las más importantes de la base de datos del EVA para la minería de los datos.

CAPITULO III:

ANÁLISIS Y SELECCIÓN DE HERRAMIENTAS Y TÉCNICAS.

3. Análisis y selección de herramientas y técnicas.

3.1. Introducción.

En esta sección, se pretende establecer una descripción clara y detallada del proceso de selección y aplicación de cada una de las herramientas, tanto para la extracción de datos, como para el análisis y descubrimiento de patrones, y las técnicas que serán utilizadas para el desarrollo del proyecto. El conjunto de datos que serán aplicados para la abstracción del conocimiento, proviene de los ficheros dentro de la base de datos, que son los Logs, los mismos que se encuentran en tablas en el servidor Web, y es conocido como dominio de conocimiento genérico, a partir de esta información será posible establecer relaciones y asociaciones posteriores a partir del conocimiento subyacente en la información de datos contenidos en los diversos archivos que registran todos y cada uno de los eventos o sucesos relacionados con los objetos requeridos en el servidor Web como consecuencia de la navegación de los estudiantes dentro del sistema; es debido a eso que estos conocimientos son de suma importancia para la extracción y la eficiente obtención de conocimientos, en la búsqueda de patrones de navegación a partir de este dominio de datos [Valenzuela R., 2006].

Los procesos asociados con el Data Mining que serán aplicados a los datos del servidor, consideran tareas de alto nivel para realizar la limpieza y extracción de datos, y la identificación de cada una de las tareas que intervienen con la información suministrada del servidor como son: las páginas visitadas y la ruta completa, previo al proceso de minería de datos.

3.2. Selección de herramientas.

Para el proceso de extracción de los datos y minería Web se requiere de ayuda de herramientas que permitan realizar este proceso de manera rápida y eficiente, la cual permita obtener de toda la información proporcionada dentro de la base de datos sólo aquella a la cual se le realizará el proceso de minado, para lo cual se consideraron luego de un previo estudio las siguientes herramientas para cada una de las fases de minería Web.

3.2.1. Herramientas para la recolección de datos.

Esta actividad consiste en la recolección de los datos necesarios para nuestro análisis, es decir, elegir los cursos que se tiene interés en realizar el uso de la minería para determinar los indicadores que se están buscando, debido a la gran cantidad de datos almacenados en Moodle, se realizará una diagrama **Entidad - Relación** con la herramienta **Sybase PowerDesigner v15.1.0**, y de esta manera filtrar sólo los datos necesarios que serán procesados en la presente

tesis; se tomarán todos los datos que pertenecen al noveno ciclo de la carrera de Informática de la UTPL, Modalidad de la Abierta, para ponerlos en un formato adecuado y entendible y proceder luego a la recuperación automática de la información relevante del EVA, para su posterior procesamiento. Para ello, se toma una muestra de estudiantes en determinado periodo académico de estudio (**abril 2013 – agosto 2013**), dentro de la base de datos, para poder inferir conocimiento de esos datos, para su posterior análisis.

3.2.2. Herramientas para la extracción y procesamiento de datos del EVA.

Para facilitar la estructura y la navegación basada en la información de uso de los estudiantes, necesitamos métodos y herramientas de análisis para observar el comportamiento de los estudiantes cuando estos interactúan con el EVA, y de esta manera poder ayudar a los tutores en la detección de posibles errores, deficiencias y posibles mejoras, para ello se trabajará con la herramienta de *Navicat for MySQL* para la consultas en las tablas y extracción de conocimiento necesario para el análisis y la posterior minería de datos a los mismos. Los Log, son datos que se generan automáticamente en tablas dentro de la base de datos y contiene información de alto valor para inferir dichos datos.

En esta sección se consideraron varias herramientas para la extracción y procesamiento de los datos dentro de la plataforma, para lo cual se configuraron las herramientas presentadas a continuación:

3.2.2.1. XAMPP Control.

La herramienta *XAMPP Control* es una tecnología que combina muchos paquetes de software diferentes en uno sólo, y para este caso el que se utilizara será el Apache para el funcionamiento del servidor localhost y MySQL para gestionar la base de datos.

En el archivo *phpMyAdmin* podemos encontrar nuestra base de datos, la misma que fue subida mediante la herramienta del XAMPP para la extracción de los datos que se encuentran dentro de la base de datos del EVA.

3.2.2.2. Navicat for MySQL.

Navicat for MySQL es una herramienta de administración y manejo de base de datos, rápida, fiable y asequible que fue diseñada específicamente para simplificar la administración de las bases de datos y reducir los costos de administración, esta tecnología es muy potente para gestionar base de datos, es por lo cual se la escogió para el proceso de extracción de los datos

del EVA, mediante consultas en lenguaje SQL, y además servirá para crear la base de datos, con los nuevos datos extraídos del EVA. Una vez instalada la herramienta procedemos con las consultas que permitirán posteriormente descubrir los indicadores con un análisis previo de los datos obtenidos.

3.2.2.3. MySQL-connector-java-5.1.17.

La herramienta *MySQL-connector-java-5.1.17* permite que *MySQL* provea conectividad para aplicaciones cliente desarrolladas en el lenguaje de programación Java a través de un **driver JDBC**, llamado **MySQL Connector / J**; MySQL Connector / J es un driver JDBC tipo 4, lo cual significa que el controlador es puro Java implementación del protocolo de MySQL y no se basa en las bibliotecas de cliente MySQL. Esta herramienta permite conectar nuestra base de datos con la herramienta Weka para la implementación del algoritmo para la minería de los datos.

Una vez extraídos los datos con los cuales se va a trabajar para el minado, se procederá a ordenarlos de forma estructurada y prepararlos para la próxima etapa; para ello, se desarrollara cada uno de los pasos dentro de las fases para llegar a la obtención de estos datos, para lo cual se utilizaran herramientas propias para obtener esta valiosa información de una forma automática, las cuales se detallaron anteriormente ya que se trabajará con las mismas herramientas dentro de esta fase.

El procesamiento de los datos se lo va a realizar de forma directa mediante consultas en lenguaje SQL dentro de la base de datos con la herramienta *Navicat for MySQL*, la misma que permite conectarse directamente con la base de datos del EVA, para lo cual está la herramienta de *mysql-connector-java-5.1.17*, la cual permite una conexión directa de la base de datos del EVA con la herramienta de minado Weka; y de esta manera poder visualizar las tablas y atributos de la plataforma y poder realizar las consultas necesarias para el posterior análisis.

3.3. Herramientas para la minería de datos.

Para el análisis de la minería de los datos extraídos de la plataforma del EVA previos a la minería, se consideraron varias herramientas; y para ello, se realizó una revisión previa de algunas herramientas enfocadas especialmente en el análisis de archivos Logs. Existen algunas herramientas que trabajan incorporadas al servidor de navegación o de correo, entre ellas se encuentran: *OmniAnalyzer*, *AWStats*, *Deep Log Analyzer V 3.1*, *Advanced Log Analyzer*, y *Weblog Expert*. Entre los comerciales, están *DB Miner* y *SpeedTracer*, entre los públicos: *STstat* y *Analog*.

Por la gran variedad de herramientas existentes para el análisis de las rutas de navegación dentro de los Log dentro de la base de datos, se creyó conveniente tomar tres de estas herramientas y realizar un previo estudio que permita cuál de ellas es la mejor para realizar el análisis según las objetivos planteados en la presente tesis, no se han incluido más, ya que, o bien porque su aparición es muy reciente, resulta difícil encontrarlas libremente en la Web o no son adecuadas para el tipo de análisis que se desea realizar. Por esta razón, se ha elegido las herramientas de las cuales existe mayor información y con resultados eficientes para posteriormente elegir con cual trabajar según sus características y resultados. Las herramientas seleccionadas para el proceso de minería se presentan a continuación con una descripción detallada de cada una de ellas, además luego se realizará una tabla comparativa de las tres herramientas para determinar cuál es la que más se adapta a los requerimientos planteados.

3.3.1. Webalizer⁴.

Webalizer, es un software para el análisis de los ficheros Logs de los servidores Web, esta herramienta ofrece reportes de datos que ayudan a conocer el funcionamiento del mismo, ya sean: números de accesos al sitio, los 30 Urls más accedidos entre los disponibles en el sitio, incluye todos los tipos de ficheros que forman la página Web disponible en el URL, además, las principales páginas de entrada y salida, como los de imágenes y los de sonido, etc. Estos datos se procesan y almacenan mensualmente en una base de datos Access Microsoft para facilitar su posterior análisis. Para el proceso de datos se toma una muestra, ya sea un periodo de tiempo, y se selecciona una muestra de ficheros Logs para identificar los URLS externos desde los cuales accedieron los estudiantes al sitio.

Esta herramienta no ofrece la información de los sitios de procedencia, por ello, es necesario procesar los ficheros Logs donde quedan registradas las URLs de donde provienen los usuarios que acceden al sitio, incluidos los motores de búsqueda. El procesamiento de los ficheros Logs no es una tarea sencilla. El volumen de los datos y de los ficheros a procesar es alto. **[Domínguez A., 2003].**

El programa permite:

- Buscar Urls y que no muestren ningún acceso de acuerdo con el diagrama de navegación del sitio.
- Identificar los enlaces externos utilizados por la audiencia para navegar hasta el sitio.

⁴ http://scielo.sld.cu/scielo.php?pid=S1024-943520030005000008&script=sci_arttext

- Identificar los enlaces externos correspondientes a motores de búsqueda.
- Crear un fichero con todos los enlaces externos identificados con el URL del sitio que accedieron.

Webalizer ofrece un conjunto de datos generales que facilitan los trabajos de administración de un sitio. Dichos datos permiten medir el uso de sus páginas y analizar su funcionamiento. Del conjunto de datos que ofrece el software en esta categoría, son relevantes, para los estudios sobre la utilización de un sitio por sus usuarios, las siguientes variables:

- Total de accesos
- Total de páginas
- Total de visitas
- Total de Urls
- Total de tiempo

3.3.2. Sawmill⁵.

Sawmill, es una herramienta para el análisis de minería de texto de los archivos Log, la herramienta disponible es el Sawmill7.2.9_x86_win32 (Demo), que puede utilizarse en servidores de navegación con ISA SERVER Proxy, es decir, que puede emplearse en servidores que generan Log con una estructura diferente a los que genera Internet Information Server.

Sawmill es una potente herramienta de análisis de Log, con la cual se han solucionado numerosos bugs y se ha añadido ayuda para una gran variedad de nuevos formatos de Log. Está especialmente diseñado para analizar Log de acceso a servidores Web, pero puede procesar casi cualquier Log. Las estadísticas son jerárquicas, atractivas, y llenas de enlaces que facilitan la navegación. El programa incorpora una completa documentación y se ejecuta como un programa CGI en un servidor Web, que publica una intuitiva interfaz gráfico de usuario, que puede utilizarse desde cualquier navegador para configurar y ejecutar Sawmill o para ver estadísticas de páginas. Las estadísticas son jerárquicas, atractivas y poseen enlaces que facilitan la navegación. El programa incluye una completa documentación [Fuentes S. & Ruiz M., 2007].

Sawmill ofrece una gran cantidad de opciones, incluida una base de datos persistente, el control sobre la apariencia de las páginas de estadísticas y diversas opciones de filtrado sobre el Log. Este software muestra, tras su instalación, una interfaz amigable en Windows Internet Explorer y

⁵ http://bvs.sld.cu/revistas/aci/vol16_4_07/aci111007.html

presenta, en un cuadro de selección de opciones ubicado a la izquierda, una serie de estadísticas posibles:

- Cantidad de visitas por hora, por día, por mes, etc.
- Horas pico y horas de baja audiencia.
- Páginas más visitadas.
- Páginas de entrada y salida más frecuentes del sitio.
- Utilización de buscadores, clasificación de palabras clave empleadas para buscar.

3.3.3. Weka⁶.

WEKA es un programa de código abierto distribuido bajo la licencia pública general GNU, se considera como una colección de algoritmos de aprendizaje automático escritos en java para tareas de minería de datos de aprendizaje máquina para tareas de minería de datos. Los algoritmos pueden ser aplicados a un conjunto de datos ya sea directamente o llamados desde el propio código del usuario. *Weka* contiene herramientas para pre-procesamiento, clasificación, regresión, aglomeración, reglas de asociación y visualización. Puede ser también utilizado para el desarrollo de nuevos sistemas de aprendizaje, máquina.

Weka soporta varias tareas estándar de minería de datos, especialmente, pre-procesamiento de datos, clustering, clasificación, regresión, visualización, y selección. Todas las técnicas de *Weka* se fundamentan en la asunción de que los datos están disponibles en un fichero plano (flat file) o una relación, en la que cada registro de datos está descrito por un número fijo de atributos (normalmente numéricos o nominales, aunque también se soportan otros tipos). *Weka* también proporciona acceso a bases de datos vía SQL gracias a la conexión JDBC (Java Database Connectivity) y puede procesar el resultado devuelto por una consulta hecha a la base de datos. No puede realizar minería de datos multi-relacional, pero existen aplicaciones que pueden convertir una colección de tablas relacionadas de una base de datos en una única tabla que ya puede ser procesada con *Weka* [Witten; Frank; Trigg; Hall; Holmes & Cunningham, 1999].

3.4. Características de Weka.

- Es una herramienta disponible libremente bajo la licencia pública general de GNU.
- Es muy portable porque está completamente implementado en Java y puede correr en casi cualquier plataforma.

⁶ http://es.wikipedia.org/wiki/Weka_%28aprendizaje_autom%C3%A1tico%29

- Contiene una extensa colección de técnicas para el pre-procesamiento de los datos y el modelado.
- Es fácil de utilizar por un principiante gracias a su interfaz gráfica de usuario.

Una de las características más importantes que ofrece Weka es que proporciona acceso directo con la base de datos del EVA, mediante lenguaje SQL con la herramienta de *Navicat for MySQL* que será el gestor de la base de datos, y se lo realizará con la ayuda de la herramienta de conexión *JDBC* (Java Database Connectivity) que puede procesar el resultado devuelto por una consulta hecha a la base de datos desde Weka. No se puede realizar minería de datos multi-relacional, pero existen aplicaciones que pueden convertir una colección de tablas relacionadas de una base de datos en una sola o tabla para que pueda ser procesada con Weka.

3.5. Filtros de atributos en Weka.

La herramienta de Weka cuenta con varios filtros para el proceso de los datos, entre ellos están los filtros aplicados a los atributos de los datos, entre los cuales se mencionan algunos de los filtros más implementados, filtros para eliminar atributos, para discretizar atributos numéricos, y para añadir nuevos atributos con expresiones, por la frecuencia con la que se realizan estas operaciones, de los cuales se aplicaran en nuestros datos los siguientes filtros:

3.5.1. Filtro de discretización.

Los filtros son muy útiles cuando se trabaja con atributos numéricos, muchas herramientas de análisis requieren datos simbólicos, y por tanto se necesita aplicar esta transformación antes de aplicar algún algoritmo. También son necesarios cuando se desea hacer una clasificación sobre un atributo numérico. Este filtrado transforma los atributos numéricos en simbólicos, con una serie de etiquetas resultantes de dividir la amplitud total del atributo en intervalos, con diferentes opciones para seleccionar los límites.

3.5.2. Filtro de selección.

En algunos de los algoritmos, se aplicará el filtro de atributos "*Remove*", el cual permitirá eliminar una serie de atributos del conjunto de entrada, que no se necesiten para el modelado. Se puede eliminar atributos innecesarios o que no sirvan para el tipo de análisis que se esté realizando, en este caso se lo utilizará para eliminar información que estropea el proceso de predicción del algoritmo.

3.5.3. Filtro de instancias.

De entre todas las posibilidades implementadas para filtros de selección de instancias (selección de rangos, muestreos, etc.), se puede utilizar filtros para seleccionar distancias cuyos atributos cumplan con determinadas condiciones.

3.6. Tabla comparativa de las herramientas para el pre-procesamiento de datos.

Las herramientas para el análisis de Logs, son detallados cada una de ellas a continuación:

Tabla 8. Datos comparativos de las herramientas para el análisis de Logs.

	Webalizer	Sawmill	Weka
PROCESO	Procesa todo tipo de ficheros que forman la página Web.	Poderosa herramienta de análisis, procesa casi todo Log	Procesa cualquier tipo de archivo.
EJECUCIÓN	Se ejecuta como un programa CGI en un servidor Web.	Se ejecuta como un programa CGI en un servidor Web, o servidor independiente.	Programa de código abierto distribuido bajo la licencia pública general GNU.
VENTAJAS	Permite medir el uso de páginas y analizar su funcionamiento.	Tiene intuitiva interfaz gráfica de usuario. Sofisticada y flexible para generar informes estadísticos.	Contiene una extensa colección de técnicas para pre-procesamiento de datos y modelado.
PROCESO	Trabaja con muestras de periodos de tiempos	Diversas opciones de filtrado sobre el Log	
VENTAJAS	Un fichero Log, en dependencia del tamaño, diseño y accesos del sitio, puede tener más de 15 megas y se genera cada día.	Configurar y ejecutar desde cualquier navegador.	Es fácil de utilizar por un principiante gracias a su interfaz gráfica de usuario.
VENTAJAS	Alto volumen de los datos y de los ficheros a procesar.	Las estadísticas son jerárquicas, poseen enlaces que facilitan la navegación. Navegación ISA SERVER Proxy	Está disponible libremente bajo la licencia pública general de GNU.
DESVENTAJAS	No presenta datos de interacción con las herramientas.	Sawmill utiliza la memoria incluso cuando no esta siendo utilizado activamente	No cubren los algoritmos incluidos en Weka es el modelado de secuencias.
OPCIONES	Base de datos Access Microsoft facilita el análisis.	Base de datos muy persistente.	Desarrollo de nuevos sistemas de aprendizaje.
SISTEMA OPERATIVO	Plataformas Windows	Unix, Windows, Mac – OS, Linux.	Portable, implementado en Java y corre casi en toda plataforma.

Fuente: Ruiz, F., (2014).

La **Tabla 8**, presenta un análisis comparativo de tres herramientas que serán estudiadas para el proceso de datos de los Log, por las características presentadas en la tabla anterior y estudios ya realizados anteriormente con esta herramienta y por su eficiencia con los resultados, la herramienta con la que se trabajara es **Weka** debido a la variedad de factores que tiene esta

herramienta, tiene un gran conjunto de técnicas y algoritmos disponibles para cualquier tipo de minería y por su facilidad de manejo para el proceso de minería de los datos.

3.7. Herramientas para el descubrimiento de patrones.

Para el descubrimiento de patrones existen múltiples técnicas, las mismas que son aplicables para ese tipo de análisis. Entre ellas, están las técnicas de agrupamiento y clasificación, reglas de asociación y el hallazgo de secuencias frecuentes de patrones, etc.

3.7.1. Técnicas para la minería de datos.

Del estudio realizado en el estado del arte de las técnicas de minería de datos y según los objetivos planteados en esta tesis, el desarrollo de esta actividad se realizará con la herramienta **Weka**, y con ella se seleccionaron varias de las técnicas de las cuales dispone esta herramienta: **Reglas de Asociación**, que permiten encontrar mediante reglas, sucesos con mayor frecuencia en los datos de entrenamiento. La técnica de **Clasificación** mediante **Árboles de decisión** que permiten obtener un diagrama organizado a partir de los con mayor frecuencia por parte de los estudiantes y clasificarlos en un orden lógico según su frecuencia de ingresos; y la técnica de **Clustering y segmentación** que permitirán clasificar mediante grupos los datos según las características más comunes entre ellas, es decir, agrupa a los estudiantes según su comportamiento o interés en las herramientas del EVA. La experimentación con estas técnicas permitirá encontrar patrones de navegación en la interacción de los estudiantes en los datos extraídos del EVA.

A continuación se presenta una tabla con las técnicas y algoritmos seleccionados para realizar la minería de los datos extraídos del EVA; más adelante se describe cada una de ellas.

Tabla 9. Técnicas para la minería de datos del EVA.

Herramienta de Minería	Técnicas	Algoritmos
Weka	Reglas de Asociación	GSP (GeneralizedSequentialPatterns)
	Reglas de Asociación	A priori
	Algoritmos de Clustering	Algoritmo SimpleKmeans
	Algoritmos de Clustering	EM
	Algoritmo de Clasificación	J48 (Árboles)

Fuente: Ruiz, F., (2014).

Para el presente estudio de minería, se consideró realizar una comparación para el desarrollo de la misma con varias de las técnicas y algoritmos de minería propuestos en la **Tabla 9**, la cual presenta cada una de las herramientas que se utilizarán para la posterior validación de los

resultados obtenidos mediante la aplicación de cada una de estas técnicas, con diferentes experimentos para cada una de las materias propuestas.

Se trabajará con la técnica de **Reglas de asociación** ya que permite descubrir casos que suceden en común a partir de sucesos que realiza el estudiante con mayor frecuencia durante la interacción con el sistema dentro del conjunto de datos de entrenamiento, y con ello poder encontrar las páginas a las cuales los estudiantes acceden con mayor frecuencia.

Se trabajará con la técnica de **Agrupamiento o Clustering** ya que permite agrupar dentro de una serie de vectores de un mismo grupo (*clústeres*) estudiantes que comparten características comunes o comportamientos similares y los agrupa en un sólo clúster, y con ello se puede determinar según las agrupaciones encontradas cuales son las preferencias del estudiante al momento de interactuar con el sistema.

Se trabajará con la técnica **Clasificación** con **Árboles de decisión** ya que es un modelo de predicción, esta técnica permite construir diagramas de construcciones lógicas a partir de los datos de entrenamiento de la base de datos, sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva dentro de la interacción del estudiante con el sistema para encontrar valores de los atributos e indicadores de navegación de los estudiantes dentro de la plataforma del EVA.

Para el análisis de los ficheros Log, se utilizarán varias técnicas que permiten el descubrimiento de patrones de navegación en los estudiantes y ayudan a determinar según el comportamiento de los estudiantes los indicadores y atributos para representarlos en un modelo de datos del usuario de acuerdo a su interacción con cada una de las herramientas de la plataforma del EVA, y para ello, se cree conveniente realizarlo mediante el análisis de varias técnicas de minería como son las *Reglas de Asociación*, la técnica de *Clustering* y *Segmentación*, y la técnica de *Clasificación* mediante árboles de decisión.

3.7.2. Herramienta para el análisis gráfico.

Para el análisis gráfico de la interacción de los estudiantes con las herramientas y servicios dentro de la plataforma y navegación por cada una de las páginas del sistema se lo realizará con la herramienta **Gephi**, ya que es una herramienta muy potente que permite visualizar mediante grafos el nivel de interacción y la navegación de los estudiantes por las diferentes actividades dentro de los cursos en los que se encuentran enrolados.

3.7.2.1. Herramienta Gephi para la visualización de grafos.

Gephi es un programa de código abierto para la visualización y consulta de grafos. Permite manejar grafos grandes (de miles de nodos) con un buen desempeño, que permite agrupar nodos del grafo, pintarlos de diferentes colores, darle tamaños proporcionales a indicadores, hacer los arcos entre nodos más gruesos dependiendo de diverso factores, etc. A pesar que está en una versión alpha, muestra muy buena calidad, y permite la importación de archivos en los formatos más comunes para grafos; entre otras cosas, los archivos pueden ser separados por comas, haciendo más fácil aun la lectura de dichos grafos.

Esta herramienta permite realizar la vista del escenario de los patrones de navegación de los estudiantes que interactúan con las herramientas del EVA y las actividades con el tutor, es decir, permite identificar cuáles son los estudiantes que más interactúan con la plataforma y cuáles son las acciones que realiza con mayor frecuencia dentro del sistema, además cuales son las herramientas a las que accede más y las páginas más visitadas por los estudiantes, estos datos ayudaran más adelante para determinar indicadores y atributos de interacción e intereses con las herramientas de los estudiantes que interactúan con el sistema y descubrir cuáles son las herramientas más afines de los estudiantes cuando navegan en el EVA.

CAPITULO IV:

PRE-PROCESO: SELECCIÓN, LIMPIEZA Y TRANSFORMACIÓN DE DATOS.

4. Pre-proceso: Selección, limpieza y transformación de datos

4.1. Introducción

En el pre-proceso, se pretende dar sentido a los datos y comportamientos generados en las sesiones de navegación que realiza el estudiante cuando interactúa con el EVA, estos eventos se generan cada vez que el usuario realiza una petición al sistema, se almacenan en la tabla *Log (mdl_log)*, dentro del EVA. En esta sección se presenta la metodología empleada para la extracción de conocimiento, y el proceso de cada una de las fases de la minería ya mencionadas en la primera fase de esta investigación. A esta información no estructurada se la recolecta para realizar en ella una serie de tareas generales de procesamiento de datos, ya sea (*limpieza de datos previa, la identificación del usuario, identificación de sesión, interacción con las herramientas, reducción de datos, etc.*), y de esta manera realizar el pre-procesamiento de datos, el cual permite que los datos originales se transformen en una forma adecuada y con un formato entendible con la ayuda de herramientas propias para este tipo de análisis, y luego a esta información procesada se le aplica minería de datos mediante técnicas e-Learning con algoritmos específicos para el descubrimiento de patrones de navegación.

El análisis de las acciones realizadas en las herramientas por los estudiantes ayudarán a descubrir y entender el comportamiento de los estudiantes, y la colaboración del aprendizaje del mismo con la estructura actual de la Web, cuando estos navegan por el EVA, y a partir de ello en trabajos futuros se podrá determinar si se necesita realizar cambios y mejoras para el conjunto de recursos de los cuales dispone el estudiante para su aprendizaje.

4.2. Información de la base de datos del EVA

Uno de los primeros pasos dentro de las fases para la extracción de los datos, consiste en la selección de los datos que serán analizados previos a la minería de datos, es decir, la información a la cual se requiere realizar la minería, ya que no toda la información que se encuentra en la base de datos del EVA es de utilidad para el proceso de minería, para la realización de este análisis parte de la información que se encuentra en la tabla de Logs, y las tablas con las que estas se relacionan.

Dentro de la tabla Logs, se registran todos los eventos realizados por los estudiantes cuando estos interactúan con el EVA, es decir, se registran cada una de las acciones realizadas por los estudiantes cuando navegan por el sistema, con indicadores de interés para el análisis.

La tabla de los Logs de la cual se extrae los datos, y a la cual se le aplicara la minería de datos, es la base de datos del EVA, que es el sistema virtual de aprendizaje, y administrador de contenidos que usan los estudiantes para interactuar con el tutor.

En la **Fig. 4**, se presenta las tablas de la base de datos dentro de Moodle (EVA) implementadas en Navicat for MySQL, para la selección de datos previas al estudio de minería de datos, esta herramienta de Navicat for MySQL permite visualizar todos las tablas con los datos que contiene la base de datos, y poder seleccionar los datos necesarios para el análisis, además permite hacer consultas hacia las diversas tablas que están dentro de la base de datos. En este caso los principales datos que serán analizados para la presente tesis son la tabla de Logs y varios atributos de otras tablas con las que se relacionan entre sí, y de esta manera poder discernir sólo la información de los estudiantes de Informática de la modalidad Abierta de noveno ciclo de la Universidad Técnica Particular de Loja.

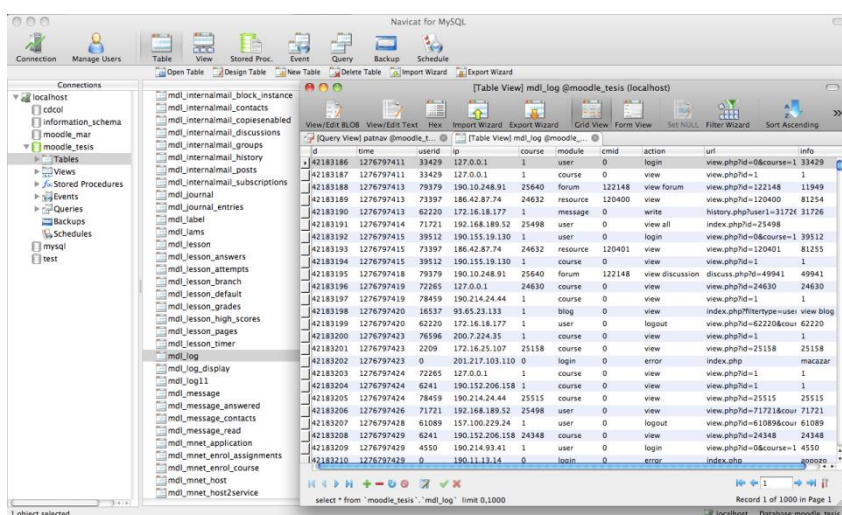


Fig. 4. Tablas de Moodle (EVA) implementadas en Navicat for MySQL.

Para la extracción, transformación y carga de los datos que provienen de la base de datos del EVA de la **Universidad Técnica Particular de Loja**, debido a su gran cantidad de información almacenada en la misma, se consideró conveniente tomar como muestra de estudio para el proceso de minería, el cual será del periodo académico **Abril 2013 – Agosto 2013**, de todos los estudiantes de la **Modalidad Abierta**, de la carrera de **Ciencias de la Computación de Noveno Ciclo**, el mismo que comprende las siguientes materias: **Administración de herramientas CASE [A], Control de Calidad [A], Modelamiento de Datos [A], Redes y Sistemas Distribuidos [A], Sistemas Basados en el Conocimiento [A] y Sistemas de Control [A]**; con el análisis de estas materias se busca relacionar los resultados obtenidos con los experimentos a realizar con las diferentes técnicas de minería, y con ello, poder definir las características o

perfiles de los estudiantes y cuáles son las características comunes de aquellos estudiantes que interactúan con las herramientas del EVA,

Además se busca determinar los indicadores de las acciones de interacción de los estudiantes con los tutores en el EVA; luego a estos datos se les dará un proceso manual mediante consultas SQL a las tablas ya mencionadas en la fase de análisis de Moodle, para ir determinando los datos que serán pre-procesados de tal manera que se tenga información de manera integrada y limpia en cuanto a su estructura y semántica antes de aplicar la minería de datos.

4.3. Proceso de Análisis de los Logs

Para el proceso de análisis de los datos de minado se lo realizará con la información almacenada en tablas dentro de la base de datos del EVA, la cual almacena las acciones, eventos o sucesos de las peticiones realizadas dentro del sistema por el estudiante en consecuencia de la navegación que realizan dentro del sistema, los archivos Logs, y demás tablas que se relacionan entre sí en el momento de interacción que tienen los estudiantes con las diferentes herramientas de la plataforma. Las herramientas con las cuales se trabajara, mencionadas anteriormente en el Capítulo III: “**Análisis y selección de herramientas y técnicas**”, permitirán medir el nivel de navegación en las páginas de la plataforma, el número de accesos a las diferentes páginas por parte del estudiante cuando interactúa con las herramientas y servicios dentro de la plataforma del EVA, entre otros factores que ayuden a determinar patrones de navegación del estudiante.

4.3.1. Pre-proceso de los datos en el EVA

Una vez seleccionada la información útil para la minería de datos, con los datos crudos es necesario realizar un proceso de limpieza en los mismos, por lo cual se procederá a realizar el pre-procesamiento de datos, la misma que permite a los datos originales transformarlos en un formato adecuado para ser utilizados y poderles aplicar un algoritmo de minería de datos con la ayuda de una herramienta que permita realizar este tipo de análisis.

Las fases a realizar previas al proceso de minería de los datos son: la limpieza de datos, la identificación de datos, identificación de acciones, obtención de caminos o rutas de navegación, identificación de procesos, transformación e integración de datos, y la reducción de datos no necesarios e inconsistentes y la extracción de los mismos. Con la finalidad de obtener todos los campos donde se almacenan las principales acciones de los estudiantes cuando interactúan con los cursos dentro del sistema, se considerarán las tablas detalladas en el (**ANEXO A**), que permitirán identificar las rutas de navegación de los estudiantes durante la interacción en el EVA.

Los datos tomados de los Logs, es información almacenada que describe todas las acciones de navegación del estudiante a través del Entorno Virtual de Aprendizaje (EVA), la misma que detalla cada uno de los campos con su respectiva descripción que contiene la entidad *mdl_log*, ver **Fig. 4**, de los cuales se tomaran los siguientes campos:

Atributo	Descripción
<i>userid</i>	Identificativo único que tiene cada uno de los estudiantes, para saber quien realizó determinada acción sobre algún objeto.
<i>time</i>	Fecha y hora en la que se realiza una acción por parte del estudiante en el EVA.
<i>course</i>	Id del curso en el que esta enrolado el estudiante e interactúo el estudiante.
<i>module</i>	Herramienta o módulo sobre la cual el estudiante realiza alguna acción.
<i>action</i>	Acciones que el estudiante realiza cuando interactúa con las herramientas del curso.
<i>url</i>	URL del módulo o recurso sobre el cual se realiza la acción dentro de la plataforma.
<i>info</i>	Dirección o localización del objeto (archivo) al que accede el estudiante durante su petición en el sistema.

Tabla 10. Descripción de la tabla de Logs.

Estas entidades permiten representar dentro del sistema de manera jerárquica y organizada las acciones realizadas en la secuencia de navegación de los estudiantes que interactúan con el EVA.

Los datos de entrada para el proceso previo a la minería de las cuales se realizará la extracción de datos para poder inferir los indicadores de la navegación de los estudiantes en el EVA, y las acciones realizadas durante su interacción con las herramientas que dispone el EVA.

4.3.2. Selección, limpieza y transformación de datos

Este proceso de análisis tiene dos pasos a seguir en la extracción del conocimiento como ya fue mencionado anteriormente en el estado del arte, la limpieza y selección de datos, y para ello se tomó en cuenta algunos factores importantes como son:

- Eliminar datos erróneos o inconsistentes, es decir, la limpieza de datos irrelevantes.
- Utilización de métodos estadísticos con Weka para realizar el siguiente proceso:

Reconocimiento y Limpieza de Datos: Este paso tiene como objetivo la detección de datos anómalos e inconsistencias. Para lo cual se tomó los datos de la base para eliminar la cantidad de datos nulos y datos erróneos, con el fin de reducir las posibilidades de error cuando sean procesados por los algoritmos de minería. Se aplicó la técnica de pre-proceso para que facilite el reconocimiento y limpieza de los datos que provee Weka, pudiendo así interpretar la validez del valor de los atributos y mejorar la calidad de los

datos.

Resúmenes de Estadísticas: La herramienta proporcionó un resumen de las estadísticas para los atributos a evaluar, con el porcentaje de instancias con valores nulos, cantidad de diferentes valores para el atributo, y la cantidad de instancias que tienen un valor único para ese atributo.

Diagramas: Se emplea un diagrama de barras segmentadas para poder visualizar cada valor distinto del atributo y la cantidad de instancias que presentan este valor.

4.4. Selección de datos para el análisis

En esta actividad, dentro del análisis de los archivos Logs, se tomará un conjunto determinado de datos de la base de datos, los cuales son algunos atributos dentro de las tablas que son actividades realizadas por los estudiantes de la UTPL cuando interactúa con el EVA. Estos datos se tomarán como datos de entrada para determinar la navegación de los estudiantes dentro del Entorno Virtual, y de esta manera poder identificar las características, propiedades, preferencias e intereses de cada uno de los estudiantes.

4.4.1. Datos de usuario

Se determinan los datos que necesitaremos de la base de datos para realizar el análisis respectivo de la navegación del estudiante por el EVA. Estos atributos permiten definir los parámetros de acceso del usuario al entorno de aprendizaje y de esta manera poder descubrir los patrones de navegación durante su interacción con el EVA de la UTPL, los mismos que se detallan a continuación:

- Id de usuario
- Centro universitario al que pertenece el estudiante.
- Ciclo en el que está matriculado el estudiante.
- Modalidad de estudio a la cual pertenece.
- Periodo académico en el que está matriculado.
- Materias en las que está enrolado dentro del sistema.

4.4.2. Datos de interacción

En esta sección se detallarán las interacciones que el estudiante realiza cuando solicita algún recurso dentro del EVA, el cual define los niveles de utilidad de los estudiantes de la UTPL en los contenidos de las herramientas del curso en el que está enrolado, además la frecuencia de

interacción del estudiante con el sistema, este último es uno de los atributos que más ayudan a definir el nivel de importancia para el aprendizaje del usuario con respecto al EVA, ya que permite identificar la frecuencia de peticiones y acciones que el usuario (sujeto) realiza sobre los recursos (objetos) del EVA, los datos de interacción son:

- Fecha de acceso al sistema.
- Herramientas a los cuales ingresa en la plataforma.
- Actividades a las que ingresa en el sistema.
- Acciones que realiza el estudiante cuando navega en el sistema.
- Información de las páginas visitadas por el estudiante cuando interactúa con el EVA.
- Páginas a las que acceden con mayor frecuencia los estudiantes.

Estos datos se almacenan automáticamente dentro de los archivos Logs, los cuales se detalla más adelante el proceso de análisis de la entidad llamada *mdl_log*, dentro de la base de datos del EVA, y son algunos de los datos que servirán para el descubrimiento de las rutas de navegación de los estudiantes que navegan en el sistema.

4.5. Descripción del proceso de extracción de datos

A partir de los datos, previo a los resultados obtenidos del análisis que se realice en este proceso son la parte fundamental de este proyecto, ya que estos datos son las acciones que los estudiantes de la UTPL realizan al momento de interactuar con el EVA, lo cual ayudará a encontrar una estructura jerárquica que determine un recorrido lógico y organizado de la información a través de patrones de navegación de los estudiantes, y con ello las diferentes maneras de aprendizaje del estudiante. Para la extracción de conocimiento de los datos adquiridos, se tiene que seguir un proceso. El proceso de las fases fundamentales que constituye la minería del uso Web, para el procesamiento de los Log, la misma que produce información de alto valor para el análisis previo de los Logs.

Para el desarrollo del presente análisis y el minado se consideraron los datos expuestos en la **Tabla 11**, la cual se la presenta más adelante, esta tabla describe los datos de extracción de la base de datos del EVA, con los cursos y el total de estudiantes enrolados dentro de los mismos, de la Universidad Técnica Particular de Loja, modalidad Abierta, noveno ciclo de la carrera de Informática del periodo académico: **Abril 2013 – Agosto 2013**.

La razón por la cual se escogió la carrera de Informática es debido al nivel de interacción que los estudiantes tienen con la plataforma del EVA, y los cinco cursos de noveno ciclo fueron

seleccionados debido al número de estudiantes que se encuentran enrolados en los cursos, que en comparación con otros cursos del mismo periodo académico cuentan con un número bajo de estudiantes enrolados en el sistema. Además, por ser los cursos con mayor nivel de interacción en relación a los demás estudiantes de la misma carrera. Los estudiantes de noveno ciclo interactúan con la mayoría de las herramientas de la plataforma, ya sea para su aprendizaje y la colaboración y comunicación con su profesor. Luego con estos datos se realizará la minería de los datos para extraer patrones de navegación de los estudiantes que interactúan con el EVA y determinar indicadores y atributos que serán representados dentro de un modelo de datos del usuario.

4.5.1. Datos de extracción

Para la fase de extracción de los datos del EVA, se toman los atributos más importantes dentro de la tabla de los Logs, los mismos que se determinaron de acuerdo a la interacción que tiene el estudiante con las herramientas del EVA, y que permitan extraer las rutas de navegación de los usuarios durante la interacción con el EVA, ya sean las acciones del estudiante sobre los objetos de los cursos en los que se encuentra enrolado dentro del sistema, para almacenarlos en una tabla dentro de la base de datos, ver (**ANEXO E**).

Una vez realizadas las consultas mediante lenguaje SQL en la base de datos, y con la información clasificada y pre-procesada de la interacción de los estudiantes en el EVA, se procedió a experimentar con los algoritmos de las técnicas propuestos para el proceso de minería.

Para el proceso de extracción de los datos del EVA, se realizó una conexión directa con la herramienta de Navicat y la base de datos del Eva, ver (**ANEXO F**). Los datos de extracción que fueron obtenidos mediante consultas en lenguaje SQL desde la base de datos del EVA, se detallan en la siguiente tabla.

Fuente: Ruiz, F., (2014).

Periodo	Modalidad	Carrera	Ciclo	Curso (Asignatura)
Abril 2013 – Agosto 2013	Abierta y a Distancia	Informática	Noveno	Administración de Herramientas CASE [A]
				Control de Calidad [A]
				Modelamiento de Datos [A]
				Redes y Sistemas Distribuidos [A]
				Sistemas Basados en el Conocimiento [A]
				Sistemas de Control [A]

Tabla 11. Datos de extracción del EVA para la minería.

La metodología utilizada para la minería del uso de la Web y la extracción del conocimiento incluye cinco fases fundamentales para el análisis de Logs, que fueron descritas en el estado del arte, estas fases están relacionadas con las etapas del uso de la minería Web, descrita anteriormente. Las fases se detallan más adelante junto con las herramientas con las cuales se trabajara para realizar cada una de las fases de la minería de datos.

4.5.2. Limpieza de datos

Este proceso es fácil, pero se debe realizar con mucho cuidado, sobre todo al momento de la selección de los datos que van a ser tratados, ya que en la tabla de los Log, existen muchos datos irrelevantes e innecesarios, además hay que tomar en cuenta que los datos sean concisos, y estén libres de datos erróneos, obsoletos o con inconsistencias; ya que de estos datos depende la veracidad y eficiencia para el posterior análisis de los resultados.

Luego de realizar el análisis de minería, lo siguiente consiste en realizar una revisión minuciosa de los datos del EVA, para luego elaborar un diagrama *Entidad-Relación* con las tablas más importantes para el análisis y proceso de los datos, la selección de las tablas se la realizó de forma manual mediante consultas SQL, desde una herramienta de gestión de base de datos, *Navicat for MySQL*, propia para la plataforma de MAC, que es, en la que se está trabajando, esta herramienta permite editar, borrar y modificar todas las tablas y los datos, además permite ejecutar consultas SQL, crear nuevas tablas, hacer consultas, etc.; en el momento de realizar las consultas para extraer los datos previos que permitirán encontrar las trazas de navegación de los estudiantes que interactúan con el sistema, por tal razón, es necesario filtrar información para que el sistema no arroje datos redundantes, incoherentes e innecesarios, y sólo nos presente los datos necesarios, ver (**ANEXO B**).

4.5.3. Transformación de datos

Para la transformación de los datos, existen varias herramientas de consulta y análisis para explotación de la información, herramientas de procesamiento analítico que sirven para visualizar y manipular datos interactivamente que se encuentran almacenados en un repositorio de datos, y la minería de datos es para encontrar patrones de navegación que ayuden a determinar los indicadores de navegación de los estudiantes que interactúan con el EVA; la razón por la que se escogió una herramienta para realizar el análisis y las consultas en el EVA, es debido a la gran cantidad de datos almacenados ahí, a los cuales hay que realizarles un proceso de depuración antes de realizar la extracción de los mismos.

La fase de la transformación y limpieza de datos, se realizó con la herramienta de Weka, la cual permite aplicar una gran diversidad de filtros y todo tipo de transformaciones sobre los datos; y de esta manera poder discernir sólo la información necesaria para determinar indicadores y patrones de navegación en los estudiantes. Los datos se filtraron en una nueva tabla con los datos de interacción de los estudiantes con el EVA.

Con la información obtenida en el pre-proceso de datos, se procedió a realizar cada una de las fases previas a la minería, para luego con los datos extraídos crear una nueva entidad dentro de la base de datos con las interacciones obtenidas de los estudiantes en el EVA mediante consultas SQL, con la finalidad de realizar la minería de datos de una manera dinámica, según las acciones de los estudiantes en las herramientas (*foros, recursos, REA, RSA, chat, tareas, cuestionarios, etc.*) o la interacción en actividades (*subir tareas, revisar blogs o la Wiki, enviar mensajes, etc.*).

Para realizar el pre-proceso, que consiste en la selección, limpieza y transformación de datos, se lo realizará con la herramienta Weka. Antes de aplicar las técnicas de minería a los datos del EVA, se realizará algunos ajustes en el módulo, esta herramienta importa datos de varios formatos: formato Weka denominado *ARFF*, archivos de texto separados por comas (*CSV*), *c4.5*, binario y puede leer datos usando la especificación *JDBC* de Java; que es con la que se trabajará. Para ello, se realizará la conexión de *MySQL* con la base del EVA y desde allí realizar el proceso de minado.

Weka contiene filtros para obtener los datos útiles, estos filtros incluyen operaciones de muestreo, selección de campos y transformación y combinación de atributos. Los algoritmos de minería que incluye caen en 3 rubros que se pueden ver en su interfaz gráfica y los cuales se tomarán para el proceso de minería del EVA, dentro de los Logs.

4.6. Tablas de la base de datos del EVA

Las principales tablas del EVA para realizar el proceso de minería de datos, mencionado en la **Tabla 2**, contienen los registros detallados de todas las actividades y sucesos realizados por los estudiantes durante su navegación por el sistema, estas tablas permiten obtener un seguimiento detallado de las acciones (*ver, revisar, subir, actualizar, responder, etc.*) que los estudiantes han realizado en las herramientas: *cursos, tareas, foros, cuestionarios, twitter, blog, perfil de usuario* y demás herramientas del EVA, que ayuden a la identificación de indicadores en la navegación de los estudiantes.

La tabla con los datos que fueron extraídos con la información detallada en el **Capítulo II** “Análisis de la plataforma del EVA de la UTPL”, presenta las acciones, ver (**ANEXO D**); para representarlo mediante datos estadísticos las acciones dentro de las herramientas más utilizadas por los estudiantes de noveno ciclo; con estos datos se puede determinar las rutas de navegación que los estudiantes tienen cuando acceden al EVA.

Una vez realizada la abstracción de los datos de los estudiantes de noveno ciclo de la Modalidad a Distancia de la carrera de Informática de la UTPL, se procedió a realizar una entidad adicional con la información de interacción de los estudiantes el EVA, la cual se detalla a continuación.

Tabla creada para almacenar los datos de interacción de los estudiantes

En esta tabla se almacenan las acciones realizadas por los estudiantes cuando interactúan con las actividades y herramientas del EVA, y la navegación por las diferentes páginas de los cursos.

Fuente: Ruiz, F., (2014).

TABLA mdl_interacción			
En esta tabla se almacena toda la información referente a la interacción que tiene el usuario con las herramientas de los cursos de Noveno Ciclo de la Universidad Técnica Particular de Loja en el EVA.			
Campo	Tipo	Tamaño	Descripción
usuario_id	bigint	10	Identificador único de cada usuario en el sistema.
num_pag_visit	integer	10	Páginas a las que ha ingresado el usuario en el sistema.
num_curso_ver	integer	10	Veces que el usuario ha ingresado al curso dentro del sistema.
num_foro_susc	integer	10	Número de foros que se ha suscrito el estudiante en el curso.
num_foro_adcdeb	integer	10	Veces que el estudiante agrega un debate dentro de un foro.
num_foros_verdeb	integer	10	Veces que el estudiante ingresa a ver los debates del curso.
num_foro_ver	integer	10	Número de veces que el usuario revisa un foro dentro del curso.
num_foro_adc	integer	10	Número de foros ingresados dentro del curso por el tutor.
num_foro_deldeb	integer	10	Número de veces que el usuario borra un debate dentro del foro.
num_foros_ver	integer	10	Número de foros que el usuario ingresa a ver dentro del curso.
num_foro_adnot	integer	10	Número de notas que el usuario agrega a un foro.
num_tarea_ver	integer	10	Número de veces que el usuario revisa una tarea.
num_tareas_ver	integer	10	Veces que el usuario revisa las tareas dentro del curso.
num_tarea_sub	integer	10	Número total de tareas subidas por los usuarios dentro del curso.
num_recurso_ver	integer	10	Veces que el usuario ingresa a revisar un recurso en el EVA
num_recurso_adc	integer	10	Número de veces que el tutor agrega un examen dentro del curso.
num_quiz_ver	integer	10	Número de veces que el usuario ingresa a revisar un examen.
num_quiz_ver	integer	10	Veces que el usuario ingresa al curso a revisar los cuestionarios.
num_quiz_int	integer	10	Número de intentos que el estudiante realiza en un cuestionario.
num_quiz_contint	Integer	10	Veces que el usuario continúa con un intento en el cuestionario.
num_quiz_intcer	Integer	10	Número de veces que el usuario finaliza un cuestionario.
num_quiz_rev	integer	10	Veces que el usuario revisa los cuestionarios dentro del curso.

Tabla 12. Tabla con la descripción de los campos de la entidad mdl_interacción.

La **Tabla 12**, fue creada para almacenar los datos de interacción (actividades en las cuales realiza alguna acción) de los estudiantes en las herramientas cuando navegan por la plataforma del EVA de los cursos en los cuales se encuentran enrolados. Los datos se muestran con una breve descripción del tipo de dato y el atributo.

Los datos de interacción de los estudiantes con las herramientas del EVA se detallan en la siguiente gráfica.

The screenshot shows a database table view titled "[Table View] mdl_interaccion @moodle_tesis (localhost)". The table contains 20 columns representing different interaction metrics for each user. The data is presented in a grid format with a toolbar at the top for various operations like sorting and filtering.

usuario_id	num_pag	num_curs	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for	num_for		
47	16	94	0	0	11	65	5	0	0	0	0	9	8	0	2	0	4	2	0	0	3	
54	23	22	2	4	19	24	0	0	4	0	6	0	0	2	3	0	0	4	1	2	1	2
191	12	29	1	4	13	22	0	0	4	4	0	0	0	0	6	0	0	0	0	0	0	0
453	11	22	0	4	14	24	2	0	18	0	7	4	0	3	5	1	0	7	2	3	0	3
1419	7	31	0	1	5	11	0	0	1	0	0	0	0	0	3	0	0	2	1	0	0	0
1492	22	18	1	2	10	20	0	2	6	2	1	0	0	0	3	1	0	0	0	0	0	0
1763	18	26	0	12	47	30	0	0	12	0	0	2	0	0	5	1	0	0	0	0	0	0
2688	10	21	2	4	38	23	2	0	18	0	0	0	0	0	3	0	0	1	0	0	0	0
2875	21	34	0	5	26	30	1	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0
4222	8	34	0	2	22	16	4	0	6	10	0	4	0	0	5	0	0	0	3	0	0	0
4235	10	17	2	6	30	24	0	0	10	0	4	2	0	2	0	2	0	6	2	2	0	2
6086	9	58	1	4	42	43	16	0	2	0	4	0	0	2	8	0	0	5	0	2	0	2
7052	12	23	0	4	21	20	0	0	2	6	0	0	0	0	3	0	0	0	0	0	0	0
67106	3	0	1	1	1	20	2	1	0	0	0	0	0	0	0	2	0	0	0	0	0	0
16057	3	1	0	1	2	10	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0
23610	8	3	2	0	20	21	3	0	1	3	3	6	8	0	0	0	1	2	6	0	0	0
28003	0	3	1	2	12	16	0	0	2	0	3	0	0	1	2	0	0	2	0	1	0	1
30416	10	4	0	2	14	11	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0
32748	0	0	0	2	32	16	0	0	0	1	0	0	2	0	1	0	0	0	0	0	0	0
33299	20	2	1	2	14	22	0	0	2	0	2	0	0	1	4	0	0	4	0	1	0	1
37251	0	10	0	2	7	12	0	0	9	0	0	0	0	0	0	0	0	1	0	0	0	0
40224	0	0	0	1	0	0	0	2	1	2	0	0	0	0	0	0	0	0	0	0	0	0
41555	0	0	0	2	2	3	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
41915	0	2	1	2	3	5	0	0	0	0	2	0	0	1	1	0	0	2	0	1	0	1
42631	0	0	1	2	13	16	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0
45798	0	4	1	1	8	18	0	0	15	0	4	2	0	1	2	1	0	2	2	1	0	1
46746	0	9	0	2	17	16	0	0	5	0	9	5	0	2	1	0	0	4	0	2	0	2
48220	0	3	0	3	16	25	0	0	27	0	3	1	0	2	4	0	0	3	6	1	1	1
50029	0	0	0	1	1	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
50705	0	0	0	3	15	8	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
54159	0	1	0	2	54	39	0	0	25	1	1	6	0	0	1	0	0	0	4	0	0	0
5636	0	0	0	2	9	13	0	1	3	0	0	1	0	0	0	0	0	0	2	0	0	0

Fig. 5 Vista de la entidad *mdl_interacción* con sus campos

Fuente: Ruiz, F., (2014).

La **Fig. 5** muestra los datos de interacción de los estudiantes enrolados en las materias de noveno ciclo, y el total de interacciones realizadas en las diferentes herramientas dentro del sistema, estos datos más adelante mediante el proceso de minería permitirán encontrar los patrones que tienen los estudiantes durante su navegación con el EVA.

Una vez cargados los datos, se procedió a realizar la conexión de la base del EVA con MySQL en Weka, ver (**ANEXO F**), para empezar el proceso de minería y luego realizar un análisis comparativo con los resultados obtenidos de los datos extraídos. Y a partir de los resultados establecer los indicadores para diseñar el modelo de datos del usuario con los indicadores.

4.7. Resultados obtenidos de la interacción con las herramientas

Los datos obtenidos del estudio que se realizó con el total de 132 usuarios enrolados en las

asignaturas de noveno ciclo de Informática de la UTPL, Modalidad a Distancia; se puede detallar que 126 estudiantes ingresaron al EVA e interactuaron con las herramientas.

Las herramientas con mayor frecuencia se resumen en la **Tabla 13**, la cual fue recopilada de los datos obtenidos mediante consultas en lenguaje SQL, que se muestran en el (**ANEXO D**), esto se realizó mediante consultas SQL, en el cual se muestra un detalle general de la interacción con las herramientas en los cursos que están enrolados los estudiantes.

```
SELECT mdl_log.userid,
       mdl_log.course,
       mdl_log.module,
       mdl_log.action,
       mdl_log.url
FROM mdl_log
WHERE course = 51749 AND (module = 'course' AND (action = 'view'))
   || course = 51751 AND (module = 'course' AND (action = 'view'))
   || course = 51763 AND (module = 'course' AND (action = 'view'))
   || course = 51755 AND (module = 'course' AND (action = 'view'))
   || course = 51758 AND (module = 'course' AND (action = 'view'))
   || course = 51746 AND (module = 'course' AND (action = 'view'))
```

Fuente: Ruiz, F., (2014).

HERRAMIENTA	ACCIONES REALIZADAS	PÁGINAS VISITADAS	NÚMERO DE INTERACCIONES
Course	view	view.php?id=51755	4378
Course	add mod	../mod/label/view.php?id=137271	34
Course	update mod	../mod/quiz/view.php?id=139670	12
Course	editsection	editsection.php?id=495848	41
Course	user report	user.php?id=28989&user=34361&mode=outline	75
Assignment	view	view.php?id=139722	330
Assignment	view all	index.php?id=51755	175
Assignment	upload	view.php?a=16289	73
Assignment	view submission	submissions.php?id=16282	43
Assignment	update grades	submissions.php?id=16282&user=47013	41
Assignment	add	view.php?id=154391	2
Assignment	update	view.php?id=154391	4
Quiz	preview	attempt.php?id=139669	7
Quiz	view	view.php?id=139682	160
Quiz	review	review.php?id=139675&attempt=67431	11
Quiz	view all	index.php?id=51755	89
Quiz	attempt	review.php?attempt=66173	61
Quiz	continue attemp	review.php?attempt=80872	14
Quiz	close attempt	review.php?attempt=66173	61
Quiz	report	report.php?id=139670	4
Quiz	report_quiz	report_quiz	2
Quiz	editquestions	view.php?id=139675	171

Quiz	update	view.php?id=139670	1
Forum	view forum	view.php?id=135428	2606
Forum	view forums	index.php?id=51755	761
Forum	add	view.php?id=135428	16
Forum	view discussion	discuss.php?d=56981	2231
Forum	subscribe	view.php?f=13159	9
Forum	view subscribers	subscribers.php?id=14288	1
Forum	add discussion	discuss.php?d=56981	301
Forum	add post	discuss.php?d=56981&parent=130155	105
Forum	delete discussion	view.php?id=141631	3
Forum	user report	user.php?id=29002&user=28456&mode=posts	24
Forum	unsubscribe	view.php?f=13085	1
Forum	update	view.php?id=154385	1
Forum	update post	discuss.php?d=56981#p128487&parent=128487	88
Forum	search	search.php?id=51755&search=	3
Blog	view	index.php?filtertype=user&filterselect=4235&postid=0&tagid=0&tag=	1
Resource	view	view.php?id=137179	507
Resource	view all	index.php?id=51750	76
Resource	add	view.php?id=146523	6
User	view	view.php?id=48077&course=51755	443
User	view all	index.php?id=51755	438
User	update	view.php?id=30035&course=51763	6
Utpl	role_assign	admin/roles/assign.php?contextid=450106&roleid=5	70
Total General.			13486

Tabla 13. Resumen de las páginas visitadas en las herramientas del EVA.

El resultado obtenido del análisis de los usuarios que interactúan cuando navegan a través del EVA, almacenados en los Logs, se presenta en la **Tabla. 13**; la cual detalla las paginas a las cuales ingresan los estudiantes y las acciones que realizan dentro de las mismas cuando interactuar con las herramientas del EVA, en el campo de *páginas visitadas* se muestran aquellas páginas de las actividades que realizan dentro de los cursos con mayor número de acceso por parte de los estudiantes en el sistema. Las actividades más frecuentadas son las siguientes:

- Las herramientas con mayor índice de interacción los estudiantes son: foros, perfil del usuario, tareas, cuestionarios, recursos, la subida de archivos, y con un porcentaje mínimo ingresan a los blogs, notas y role.
- Los estudiantes usan la herramienta del RSA (*Red Social de Aprendizaje*) para consultar sobre sus dudas e inquietudes a los docentes de los diferentes cursos.
- Los estudiantes antes de crear o realizar una discusión en los foros, primero ingresan y

revisar los foros de los compañeros ya ingresados en el sistema.

- Las acciones más frecuentes del estudiante que realizan cuando interactúa con el EVA son revisar tareas, recursos, foros, archivos y enlaces subidos por el tutor del curso dentro del EVA.
- En la herramienta de los foros revisan, postean y contestan los foros con mayor frecuencia en la mayor parte de las asignaturas.
- Los estudiantes ingresan a los cursos con mayor frecuencia para revisar las actividades propuestas por el docente.
- Una vez que el estudiante ingresa al sistema revisa en el curso, los recursos y comentarios brindados por el tutor.
- La mayor parte de los estudiantes participan de los foros propuestos por cada uno de los tutores en los cursos dentro del EVA.
- La mayor parte de los estudiantes revisaron, contestaron y finalizan los exámenes propuestos por el tutor de cada curso.

4.8. Promedio de interacción de los estudiantes con las herramientas del EVA.

El promedio total de accesos a cada una de las herramientas cuando el estudiante navega por el sistema e interactúa con cada una de estas herramientas y servicios presentados en la **Tabla 14**, se obtuvieron a partir de la base de datos, mediante consultas en lenguaje SQL, ver (**ANEXO C**) dentro de la plataforma del EVA, detallados estadísticamente en la siguiente tabla:

Fuente: Ruiz, F., (2014).

	Assignment (tareas)	Blog	Course (curso)	Forum (foros)	Quiz cuestionario	Resource recurso	Role	utpl	User (perfil)	Upload (subida)
TOTALES	668,00	4,00	4731,00	6151,00	581,00	589,00	70,00	70,00	887,00	104,00
MEDIA	17,58	0,11	124,50	161,87	15,29	15,50	1,84	1,84	23,34	2,74
MEDIANA	10,02	0,06	122,50	94,00	8,72	8,84	1,05	1,05	24,22	1,56
MAX.	668,00	4,00	4731,00	6151,00	581,00	589,00	70,00	70,00	23,34	104,00
MIN.	10,00	4,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00

Tabla 14. Estadísticas de interacción con las herramientas del EVA.

Según el análisis realizado anteriormente, se puede observar en la **Tabla 14**, los datos estadísticos del *promedio total de interacciones* que tienen los estudiantes en los cursos, los cuales permiten determinar las herramientas con las cuales los estudiantes y docentes usan para

su aprendizaje. Los *foros* con un promedio cualitativo de **alto**, es la herramienta con mayor nivel de aporte por parte de los estudiantes, seguido de ella, con un promedio de contribución cualitativo de **medio-alto** están las herramientas del *perfil de usuario* y *las tareas*; luego están los *recursos* y *cuestionarios* con un promedio de aporte **medio**. También están las herramientas de *subida de archivos*, *rol* y *utpl*, con un porcentaje menor de interacción y un promedio de **medio-bajo**; y finalmente un porcentaje casi nulo de interacción y un promedio cualitativo de **bajo** están el *blog* y *las notas*.

La media de la interacción del estudiante en el EVA son las actividades realizadas en los **cursos**, el cual se califica como **media-baja**, ya que no interactúan con todas las herramientas del EVA. La media indica que el mayor número de accesos en los cursos con **87,55%**, con la interacción más alta en **páginas de las Actividades** (*foros, cuestionarios, tareas*); con un porcentaje de **5,51%** en las **páginas de Apoyo** (*recursos, recursos educativos abiertos, utpl y upload*); con menor porcentaje **6,91%** de interacción en las **páginas de Información general** (*perfil de usuario, rol*); y Finalmente con **0,03%** de accesos interactúan en las **páginas de Colaboración y comunicación** (*chat, mensajería, foros, red social colaborativa, twitter, blog, etc.*).

Con la extracción de los datos se creará una entidad en la base de datos, con los indicadores obtenidos en el pre-proceso que se realizó, e implementar en ella los datos extraídos mediante consultas SQL, en la base de datos del EVA para el posterior análisis en Weka.

4.9. Datos de navegación de los estudiantes en el EVA.

Una vez obtenidos los datos en la base del EVA, se realizó las consultas SQL, para determinar las rutas de navegación de los estudiantes, las cuales permitirán hacer un análisis posterior para obtener los indicadores de navegación que tienen los estudiantes con las herramientas dentro de los cursos durante su navegación en el EVA. Estos datos permitirán discernir las herramientas a las cuales los estudiantes acceden y las acciones que realizan dentro de cada una, para con el resultado de esta información diseñar un modelo de estudiante con los indicadores obtenidos en este análisis. Las páginas más frecuentadas por los estudiantes cuando interactúan con las herramientas del EVA se detallaran más adelante en el *Capítulo: 'Minería de datos: Experimentación y Resultados'*, en la sección de la experimentación con el algoritmo de Árboles de decisión **J-48**, en el cual se describen las rutas de navegación de los estudiantes por medio de los accesos a las herramientas del EVA.

CAPITULO V:
MINERÍA DE DATOS:
EXPERIMENTO Y RESULTADOS

5. Minería de los datos: Experimentación y resultados

5.1. Introducción.

Para resolver la problemática planteada al inicio de la presente tesis, se pretende realizar minería de datos, para lo cual se ha considerado que, desde un determinado conjunto de datos ya sean estos de tipo nominal o reales, se pretende evaluar varias técnicas y algoritmos de inteligencia artificial, los cuales permitan realizar un aprendizaje automático para establecer patrones (enfocados a la navegación del usuario) y modelos sobre estos datos, y de esta manera poder extraer conclusiones sobre los mismos; este proceso de minería será desarrollado con la ayuda de la herramienta, *open source*, de minería de datos y aprendizaje automático *WEKA*, el mismo que dispone de un conjunto importante de varios algoritmos que permitirán descubrir patrones de navegación en el comportamiento de aprendizaje de los estudiantes en el EVA.

De acuerdo a los objetivos planteados para la presente tesis, en los cuales principalmente esta encontrar las rutas de navegación que los estudiantes tienen cuando interactúan con las herramientas disponibles en la plataforma de la UTPL (EVA), mediante técnicas básicas de aprendizaje automático, y extraer indicadores y atributos según patrones de navegación del estudiante mediante la aplicación de estas técnicas de minería, las cuales se detallaron anteriormente en la sección de "*Selección de herramientas y técnicas para la minería de datos*", y con los algoritmos de aprendizaje propuestos anteriormente para el presente análisis.

Para descubrir patrones de navegación en los datos de interacción de los estudiantes, una técnica muy eficaz en este tipo de análisis es la aplicación de la Identificación de Patrones Secuenciales en el proceso de extracción de los datos de navegación. Este proceso de patrones secuenciales permitirá aplicar técnicas de reglas de asociación secuenciales, reglas que expresen patrones de comportamiento secuencial en la navegación de los estudiantes que interactúan con el EVA, en distintos instantes del tiempo y de esta manera clasificar a los estudiantes que visitan las diferentes páginas del sistema en grupos con características similares según las rutas de navegación que tengan, es decir que permita agrupar estudiantes con una conducta similar y que tengan mismos intereses al momento de acceder al sistema.

5.2. Clasificación de páginas del EVA.

Para determinar las rutas de navegación de los estudiantes en la plataforma del EVA, se ha creído conveniente clasificarlas a las diferentes páginas del sistema según las actividades que realiza el estudiante dentro de los cursos para su aprendizaje y comunicación, para ello se las clasificara en cuatro tipos de dominios o categorías, en las cuales se encuentran las herramientas

y servicios disponibles en el sistema para la interacción de los estudiantes con su docente y demás compañeros, la misma que se detalla a continuación:

Fuente: Ruiz, F., (2014).

CATEGORÍA	SERVICIOS	ACCIONES
Páginas de Información General	<ul style="list-style-type: none"> • Perfil de usuario • Calendario • Library • Glossary • Etc. 	view, view all, update add, delete, edit. mailer. view, view all, view entry, add, add comment, add entry, delete entry, update, update entry.
Páginas de Actividades	<ul style="list-style-type: none"> • Foros • Recursos • Cuestionarios • Tareas 	view forum, view discusión, subscribe, add post, add discusión view, view all. view, view all, attempt, close attempt, continue attemp. view, view all, upload.
Páginas de Apoyo	<ul style="list-style-type: none"> • Recursos Educativos Abiertos • Subir archivos • utpl 	view, view all. upload role_assign
Páginas de Comunicación y Colaboración	<ul style="list-style-type: none"> • Twitter • Chat • Blogs • Mensajería • Red Social de Aprendizaje 	view all view, view all, talk, report, add, report, update, view, add, comment, delete, update add contact, block contact, history, remove contact, unblock contact, write Comentar, eliminar, bloquear, seguir.

Tabla 15. Tipos de páginas del EVA, con las herramientas y acciones.

En la **Tabla 15**, se detallan las categorías que se clasificaron para determinar las páginas por las cuales navega el estudiante cuando interactúan con las herramientas y servicios dentro del EVA, y de esta manera determinar cuáles son las preferencias e intereses de los estudiantes según las acciones que realice dentro de las páginas por las que navega, y de esta manera encontrar las rutas de navegación de los estudiantes por el sistema para su aprendizaje e interacción y comunicación con los demás estudiantes y tutor/profesor, y a partir de estos resultados obtenidos del proceso de minería los determinar indicadores y atributos para representarlos en un modelo de datos del usuario.

Luego de entender y obtener todos los datos necesarios para el pre-proceso de minería y el minado en sí, para cumplir con los objetivos propuestos en la presente tesis y las herramientas que nos permita realizar dicho proceso, el cual se detalla más adelante con cada una de las técnicas y algoritmos propuestos para obtener la información necesaria que nos ayuden a determinar los indicadores y presentar un pequeño modelo con los indicadores obtenidos en el minado de los datos.

5.3. Descripción del escenario.

Según la definición de las técnicas de minería ya mencionadas en el Capítulo III: “*Análisis y selección de herramientas*” se trabajará directamente desde la base de datos del EVA, ya que contiene los datos de interacción de los estudiantes con el EVA, y permitirán determinar las rutas de navegación de los estudiantes de noveno ciclo en el sistema, este proceso se lo realizará con **WEKA**, con tres de los algoritmos supervisados: (**Asociación**: *GSP* y *A-priori* y **Clustering**: *SimpleKMeans* y *EM*) y uno de los algoritmos no supervisados: **Clasificación**: *Árboles de decisión J48*, los cuales vienen incorporados dentro del conjunto de algoritmos de Weka.

Las materias seleccionadas para el experimento cuentan con un total de 132 usuarios, entre los cuales están estudiantes y docentes; de los cuales 6 son docentes, 91 son estudiantes de los diferentes cursos, y 35 están enrolados en más de una materia de noveno ciclo de Informática. El total de estudiantes enrolados en los cursos de noveno ciclo son de 132 estudiantes, de los cuales 91 son los estudiantes que interactuaron directamente con las herramientas del EVA (*foros, tareas, recursos, blog, cuestionarios, chat, REA, mensajería, fuentes de información, y demás*) en el sistema. Las páginas (*Informativas, Actividades, Colaboración o páginas de Apoyo*) en las cuales realizaron acciones dentro de los cursos en la plataforma son con los cuales se realizará el experimento, es decir con 91 estudiantes. Estos datos de experimentación fueron tomados del periodo académico **Abril 2013 – Agosto 2013**, desde el 08 de Abril del 2013 hasta el 17 de Agosto del mismo año.

En los datos extraídos de forma manual mediante consultas en el lenguaje SQL, se encontró que todos los estudiantes ingresaron al sistema, y la mayor parte de ellos realizaron alguna acción dentro de los cursos de “Noveno Ciclo”, entre ellas se destacan las siguientes actividades:

- El curso con mayor número de interacciones por parte de los estudiantes en las diferentes herramientas del EVA es el curso de “REDES Y SISTEMAS DISTRIBUIDOS”.
- Los estudiantes interactuaron en varias de las herramientas que dispone el EVA dentro de los cursos, tales como: foros, cuestionarios, tareas, recursos, subir archivos, perfil de usuario, utpl, blog, entre otras.
- Todos los estudiantes de noveno ciclo de informática ingresan a revisar dentro de los cursos en los que se encuentran enrolados a revisar las actividades o sugerencias realizadas por el docente tutor.
- En los seis cursos la mayor parte de los estudiantes interactúan con las mismas

herramientas como son (cursos, foros, cuestionarios, recursos, tareas, perfil de usuario, utpl, subir de archivos) y sólo uno de los seis cursos interactúa con el blog.

- Del total de 126 usuarios que están enrolados en los cursos de noveno ciclo a realizar el proceso de minería, 91 estudiantes interactúan con el EVA, por lo tanto ese número de estudiantes será la muestra para la minería.
- Con la validación de los resultados obtenidos, se determinará el modelo que proporcionan salidas adecuadas y/o con márgenes de error admisibles.
- La muestra de 126 estudiantes, es un valor justificativo para el proceso de minería, ya que los datos a evaluar del usuario son los eventos y acciones realizadas por los estudiantes dentro de las actividades del sistema. Hay 15206 datos de los 132 usuarios, entre estudiantes y docentes en el EVA.
- En el proceso de interacción de los estudiantes con las herramientas del sistema, y navegación por el EVA, los estudiantes acceden a revisar con mayor frecuencia los recursos y actividades subidos al entorno por el tutor responsable de cada uno de los cursos.
- Durante el ingreso al EVA, los estudiantes ingresan a revisar dentro de los cursos, actividades tales como los recursos, tareas y foros, antes de realizar alguna acción sobre el mismo.
- Cuando tienen que responder los cuestionario subidos por el tutor, los estudiantes realizan más de un intento para el desarrollo de los mismos.
- Casi todos los estudiantes participaron de los foros y cuestionarios propuestos en el curso por cada uno de los tutores responsables.
- La mayor parte de los estudiantes participan en las discusiones dentro de los foros propuestos por el docente tutor del curso.
- La mayor parte de los estudiantes interactúan sólo con aquellas herramientas que los profesores consideran para el aprendizaje de los estudiantes dentro de los cursos.
- Un porcentaje mínimo de estudiantes ingresan una sola vez a revisar las actividades propuestas por el docente tutor pero no realizan ninguna actividad dentro del EVA.

5.4. Descripción del proceso de experimentación.

Con los datos detallados anteriormente para el análisis de minería de los datos se realizarán varios experimentos con cada uno de los algoritmos propuestos anteriormente para determinar un mejor ajuste en los datos y con ello obtener los mejores resultados para el posterior análisis de cada uno de ellos.

Los resultados obtenidos serán interpretados para determinar los indicadores y atributos que tienen los estudiantes según su comportamiento cuando interactúan con las herramientas y servicios dentro del EVA, para representarlos en un modelo de datos de usuario. El proceso a realizar se muestra en la **Fig. 6**.

Una vez obtenidos los datos de entrada, se realizará el proceso de minado, para ello dentro de la herramienta de *mysql-connector-java-5.1.17* se realizó la conexión con la base de datos desde Weka y en ella se realizaron las consultas presentadas en el (**ANEXO C**), las cuales se realizaron para obtener la información a la cual se le realizará el proceso de minería, además se realizarán filtros de discretización en los algoritmos de Asociación ya que se desea transformar los atributos numéricos seleccionados en atributos simbólicos; luego de tener los datos procesados y listos para el proceso de minería, se procederá a realizar el proceso de experimentación con cada uno de los algoritmos propuestos y obtener cada uno de los resultados para el proceso de evaluación de los mismos, y finalmente a partir de los resultados obtenidos, determinar indicadores y atributos de la navegación de los estudiantes y la interacción que tienen con las diferentes herramientas y servicios dentro de la plataforma, para representarlos en un modelo de datos del usuario, mediante la aplicación de una metodología, para lo cual se consideró el modelo Entidad-Relación.

Como se puede apreciar en la **Fig. 6**, la estructura del proceso de minería que se realizará con los datos obtenidos en el análisis previo realizado en cada una de los capítulos antes mencionados y en cada una de las fases de minería.



Fig. 6 Estructura del proceso de experimentación

Fuente: Ruiz, F., (2014).

A continuación se detallan los algoritmos seleccionados para el proceso de minería de los datos:

1. Las **reglas de asociación** permiten encontrar asociaciones interesantes en forma de relaciones de implicación entre los valores de los atributos de los objetos de un conjunto de datos.

- **A priori.-** Se trabajará con este algoritmo para el proceso de minado ya que se pretende descubrir patrones de navegación considerando las funciones de semejanza entre ellos y las acciones más frecuentes dentro de las rutas que tienen los estudiantes cuando interactúan con el sistema, y a partir de ellos encontrar mediante reglas de asociación las páginas a las cuales acceden los estudiantes con mayor frecuencia y las herramientas dentro de la plataforma en las cuales existe mayor número de ocurrencias.
- **GSP (Generalized Sequential Patterns).-** Se trabajará con este algoritmo ya que permite contabilizar las ocurrencias de todos los elementos dentro de los datos de entrada y filtrar o eliminar las partidas que tengan mucha frecuencia por parte de los estudiantes dentro del sistema; y de esta manera presenta el resultado sólo con los elementos frecuentes de interacción de los estudiantes dentro del sistema que originalmente contenía. *SGP* pretende descubrir patrones secuenciales dentro del conjunto de datos de navegación de los estudiantes; ya que es uno de los algoritmos que generan mejores resultados en cuanto a modelos predictivos para el descubrimiento de patrones y relaciones entre los datos que se usan para determinar las predicciones en el modelo.

2. Los **algoritmos de clustering** permitirán clasificar dentro de un conjunto de datos un determinado número de grupos con semejanzas y diferencias existentes entre los componentes del conjunto de entrenamiento.

- **SimpleKMeans.-** Se trabajará con este algoritmo para la minería de datos, ya que permite clasificar dentro de un conjunto de elementos un determinado número de grupos basándose en las semejanzas y diferencias existentes entre los componentes de determinada muestra. Es por ello, que a partir de los datos de interacción de los estudiantes con las herramientas del sistema, se busca clasificar a los estudiantes en grupos según las características de comportamiento similares que existan en el conjunto de entrenamiento, para luego con los resultados obtenidos determinar patrones de navegación de los estudiantes que interactúan con las herramientas del EVA.
- **EM.-** Se trabajará con este algoritmo ya que permitirá mediante un conjunto de clústeres agrupar aquellos estudiantes que realizan acciones similares cuando interactúan dentro del sistema, para a partir de esta información seleccionar las herramientas con las cuales interactúan y aprenden más los estudiantes cuando navegan por el sistema.

3. Los **algoritmos de clasificación** que predicen una o más variables discretas, se basan en otros atributos del conjunto de datos. El algoritmo de clasificación que se aplicará en este estudio será un algoritmo de árboles de decisión, del conjunto de algoritmos de Weka.

- **J 48.-** Se trabajará con este algoritmo de clasificación, ya que es uno de los más utilizados para este tipo de minería. Este algoritmo permite podar las acciones con poca frecuencia en los estudiantes, cuanto más baja sea esta probabilidad, se exigirá que la diferencia en los errores de predicción antes y después de podar sea más significativa para no podar.

Los algoritmos mencionados en la **Tabla 7**, se consideraron de acuerdo a resultados obtenidos con experiencias de trabajos similares a los que se está realizando en la presente tesis. Además de presentar resultados eficientes y ser muy usados para diferentes tipo de análisis.

5.5. Obtención de datos de navegación a partir de los datos del usuario.

Una vez realizado el análisis de las entidades que almacenan la información de las actividades y eventos generados por los estudiantes durante su navegan por el sistema, se obtuvieron los datos de interacción de todos los estudiantes, los cuales fueron realizados mediante consultas en lenguaje SQL dentro de la base de datos con la ayuda de la herramienta Navicat for MySQL, la cual permite determinar las acciones realizadas por el estudiante cuando interactúa con el EVA.

Fuente: Ruiz, F., (2014).

<i>DATOS DE INTERACCIÓN DE LOS ESTUDIANTES</i>
<i>Páginas en las que el estudiante inicia sesión de usuario dentro del EVA.</i>
<i>Páginas visitadas por el estudiante en cada uno de los cursos dentro del sistema</i>
<i>Páginas visitadas por el estudiante en el perfil del usuario dentro del sistema</i>
<i>Páginas visitadas en las que el estudiante realiza vistas de los recursos del curso.</i>
<i>Páginas en las que el estudiante ingresa a revisar los recursos disponibles en el curso.</i>
<i>Páginas donde los estudiantes ha visitado los foros propuestos por el tutor en los curso.</i>
<i>Páginas en las que el estudiante se ha suscripto en los foros del curso</i>
<i>Páginas en las que el estudiante se ha quitado la suscripción en el curso</i>
<i>Páginas en las que el estudiante ha iniciado una conversación dentro de los foros.</i>
<i>Páginas en las que el estudiante da respuesta en los debates propuesto dentro del foro</i>
<i>Páginas donde el estudiante ha agregado un debate en los foros dentro del curso.</i>
<i>Páginas en las que el estudiante ha borrado un debate dentro de los foros del curso.</i>
<i>Páginas en las que el estudiante ha adicionado una nota dentro de los foros</i>
<i>Páginas en las que el estudiante ha actualizado las notas dentro de los foros</i>
<i>Páginas donde el estudiante postean un mensaje en el twitter en los curso en el EVA.</i>
<i>Páginas donde el estudiante da respuestas en los foros en los que ha ingresado</i>
<i>Páginas en las que el estudiante revisa las tareas propuestas por el tutor en el curso</i>
<i>Páginas en las que el estudiante sube las tareas en el EVA en los cursos.</i>
<i>Páginas en las que el estudiante actualiza las tareas propuestas por el tutor</i>
<i>Páginas en las que el estudiante ingresa al chat para enviar un mensaje en el EVA.</i>
<i>Páginas en las que el estudiante ingresa a revisar los cuestionarios.</i>
<i>Páginas en las que el estudiante realiza un intento en los cuestionarios.</i>
<i>Páginas en las que el estudiante cierra el cuestionario que realizó.</i>
<i>Páginas en las que el estudiante ingresa a la mensajería dentro del EVA.</i>
<i>Páginas en las que el estudiante ingresa al blog dentro del sistema.</i>
<i>Páginas en las que el estudiante ingresa a revisar los blogs.</i>
<i>Páginas en las que el estudiante accedió a las páginas referentes a Faqs en el sistema.</i>
<i>Páginas en las que el estudiante realizó accesos al calendario dentro del sistema.</i>
<i>Páginas en las que el estudiante ingresa a información general dentro del EVA.</i>

Tabla 16. Datos de navegación de los estudiantes.

Este procesamiento de datos se lo realizó en las tablas *mdl_log*, *mdl_user*, *mdl_course* y *mdl_course_category* almacenados en el repositorio del EVA. Los campos se extraerán a partir de las tablas ya mencionadas del usuario, y las acciones se detallan más adelante, las mismas que describe cada uno de los datos de interacción de los estudiantes con las herramientas del EVA.

Los datos mencionados en la **Tabla 16**, son los datos que permitirán determinar la identificación de indicadores y atributos en las rutas de navegación e interacción de los estudiantes dentro del sistema, luego de haber realizado la aplicación de técnicas de minería mencionadas en la parte de análisis del EVA, con la finalidad de según la interpretación de los resultados de cada uno de

los algoritmos con los que se experimentó, representarlo en un modelo de datos con indicadores y atributos obtenidos del análisis.

Las consultas en el lenguaje SQL de los datos de interacción de los estudiantes con el sistema se pueden apreciar en el (**ANEXO D**), el cual describe cada uno de los datos mencionados en la **Tabla 20**, realizadas en la base de datos del EVA.

La técnica utilizada permitirá encontrar las rutas de navegación de los estudiantes que interactúan con las herramientas de los cursos en los que se encuentran enrolados, las variables que fueron consideradas para el presente análisis son: la interacción con los recursos, tareas, foros y cuestionarios; nivel de interacción con las herramientas, rutas de acceso y acciones realizadas dentro de cada una de las herramientas de los cursos, nivel de proceso de aprendizaje del estudiante al interactuar con estas herramientas.

5.6. Aplicación de los algoritmos seleccionados.

Para la obtención de resultados de la minería aplicada a los datos de entrada y previa al diseño del modelo de datos con los indicadores y atributos encontrados después de la minería, se consideró conveniente aplicar como se mencionó anteriormente los algoritmos de (*Asociación, Clustering y Clasificación*), los cuales se los aplicara a los datos seleccionado dentro de la base de datos del EVA, con las interacciones de los estudiantes dentro de las herramientas y servicios a los cuales accede con mayor frecuencia cuando navega en la plataforma, una vez obtenidos los resultados de cada una de las experimentaciones con los algoritmos se realizará una tabla comparativa con los resultados de los diferentes algoritmos propuesto, y a partir de estos realizar un modelo con indicadores y atributos encontrados del análisis.

Las páginas de la plataforma del EVA a las cuales acceden los estudiantes de Informática de la UTPL Modalidad Abierta, se clasificaron en cuatro categorías: **Actividades, Información, Apoyo y Colaboración** sobre los cuales los estudiantes navegaron por **15206** páginas del EVA, a las cuales se les clasificó su accesos de acuerdo a las herramientas y servicios a los que ingresaron a realizar alguna acción, las páginas de tipo *Informativas* son (Perfil de usuario, introducción, calendario, etc.), son de tipo *Actividades* (tareas, cuestionarios, foros, etc.); las páginas de *Apoyo* (recursos, REA, links, utpl, perfil de usuario) y por último las páginas de tipo *Comunicación y Colaboración* (twitter, chat, mensajería, blog, etc.); las cuales son páginas con las que el estudiante interactúa dentro de la plataforma.

5.6.1. Experimentación con la técnica de Asociación.

Estos algoritmos permiten encontrar patrones de comportamiento en los estudiantes frente al sistema cuando interactúan con el mismo, se creyó necesario antes de realizar la minería buscar algoritmos que permitan implementar los indicadores seleccionados, como ya se lo considero en el estado del arte, que se trabajaría con la herramienta de WEKA (cual cuenta con un amplio conjunto de algoritmos que sirven para la generación de indicadores), la técnica empleada será la de Asociación y los algoritmos seleccionados fueron el GSP y A priori, los mismos que permite encontrar patrones en los datos ingresados.

En las Reglas de asociación el interés de una regla de asociación está dada por su soporte, que determina la frecuencia de aparición en la colección de la combinación de productos involucrados en dicha regla, y además de otro valor importante que es la confianza, la cual, determina cuánto representa el soporte de la regla, del soporte del antecedente de la regla; y de acuerdo a lo se puede decir que una regla es interesante si su soporte y confianza son mayores o iguales que ciertos umbrales de mínimo soporte y confianza especificados. Según el análisis que ya se realizó con anticipación y fueron descritos en el Estado del Arte, de cada una de las técnicas de las que dispone E-Learning, se consideró como la mejor técnica según las necesidades del presente estudio, la técnica de **GSP (Generalized Sequential Patterns)**, que permite encontrar patrones de navegación en los datos ingresados de los estudiantes de acuerdo al nivel de interacción cuando navega en el EVA.

Se consideró esta técnica ya que las técnicas de clasificación permiten desarrollar un perfil para usuarios que acceden a páginas o archivos particulares, en este caso serán analizadas las rutas de navegación del estudiante con las herramientas del sistema.

5.6.1.1. Aplicación del algoritmo GSP (*Generalized Sequential Patterns*).

El análisis de los datos con **GSP**, se lo realizó con varios experimentos, el primer experimento se lo realizó con los parámetros por defecto del algoritmo, con **S=0.9** (*valor de semilla*), **F=1** (*número de filtros en los atributos*) y **I=0** (*número o la longitud de las secuencias*). En el segundo experimento sólo se modificará el valor de **F=1**, y en el tercer experimento con **F=2**. La razón por la cual se realizó varios experimentos con los datos de entrada es con la finalidad de encontrar resultados más precisos en las secuencias más frecuentes en la navegación de los estudiantes en el EVA.

Mediante la implementación de este algoritmo lo que se busca es encontrar dentro de un conjunto de entrenamiento completo, las secuencias de navegación más frecuentadas por los

estudiantes; y a partir de estos resultados obtener indicadores de navegación y patrones que tienen los estudiantes cuando interactúan con las Actividades del EVA.

```
==== Run information ====
Scheme:      weka.associations.GeneralizedSequentialPatterns -S 0.9 -I 0 -F -1
Relation:    Cursos Noveno Ciclo-weka.filters.unsupervised.attribute.Remove-R1
Instances:   13874
Attributes:  4
             CURSO
             ACTIVIDAD
             ACCION
             PAGINA INGRESADA
==== Associator model (full training set) ====

GeneralizedSequentialPatterns
=====

Number of cycles performed: 4
Total number of frequent sequences: 12

Frequent Sequences Details (filtered):

- 1-sequences
[1] <{course}> (6)
[2] <{add mod}> (6)

- 2-sequences
[1] <{course}{course}> (6)
[2] <{course}{add mod}> (6)
[3] <{add mod}{course}> (6)
[4] <{course,add mod}> (6)
[5] <{add mod}{add mod}> (6)

- 3-sequences
[1] <{course}{course,add mod}> (6)
[2] <{add mod}{course,add mod}> (6)
[3] <{course,add mod}{course}> (6)
[4] <{course,add mod}{add mod}> (6)

- 4-sequences
[1] <{course,add mod}{course,add mod}> (6)
```

Fig. 7. Resultados de los datos con el Algoritmo GSP para la minería.

Fuente: Ruiz, F., (2014).

Una vez realizado el proceso de experimentación de los datos con la implementación del algoritmo *GSP*, se puede apreciar en la **Fig. 7**, una lista ordenada de secuencias (*ítem-sets*) optimizadas por el algoritmo, la cual presenta las interacciones con las herramientas del sistema.

Resultados obtenidos de la experiencia.

Los resultados más óptimos con los datos ingresados en el algoritmo evaluador fueron los del primer experimento, con los parámetros por defecto que optimiza el propio algoritmo. Al analizar el resultado mostrado en la **Fig. 7.**, la cual presenta una lista ordenada de ítem-sets que vendrían a ser las secuencias de navegación de los estudiantes en la plataforma (**S=4**), la longitud de las secuencias se las determina con el número de elementos o ítem-sets que pertenecen a determinada secuencia, que en este caso es (**I=0**), y uno de los cuatro estados representan las acciones realizadas por los estudiantes dentro de los cursos del EVA, es decir, un ítem de la secuencia es una actividad a la cual accedió el estudiante para realizar algún tipo de acción sobre la misma, y una secuencia es el total de páginas Web visitadas por el estudiante cuando navega por el sistema en determinado tiempo.

Según los resultados de la aplicación del algoritmo GSP, los datos resultantes de la secuencia frecuente de acceso dentro de la navegación de los estudiantes en el sistema muestran el ingreso frecuente a los cursos en los que se encuentran enrolados dentro del sistema.

5.6.1.2. Aplicación con el algoritmo A-priori.

Los algoritmos de asociación son una técnica de minería muy importante, ya que permiten la búsqueda automática de reglas que relacionan conjuntos de atributos entre sí, y encontrar asociaciones interesantes en forma de relaciones de implicación entre los valores de los atributos de los objetos de un determinado conjunto de datos. Estos algoritmos no supervisados, en el sentido de que no existe ningún tipo de relaciones conocidas a priori con las que contrastar la validez de los resultados, sino que se evalúa si esas reglas son estadísticamente significativas.

El principal algoritmo de asociación implementado en el presente trabajo es el algoritmo "A priori". Este algoritmo únicamente puede buscar reglas entre atributos simbólicos, razón por la que primeramente se aplicó en el pre-proceso de los datos un filtro '*Discretize*' no supervisado sobre los atributos de los datos, ya que en ellos las operaciones son independientes del análisis del algoritmo. La función del filtro es transformar los atributos numéricos seleccionados en atributos simbólicos, con una gran variedad de etiquetas que resultan de dividir la amplitud total del atributo en intervalos.

Se realizaron tres experimentos con el algoritmo A priori, el primer experimento se lo realizó con los parámetros por defecto del modelo $N = 10$, (*número de reglas*) $C = 0.9$, (*métrica mínima aceptable*) $M = 0.1$ (*límite inferior de apoyo mínimo*) y $S = -1.0$ (*valor del límite superior de apoyo mínimo*); el segundo experimento se modificó el valor del límite inferior de apoyo mínimo, quedando ($N = 10$, $C = 0.9$, $M = 0.0$ y $S = -1.0$); y el tercer experimento se modificó el umbral mínimo aceptable para una regla de 90% a 50%.

El número de reglas obtenidas por el modelo depende directamente del valor que se le asigne al parámetro *número de reglas* y de los datos de entrada que se tiene, según estos dos datos el total de reglas pueden ser pequeñas o enormes.

```

15:48:49 - Apriori
=== Run information ===
Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    Cursos Noveno Ciclo-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last
Instances:   13874
Attributes:  5
              USUARIO
              CURSO
              ACTIVIDAD
              ACCION
              PAGINA INGRESADA
=== Associator model (full training set) ===

Apriori
=====
Minimum support: 0.1 (1387 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 18

Generated sets of large itemsets:
Size of set of large itemsets L(1): 11
Size of set of large itemsets L(2): 12
Size of set of large itemsets L(3): 1

Best rules found:
1. ACCION=view forum 2606 ==> ACTIVIDAD=forum 2606   conf:(1)
2. ACCION=view discussion 2231 ==> ACTIVIDAD=forum 2231   conf:(1)
3. USUARIO='(-inf-7715.9]' ACTIVIDAD=course 2477 ==> ACCION=view 2458   conf:(0.99)
4. ACTIVIDAD=course 4731 ==> ACCION=view 4569   conf:(0.97)

```

Fig. 8 Experimentación con el Algoritmo A priori.

Fuente: Ruiz, F., (2014).

La **Fig. 8**, muestra el resultado de la ejecución de A priori con los parámetros por defecto de entrada del algoritmo, se puede apreciar en los resultados obtenidos que las reglas que aparecen aportan poca información. Se puede observar relaciones triviales entre asignaturas y acciones, así como un rango insignificante de estudiantes que interactúan en los cursos. Se realizaron algunos experimentos variando los valores de soporte y confianzas y se puede ver que mientras más bajos son los valores, menos significativas son las reglas o sin relevancia. Los valores de las reglas varían según el valor de **M**, mientras más alto es el valor de M en este caso 0.3 menor es el número de reglas.

Ya que las reglas obtenidas no presentan información significativa, se modificó el valor de M a un valor más bajo de 0.0 y los demás valores los mismos del modelo. Como podemos observar en la **Fig. 9.**, las reglas que presenta el algoritmo aportan mayor información significativa: se puede apreciar que un valor de confianza del 100% los estudiantes del curso de Redes y Sistemas Distribuidos ingresan a revisar las actividades dentro del mismo, e interactúan con la herramienta de los foros, a la cual ingresan a revisar las discusiones que hay en el curso y a revisar los foros. Al aplicar este procedimiento el resultado de los experimentos no varió mucho en los resultados, sino que presentó más reglas con información significativa.

```

15:36:19 - Apriori
=== Run information ===
Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 2.0 -M 0.0 -S -1.0 -c -1
Relation:    Cursos Noveno Ciclo-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last
Instances:   13874
Attributes:  5
            USUARIO
            CURSO
            ACTIVIDAD
            ACCION
            PAGINA INGRESADA
=== Associator model (full training set) ===

Apriori
=====
Minimum support: 0.05 (695 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 39

Generated sets of large itemsets:
Size of set of large itemsets L(1): 23
Size of set of large itemsets L(2): 46
Size of set of large itemsets L(3): 25
Size of set of large itemsets L(4): 4

Best rules found:
1. ACCION=view forum 2606 ==> ACTIVIDAD=forum 2606   conf:(1)
2. ACCION=view discussion 2231 ==> ACTIVIDAD=forum 2231   conf:(1)
3. USUARIO='(-inf-7715.9]' ACCION=view forum 1321 ==> ACTIVIDAD=forum 1321   conf:(1)
4. USUARIO='(-inf-7715.9]' ACCION=view discussion 1293 ==> ACTIVIDAD=forum 1293   conf:(1)
5. CURSO=Redes y Sistemas Distribuidos ACCION=view forum 1232 ==> ACTIVIDAD=forum 1232   conf:(1)
6. CURSO=Redes y Sistemas Distribuidos ACCION=view discussion 1231 ==> ACTIVIDAD=forum 1231   conf:(1)
7. PAGINA INGRESADA=view.php?id=51755 1141 ==> CURSO=Redes y Sistemas Distribuidos 1141   conf:(1)
8. PAGINA INGRESADA=view.php?id=51755 1141 ==> ACTIVIDAD=course 1141   conf:(1)
9. PAGINA INGRESADA=view.php?id=51755 1141 ==> ACCION=view 1141   conf:(1)
10. ACTIVIDAD=course PAGINA INGRESADA=view.php?id=51755 1141 ==> CURSO=Redes y Sistemas Distribuidos 1141   conf:(1)

```

Fig. 9 Resultados de los datos con el Algoritmo A priori

Fuente: Ruiz, F., (2014).

En el tercer experimento nos produce relaciones bastante evidentes, y con reglas igualmente significativas entre curso, actividad y la acción que realizan los estudiantes. Al ejecutar el algoritmo con los parámetros por defecto sólo nos presenta una regla; Esto se debe a que el modelo pone como umbral mínimo aceptable para una regla el 90%, si se pone un parámetro menor al del modelo de $C = 50\%$ y los demás valores con los mismos del modelo por defecto, y nos arroja el siguiente resultado:

```

16:05:46 - Apriori
=== Run information ===
Scheme:      weka.associations.Apriori -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:    Cursos Noveno Ciclo-weka.filters.unsupervised.attribute.Discretize-B10-M-1.0-Rfirst-last
Instances:   13874
Attributes:  5
            USUARIO
            CURSO
            ACTIVIDAD
            ACCION
            PAGINA INGRESADA
=== Associator model (full training set) ===

Apriori
=====
Minimum support: 0.15 (2081 instances)
Minimum metric <confidence>: 0.5
Number of cycles performed: 17

Generated sets of large itemsets:
Size of set of large itemsets L(1): 9
Size of set of large itemsets L(2): 7
Size of set of large itemsets L(3): 1

Best rules found:
1. ACCION=view forum 2606 ==> ACTIVIDAD=forum 2606   conf:(1)
2. ACCION=view discussion 2231 ==> ACTIVIDAD=forum 2231   conf:(1)
3. USUARIO='(-inf-7715.9]' ACTIVIDAD=course 2477 ==> ACCION=view 2458   conf:(0.99)
4. ACTIVIDAD=course 4731 ==> ACCION=view 4569   conf:(0.97)
5. USUARIO='(-inf-7715.9]' ACCION=view 3200 ==> ACTIVIDAD=course 2458   conf:(0.77)
6. ACCION=view 6016 ==> ACTIVIDAD=course 4569   conf:(0.76)
7. CURSO=Redes y Sistemas Distribuidos 5073 ==> ACTIVIDAD=forum 3002   conf:(0.59)
8. ACTIVIDAD=forum 6151 ==> USUARIO='(-inf-7715.9]' 3404   conf:(0.55)
9. ACTIVIDAD=course ACCION=view 4569 ==> USUARIO='(-inf-7715.9]' 2458   conf:(0.54)
10. ACCION=view 6016 ==> USUARIO='(-inf-7715.9]' 3200   conf:(0.53)

```

Fig. 10 Experimentación con el Algoritmo A priori.

Fuente: Ruiz, F., (2014).

La tarea de asociación permite descubrir patrones en forma de reglas, los cuales muestran las interacciones que ocurren con mayor frecuencia en el conjunto de datos de entrada. Como se puede ver el algoritmo busca relaciones entre los atributos de los datos para descubrir reglas de asociación expresadas en el soporte y confianza del modelo. La **Fig. 10** muestra que hay reglas poco interesantes (regla 5), reglas redundantes (la regla 3, que es igual a la regla 4 y 5), también hay reglas similares, (reglas con el mismo elemento en el antecedente y el consecuente, pero intercambiados, tales como las reglas 4, 9, y las reglas 1 y 2); también hay reglas que pueden ser muy importantes para los propósitos educativos, es decir, para saber las herramientas con la que a los estudiantes más les gusta trabajar, o son más aptas para su aprendizaje, como son las reglas que se ajustan o tienen mayor nivel de interacción y visitas (si un estudiante interactúa bastante con los foros, puede ser porque expresando sus conocimientos o leyendo el aporte de otros estudiantes aprende de mejor manera, como la regla 1, 2 y algunas reglas similares).

Según los resultados obtenidos con el algoritmo A-priori, el modelo busca reglas entre atributos simbólicos en varios intervalos de igual frecuencia. La necesidad de aplicar este tipo de algoritmos es que ayuda a los datos de entrada relacionarlos entre atributos, y así obtener datos más precisos en cuanto a la interacción de los estudiantes con el EVA, lo cual no se lo podría hacer mediante consultas SQL directamente en la base del EVA, ya que no dispone de estas opciones adicionales que presenta el modelo A-priori.

Los resultados con mayor significatividad en las relaciones encontradas se pueden visualizar de forma gráfica en la **Fig. 9**, ya que presentan las reglas más significativas de todas las posibles por la precisión o confianza, dada por el porcentaje de veces que instancias que cumplen el antecedente cumplen el consecuente, y el segundo es el soporte, dado por el número de instancias sobre las que es aplicable la regla. En todo caso, son reglas de muy baja precisión y que habría que considerar simplemente como ciertas tendencias.

Resultados obtenidos de la experiencia

Como se puede apreciar en la **Fig. 9**, el resultado del algoritmo presenta 10 reglas descubiertas. Según los resultados obtenidos de la implementación de *A priori* se puede concluir que el modelo tiene la ventaja de ser de fácil utilización para el usuario, ya que sólo requiere como parámetro inicial el número de reglas que se desee obtener. A su vez este algoritmo posee varias desventajas para este tipo de análisis, y es que en los resultados el modelo dispara gran cantidad de datos redundantes, es decir, reglas con la misma información en reglas con bajo soporte, además, el tiempo de ejecución es muy alto.

A partir de la información detallada anteriormente, el profesor puede tomar decisiones con respecto al manejo de las herramientas y prestar más atención a las herramientas y actividades con las que los estudiantes interactúan y trabajan sin problemas de aprendizaje.

5.6.2. Experimentación con la técnica de Clustering

Con la finalidad de encontrar patrones de comportamiento en los estudiantes frente al sistema, se vio necesario antes de realizar la minería buscar algoritmos que permitan implementar los indicadores seleccionados con *WEKA* (la cual cuenta con un amplio conjunto de algoritmos que sirven para la generación de indicadores).

Los algoritmos empleados dentro de la técnica de clustering serán el ***SimpleKMeans*** y ***EM***, el primero por ser uno de los algoritmos más veloces y eficientes y el segundo por ser estadístico y es bastante más elaborado que el anterior, por lo cual, permitirán encontrar las herramientas/actividades con las cuales más interactúan los estudiantes durante su navegación por el EVA, en grupos separados de instancias con características o comportamiento similares o en común que tengan los estudiantes, según el criterio de comparación entre los valores de atributos de las instancias que se defina en el algoritmo.

5.6.2.1. Aplicación del algoritmo SimpleKMeans

En el análisis con el algoritmo *SimpleKMeans* se realizó tres experimentos con diferentes valores en los parámetros del modelo. El primer experimento se lo realizó con los parámetros que tiene por defecto el algoritmo, centros iniciales del clúster (***N=2***), número máximo de interacciones (***I=500***) y la *semilla* (***S=10***), y el modo de evaluación sobre el mismo conjunto de entrenamiento, es decir, sobre los mismos datos (*training set*). Los otros dos experimentos se los realizó con diferentes parámetros a los del modelo, la semilla ***S=100***, y el valor de las iteraciones ***I=200***, y el número de clústeres se lo configuró en 5 clústeres, los demás parámetros son los que tiene por defecto el modelo, y el modo de evaluación se la realizará para que el modelo los clasifique a los estudiantes por la interacción en el atributo de los cursos.

La razón por la cual se realizó varios experimentos con el valor de la semilla es con el objetivo de encontrar el error mínimo en la *suma de los cuadrados* en cada uno de los resultados, y el *mínimo valor de clústeres mal clasificados*, y con ello, determinar el mejor resultado de los tres experimentos.

Fuente: Ruiz, F., (2014).

Evaluación de las herramientas									
Algoritmos	SimpleKMeans (S=500)			SimpleKMeans (S=1000)			SimpleKMeans (S=2000)		
	Clúster	0	1	2	0	1	2	0	1
# Iteraciones	5			5			6		
# Instancias	4127	1501	2075	1978	4326	1399	3485	1936	2282
Porcentaje	54 %	19 %	27 %	26 %	56 %	18 %	45 %	25 %	30 %
Error	9185			9668			9414		
Clústers mal clasificados	3944.0 51.2 %			4009.0 52.0 %			3426.0 44.5 %		
Herramienta	C0 = Curso (course) C1 = Foro (forum) C2 = Recursos (resource)			C0 = Cuestionario (quiz) C1 = Curso (course) C2 = Foro (forum)			C0 = Curso (course) C1 = Cuestionario (quiz) C2 = Foro (forum)		

Tabla 17. Análisis de los resultados con el algoritmo SimpleKMeans

La **Tabla 17**, muestra los resultados obtenidos con el modelo *SimpleKMeans*. Para determinar la agrupación con los mejores resultados se consideró el valor del número de instancias mal clasificadas y el valor de error de la suma de los cuadrados, el experimento con semilla ($S=2000$) tiene los resultados más eficientes, debido a que los clústeres son los mejor definidos y más claros, y con un valor medio-bajo en las instancias clasificadas incorrectamente con **44.5%**, y presenta el valor de error de la suma de los cuadrados medio (*9414*).

Los valores de semilla se pusieron más altos, debido a que los datos de semilla con valores de **(10, 50 y 100)** presentan un valor muy alto en el error de la suma de los cuadrados y en el número de clústeres mal clasificados, por ello se realizaron experimentos con valores altos para determinar si existe un decremento en los datos considerados en la experimentación.

```

Cluster output
-----
kMeans
=====
Number of iterations: 6
Within cluster sum of squared errors: 9414.374062676012
Missing values globally replaced with mean/mode

Cluster centroids:
Attribute                               Full Data          Cluster#
(7703)                                (3485)              0              1              2
-----
USUARIO                                17440.653           12513.3022          34326.905           10639.6161
CURSO                                   28992.9085           28993.7776          28993.6674          28990.9373
ACCION                                  view                 view                 view                 view forum
PAGINA VISITADA                         view.php?id=28990   view.php?id=28990   view.php?id=28991   index.php?id=28985

Clustered Instances
0  3485 ( 45%)
1  1936 ( 25%)
2  2282 ( 30%)

Class attribute: HERRAMIENTA
Classes to Clusters:
  0  1  2  <-- assigned to cluster
111 145 107 assignment
  3  0  0 blog
2395 1074 1 course
328 202 1710 forum
182 172 170 quiz
302 155 34 resource
  0  4  15 role
  0  1  0 twitter
 14  36  40 upload
150 143 190 user
  0  4  15 utpl

Cluster 0 <-- course
Cluster 1 <-- quiz
Cluster 2 <-- forum

Incorrectly clustered instances :      3426.0  44.4762 %

```

Fig. 9. Resultados con el algoritmo SimpleKMeans con S=2000.

Fuente: Ruiz, F., (2014).

En la **Fig. 9**, se pueden apreciar los resultados obtenidos de la clasificación de los tres clústeres con los grupos evaluados según la interacción con las herramientas del EVA.

En el proceso de agrupación que realiza el algoritmo *SimpleKMeans* se puede filtrar dependiendo de la información que se desee obtener, es decir, en el modelo se pueden filtrar los datos según los atributos de los datos de entrenamiento, en este caso se escogió el valor nominal de **herramientas** para que de esta manera clasifique a los estudiantes según la interacción que tiene con las herramientas dentro del EVA, y se detalla de la siguiente forma:

Clúster 0 herramienta más frecuentada <-- **curso** con (3485) accesos, correspondiente al 45%

Clúster 1 herramienta menos frecuentada que la anterior <-- **quiz** con (1936) accesos, correspondiente al 25%

Clúster 2 herramienta más frecuentada que la anterior <-- **foros** con (2282) accesos, correspondiente al 30%

Clúster 0. Se caracteriza por agrupar a los estudiantes con una mayor frecuencia de acceso a las páginas dentro de los **Cursos**, en la cual navegan para revisar las actividades propuestas por el profesor, la herramienta que presenta un mayor nivel de interacción por parte de los estudiantes es en los **foros**, **recursos** y **tareas** en el orden respectivo; luego con un nivel de interacción menor al anterior están los **cuestionarios** y el **perfil de usuario (user)**; y finalmente con un valor muy bajo los estudiantes interactúan con las herramientas de **subida de archivos** y **el blog** y un valor nulo se clasificaron las herramientas de **rol**, **twitter** y **utpl** en las cuales no existe actividad por parte de los estudiantes.

Los estudiantes interactúan más con las páginas de las *Actividades* (foros, cuestionarios, tareas), y las de *Apoyo* (recursos, upload) y páginas de *información general* (perfil de usuario) y con menor interacción las páginas de *Colaboración* (blog). Los *foros* y los *recursos*, tienen un nivel **medio-alto** de interacción. Los *cuestionarios*, *perfil de usuario* y *tareas* tienen un nivel de interacción **medio**. La subida de archivos y el blog tienen un nivel **bajo** de aporte, y con un nivel nulo de interacción están el *rol*, *twitter* y *utpl*, que no existe aporte por los estudiantes.

Clúster 1 En este grupo se agruparon los estudiantes que tienen una mayor frecuencia de acceso en las páginas dentro de los **cursos**, es decir ingresan a revisar las actividades propuestas por el profesor en cada una de las cinco materia escogidas para la minería, el mismo que tiene un nivel de interacción **medio-alto**; la herramienta con mayor nivel de contribución por parte del estudiante son los **foros**, el cual cuenta con un total de 202 accesos, es decir tiene un aporte **medio**; luego están con un nivel de interacción menor al anterior los **cuestionarios**,

recursos, tareas y perfil del usuario, en el orden respectivo, los cuales tienen un valor cualitativo de interacción **medio-bajo**; y finalmente están las herramientas con un nivel bajo de interacción por parte de los estudiantes que son **rol** y **utpl**; y sin ningún aporte se clasificaron en última instancia los **blogs** con un valor nulo.

Con mayor nivel de contribución por parte de los estudiantes son las páginas de las actividades, los foros son la herramienta con mayor nivel de aporte, luego están los cuestionarios, recursos, tareas y el perfil del usuario con un nivel medio; y con un nivel de interacción bajo esta la subida de archivos, rol, utpl, el twitter y el blog.

Clúster 2 Este clúster agrupó a los estudiantes con mayoría de accesos en los **foros** con un total de **1710** ingresos, y un valor cualitativo de **alto**, demás está el **perfil de usuario, cuestionarios y tareas**, con un promedio **medio-bajo** de interacción. Y finalmente están las herramientas con un promedio **bajo** de interacción como **subida de archivos** (upload), **recursos, rol, utpl y cursos**; y con valor nulo el **blog** y **twitter**.

Las páginas de las **Actividades** (*foros, cuestionarios y tareas*) tienen un nivel **alto** de contribución por parte de los estudiantes; y las páginas de **Apoyo** (*recursos, utpl y subida de archivos*) un nivel **medio** de aporte. Finalmente ingresan a las páginas de **Colaboración y Comunicación** (*foros, blog, twitter*) que a excepción de los foros también considerada como una herramienta de colaboración tienen un nivel **bajo** de interacción los estudiantes.

Resultados obtenidos de la experiencia

Con los resultados obtenidos con el modelo **SimpleKMeans** se pueden determinar atributos e **indicadores de interés** que los estudiantes tienen con las herramientas del EVA durante el proceso de aprendizaje en los cursos que el estudiante está enrolado, la cual se medirá según el uso o grado de interacción que tienen con las mismas. Y se la presenta más adelante en el capítulo '*Representación de atributos e indicadores en un modelo de datos del usuario*'.

5.6.2.2. Aplicación con el Algoritmo EM (Expectation Maximization)

El algoritmo EM es un método probabilístico que permitirá encontrar indicadores estadísticos. El algoritmo EM es más elaborado que el SimpleK-means, ya que requiere muchas más operaciones, y es más apropiado cuando los datos tienen una variabilidad estadística de modelo conocido. Este asigna a cada instancia una distribución de probabilidad de pertenencia a cada uno de los clúster.

El algoritmo puede decidir cuántos clústeres crear basado en validación cruzada o se le puede especificar a priori cuantos debe generar, y para ello, es necesario realizar algunos ajustes en los parámetros por defecto que tiene el algoritmo, ya que, si se desea obtener un número pequeño de datos representativos es necesario incrementar el número de clústeres, para que haya menos datos por clúster, se tomó una entrada de dimensión 3, y poder alargar el tiempo de procesado y obtener clasificaciones con datos de mayor frecuencia.

Para el proceso de experimentación se realizó varios experimentos, del cual el primero se lo realizó con los parámetros por defecto del algoritmo ($I=100$, $N=1$, $S=100$), y el modo de evaluación '*Use training set*' que utiliza el propio conjunto de entrenamiento para que indique el porcentaje de instancias que se van a cada grupo. Los resultados del experimento presentaron datos poco significativos para la evaluación por lo cual, se modificó los valores de *semilla*, el *número de clúster* y el *modo de evaluación*, el cual se lo realizó por *Classes to clusters evaluation* para que realice la comparación de los clústeres con varios atributos de clasificación de los datos.

```

09:53:11 - EM
Clústered Instances
0      3731 ( 27%)
1      2640 ( 19%)
2      2212 ( 16%)
3      5089 ( 37%)
4         202 ( 1%)

Log likelihood: -6.29135

Class attribute: CURSO
Classes to Clusters:
   0   1   2   3   4 <-- assigned to cluster
1843 1245 704 1174 107 | Redes y Sistemas Distribuidos
316  286 231  474  10 | Sistemas de Control
533  291 484 1163  73 | Control de Calidad
467  301 295  852   1 | Modelamiento de Datos
137  211 360  989   4 | Sistemas Basados en el Conocimiento
435  306 138  437   7 | Administraci3n de Herramientas CASE

Cluster 0 <-- Redes y Sistemas Distribuidos
Cluster 1 <-- Administraci3n de Herramientas CASE
Cluster 2 <-- Sistemas Basados en el Conocimiento
Cluster 3 <-- Control de Calidad
Cluster 4 <-- Sistemas de Control

Incorrectly clustered instances :      10192.0  73.4612 %

```

Fig. 10. Resultados con el algoritmo EM, evaluación por cursos

Fuente: Ruiz, F., (2014).

Debido a la complejidad que tiene este algoritmo, y a que se requiere más operaciones, se realizó varios experimentos con el valor del clústeres, la **Fig. 10**, muestra que el algoritmo realizó la evaluación por cursos, en el cual se puede apreciar que el curso con mayor nivel de interacción es '*Redes y Sistemas Distribuidos*' con un total de 1843 datos bien clasificados y seguido con 1163 la materia de '*Control de Calidad*'.

```

09:54:51 - EM
Clustered Instances
0      4512 ( 33%)
1      2461 ( 18%)
2      2440 ( 18%)
3      2254 ( 16%)
4      2207 ( 16%)

Log likelihood: -7.0654

Class attribute: ACTIVIDAD
Classes to Clusters:
  0      1      2      3      4 <-- assigned to cluster
2402 1243      7 1079      0 course
 15    23      0   32      0 role
 15    23      0   32      0 utpl
328   88      6  159      0 quiz
674   392 2421  457 2207 forum
 35    45      0   24      0 upload
  0     2      0    1      0 notes
  2     1      0    1      0 blog
406  289      0  192      0 user
243  301      6  118      0 assignment
392   54      0  143      0 resource
  0     0      0   16      0 label

Cluster 0 <-- course
Cluster 1 <-- assignment
Cluster 2 <-- forum
Cluster 3 <-- user
Cluster 4 <-- No class

Incorrectly clustered instances :      8558.0  61.6837 %

```

Fig. 11. Resultados con el algoritmo EM, evaluación por actividad

Fuente: Ruiz, F., (2014).

Se puede apreciar en la **Fig. 11**, que los estudiantes navegan más por las páginas de los *foros*, y con un menor nivel de interacción las *tareas* y el *perfil de usuario* en el curso de *Redes y Sistemas Distribuidos* y *Control de Calidad* respectivamente.

Más adelante se presenta una tabla comparativa con los resultados obtenidos de cada uno de los modelos de Clustering (*SimpleKMeans* y *EM*) para determinar a partir de esos datos, buscar indicadores y atributos del comportamiento de los estudiantes cuando interactúan en las actividades del EVA.

Fuente: Ruiz, F., (2014).

Algoritmos	Class attribute = Cursos					Class attribute = Actividades					Class attribute = Acciones				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Clúster	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
Instancias	3731	2640	2212	5089	202	4512	2461	2440	2254	2207	7418	3032	881	962	1581
Porcentaje	27%	19%	16%	37%	1%	33%	18%	18%	16%	16%	53%	22%	6%	7%	11%
Instancias mal clasificadas	10192.0 73.46%					8558.0 61.68%					8893.0 64.09%				

Algoritmos	EM		
	Curso	Actividad	Acciones
herramientas	Redes y Sistemas Distribuidos	Foros Tareas Perfil de usuario	Revisar Revisar Revisar

Tabla 18. Resultados de las herramientas más frecuentadas con el algoritmo EM

En la **Tabla 18**, se puede ver que los resultados obtenidos de los experimentos, presentan derivaciones similares, esto puede ser debido a que el método del modelo permite que los clústeres se superpongan, la suma de los elementos de todos los clústeres puede superar la totalidad de los elementos existentes en el conjunto de entrenamiento, y por ello, para la evaluación se puede considerar cualquiera de los resultados obtenidos del algoritmo EM.

Los resultados del segundo y tercer experimento son similares, el valor de instancias mal clasificadas es más alto que el primer experimento. Se escogió los datos mismos datos para los tres experimentos (*semilla 500, clústeres=5, Iteraciones=100*), ya que los clústeres están mejor definidos, además de tener el valor más bajo de clústeres mal clasificados, por ello, se tomaron los mismos datos con diferentes atributos a clasificar. El tercer experimento muestra que el número de instancias significativas en la muestra de los 3 clústeres es de *1582 instancias*. El resultado del agrupamiento sobre diferentes combinaciones de atributos se escogió el tercer experimento (*semilla 500, clústeres=3, Iteraciones=100*), ya que los clústeres están mejor definidos, y además tiene el valor más bajo de clústeres mal clasificados.

```

09:56:08 - EM
Clustered Instances
0      7418 ( 53%)
1      3032 ( 22%)
2       881 (  6%)
3      962 (  7%)
4     1581 ( 11%)

Log likelihood: -7.07841

Class attribute: ACCION
Classes to Clusters:
  0   1   2   3   4  <-- assigned to cluster
13   4   0   0  17  add mod
 6   0   0   0   6  update mod
36   1   0   1  32  assign
36   1   0   1  32  role_assign
 0   3   2   0   2  preview
301  0   0   0   0  add discussion
2215 16   0   0   0  view discussion
 76  29   0   0   0  add post
 77  11   0   0   0  update post
 21   3   0   0  17  editsection
 16  51  87  13  10  upload
1366 2132 392 919 1207 view
 93  251 357   1   76  view all
609  152   0   0   0  view forums
 1   0   0   0   0  subscribeall
 0   1   0   0   1  report_quiz
 0   0   2   0   2  report
 0  31   8  12  10  close attempt
 0  31   7  12  11  attempt
 0   8   0   3   3  continue attemp
 0   0   1   0  10  review
 3   0   0   0   0  search
15   6   0   0  22  view submission
17   6   0   0  18  update grades
 0   1   0   0   0  view subscribers
84   9   0   0   6  user report
 4   5   0   0   0  subscribe
 1   0   0   0   0  unsubscribe
19   0   0   0  15  add
2401 205   0   0   0  view forum
 0   75  18   0  78  editquestions
 5   0   7   0   6  update
 3   0   0   0   0  delete discussion

Cluster 0 <-- view forum
Cluster 1 <-- view
Cluster 2 <-- view all
Cluster 3 <-- upload
Cluster 4 <-- editquestions

Incorrectly clustered instances :      8893.0    64.0983 %

```

Fig. 12. Resultados con el algoritmo EM, con evaluación por acciones

Fuente: Ruiz, F., (2014).

Como se puede apreciar en la **Fig. 12**, las herramientas con mayor nivel de interacción por parte de los estudiantes durante su navegación en el EVA están clasificadas en varios grupos, las mismas que se detallan a continuación:

De los tres grupos obtenidos el grupo con mayor número de accesos realizados dentro del sistema por parte de los estudiantes de la UTPL de Informática, de la Modalidad Abierta, de un total de 112 estudiantes, navegan con mayor frecuencia en la herramienta de los **foros** en las materias de **Redes Sistemas Distribuidos y Control de Calidad** con un total bien diferenciado en comparación a las demás herramientas, en las cuales ingresan a **revisar los foros** propuestos por el docente a **revisar las discusiones**.

Una vez aplicados los algoritmos propuestos para el análisis de los mismos, se realizó un cuadro comparativo de los resultados obtenidos de las experimentaciones realizadas con el algoritmo **SimpleKMeans** y **EM**, con la finalidad de establecer cuál de los dos algoritmos presenta los resultados más eficientes, y a partir de esos datos extraer los indicadores de navegación de los estudiantes que interactúan con el EVA. Para determinar cual de ellos es mejor se realizó una comparación de los valores con los resultados obtenidos y a partir de ello determinar el resultado con menor índice de error ver **Tabla 23**.

Luego de la implementación de los algoritmos propuestos, se procedió a realizar una evaluación de los diferentes resultados obtenidos en cada una de las experimentaciones realizadas, para con ello, determinar cuál es el resultado más eficiente, igualmente las similitudes entre ellos, además con estos resultados se pretende determinar cual de los resultados obtenidos de los diferentes algoritmos implementados son los más convenientes para determinar los indicadores de navegación a las que ingresan con mayor frecuencia los estudiantes de noveno ciclo de la modalidad Abierta de la UTPL.

Fuente: Ruiz, F., (2014).

Algoritmos	SimpleKMeans			EM				
	0	1	2	0	1	2	3	4
Clúster	0	1	2	0	1	2	3	4
Instancias	3485	1936	2282	3731	2640	2212	5089	202
Porcentaje	45%	25%	30%	27%	19%	16%	37%	1%
Clúster mal clasificados	3426.0 44.5 %			10192.0 73.46%				

Tabla 19. Comparativa de los resultados de los algoritmos SimpleKMeans y EM.

Los datos arrojados por el algoritmo de clustering **SimpleKMeans** muestra el total de interacciones de los estudiantes con las herramientas dentro del sistema, además se puede apreciar que el valor de error de la suma de cuadrados es más bajo en comparación al del

modelo *EM*, también muestran en su gran mayoría consistencia y similitud entre sus características, en los diferentes atributos de los datos cargados.

Resultados obtenidos de la experiencia

Con los resultados obtenidos de los modelos detallados en la **Tabla 19**, se puede apreciar que el modelo *SimpleKMeans* determinó mayor eficiencia en los resultados obtenidos y en un intervalo de tiempo más corto al momento de ejecutarlos, además permitió obtener el indicador del nivel de interés y colaboración que tienen los estudiantes con las herramientas del EVA para su aprendizaje. En los resultados obtenidos de la implementación del algoritmo *EM*, se agrupó a los estudiantes de los cursos en varios conjuntos de datos según las características de comportamiento que tienen con EVA, y a partir de este comportamiento análogo determinar patrones de navegación. Con los datos obtenidos del proceso del modelo se determinaron indicadores estadísticos del nivel de interacción de los estudiantes con las páginas de **Actividades** (*foros, cuestionarios, tareas*) de **Apoyo** (*recursos, subida de archivos REA*), además con la navegación de los estudiantes en las herramientas de **colaboración** (*foros, twitter, blog*), y páginas de **Información general** (*perfil de usuario, utpl, etc.*); se obtuvieron indicadores de interacción y colaboración que tienen los estudiantes con la herramienta de los foros, ya que tienen un nivel de interacción alto, a diferencia del blog y twitter, que existe un porcentaje bajo de contribución por parte de los estudiantes.

5.6.3. Experimentación con la técnica Clasificación con Árboles de decisión

Una de las técnicas de minería de datos más intuitivas es la de los Árboles de decisión, y razón por la que se la ha considerado a ésta técnica es con la finalidad de encontrar a partir los datos de evaluación del clasificador las rutas de navegación del estudiante cuando interactúa con el EVA; para ello, se ha considerado la técnica de árboles de decisión, ya que es un modelo de predicción muy utilizado en el ámbito de la inteligencia artificial, y que, a partir de una base de datos se construyen diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema.

El algoritmo que se tomó, es el J-48, El cual realiza la clasificación sobre atributos simbólicos, por lo cual se aplicó el filtro “Dizcretize” en intervalos que representarán los valores de clase en el atributo usuarios, ya que es de tipo numérico. En ésta técnica se realizaron 3 experimentaciones para la validación cruzada.

5.6.3.1. Aplicación del Algoritmo J-48.

Al igual que en las técnicas anteriores, con el algoritmo **J-48** se realizaron tres experimentos con diferentes valores en el *factor de confianza* C , que es utilizado para la poda del árbol (los valores más pequeños incurren más en la poda); *la semilla* S , usada para la aleatorización de los datos; *el valor de* M , que es el número mínimo de instancias por hoja; y *el modo de evaluación* que se lo realizó con Use training set el cual evalúa el clasificador sobre el mismo conjunto sobre el que se construye el modelo predictivo para determinar el error; y por Cross-validation el cual evalúa el clasificador por evaluación cruzada según el parámetro que se le asigne a *Folds*; y *la función de evaluación*, que usa un determinado atributo para la evaluación de los resultados.

El primer experimento se lo realizó con los parámetros por defecto del algoritmo, con un valor de confianza de ($C=25$), el número mínimo de instancias $M=2$ y el *modo de evaluación* es de **Cross-validation Folds = 10**, y la *función de evaluación* se la configuró para que clasifique los datos por cursos, este atributo es el que actuará como resultado real de la clasificación de los datos, mediante evaluación cruzada; la cual divide las instancias en 10 carpetas como lo indica el parámetro folders, y en cada evaluación se toman las instancias de cada carpeta como datos de test, y el resto como datos de entrenamiento para construir el modelo. Y los errores calculados son el promedio de todas las ejecuciones.

Además se utilizó el filtro “*Remove*” para quitar el atributo ‘*PAGINAS INGRESADAS*’, para que los datos a evaluar por el clasificador sean más precisos, además se utilizó el filtro ‘*Reorder*’ para tomar como función de evaluación el atributo ‘*CURSO*’, ya que el modelo por defecto coge el ultimo atributo de los datos para evaluar; y la evaluación se la realizará por cursos, luego se ejecuta el clasificador sobre los datos de entrenamiento y nos presenta los resultados a continuación.

En la **Fig. 13.**, se puede apreciar que en la matriz de confusión los datos de las columnas, que indican las categorías clasificadas por el clasificador y los datos de las filas que indican las categorías reales de los datos; los elementos en la diagonal principal son los elementos de nuestro interés, ya que son los datos que ha acertado el clasificador y lo demás son los errores.

```

10:11:07 - trees.J48
Number of Leaves :    1002
Size of the tree :    1003

Time taken to build model: 0.36 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      13618          98.1548 %
Incorrectly Classified Instances    256           1.8452 %
Kappa statistic                    0.9762
Mean absolute error                0.0071
Root mean squared error            0.0603
Relative absolute error            2.7441 %
Root relative squared error        16.7097 %
Total Number of Instances         13874

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      -----  -
      1         0         0.952      1         0.976      1         Redes y Sistemas Distribuidos
      0.976    0         0.999      0.976    0.987      0.999     Sistemas de Control
      0.96     0         1         0.96     0.98       0.999     Control de Calidad
      0.986    0         1         0.986   0.993      1         Modelamiento de Datos
      0.966    0         1         0.966   0.983      1         Sistemas Basados en el Conocimiento
      0.971    0         1         0.971   0.985      0.999     Administraci3n de Herramientas CASE
Weighted Avg.  0.982    0.011    0.982    0.982    0.982     0.999

=== Confusion Matrix ===
  a  b  c  d  e  f  <-- classified as
5073 0  0  0  0  0  a = Redes y Sistemas Distribuidos
31 1285 1  0  0  0  b = Sistemas de Control
100 1 2443 0  0  0  c = Control de Calidad
26  0  0 1890 0  0  d = Modelamiento de Datos
58  0  0  0 1643 0  e = Sistemas Basados en el Conocimiento
39  0  0  0  0 1284  f = Administraci3n de Herramientas CASE

```

Fig. 13. Experimentación del algoritmo J-48, con los parámetros por defecto.

Fuente: Ruiz, F., (2014).

El segundo experimento se lo realizó con los mismos parámetros del primero. El valor de confianza es de ($C=25$), el número mínimo de instancias $M=2$ y el *modo de evaluación* es de **Cross-validation**, pero con la diferencia de que en este caso se cambió el valor de Folds, para que divida las instancias ya no en 10 sino en 4 carpetas para los datos del test, y al parámetro de la *función de evaluación*, al cual se lo dejó igual para que clasifique los datos por cursos.

```

11:39:30 - trees.J48
Number of Leaves :    118
Size of the tree :    185

Time taken to build model: 0.3 seconds

=== Stratified cross-validation ===
=== Summary ===
Correctly Classified Instances      10734          77.3677 %
Incorrectly Classified Instances    3140          22.6323 %
Kappa statistic                    0.7007
Mean absolute error                0.096
Root mean squared error            0.2203
Relative absolute error            36.8858 %
Root relative squared error        61.072 %
Total Number of Instances         13874

=== Detailed Accuracy By Class ===
      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
      -----  -
      0.991    0.172    0.769    0.991    0.866    0.959     Redes y Sistemas Distribuidos
      0.74     0         1         0.74     0.851    0.989     Sistemas de Control
      0.59     0.028    0.827    0.59     0.689    0.949     Control de Calidad
      0.884    0.042    0.77     0.884    0.823    0.985     Modelamiento de Datos
      0.601    0.035    0.708    0.601    0.65     0.967     Sistemas Basados en el Conocimiento
      0.389    0.031    0.569    0.389    0.462    0.93      Administraci3n de Herramientas CASE
Weighted Avg.  0.774    0.081    0.775    0.774    0.761    0.962

=== Confusion Matrix ===
  a  b  c  d  e  f  <-- classified as
5027 0 10  4 17 15  a = Redes y Sistemas Distribuidos
335 975 6  0  0  1  b = Sistemas de Control
386  0 1502 156 207 293  c = Control de Calidad
7  0  40 1693 137 39  d = Modelamiento de Datos
141  0 209 287 1022 42  e = Sistemas Basados en el Conocimiento
641  0 49 58 60 515  f = Administraci3n de Herramientas CASE

```

Fig. 14. Resultados con el algoritmo J-48, con $C=0.25$

Fuente: Ruiz, F., (2014).

La **Fig. 14.**, muestra los resultados obtenidos del segundo experimento. Se puede ver en el resumen, que el porcentaje global de errores cometidos en la evaluación es bajo, pero las instancias mal clasificadas es más alto con relación al primer experimento, también se puede apreciar en la matriz de confusión que el porcentaje de errores en las clases **c**, **d**, **e** y **f** es muy alto y sólo la clase **a** presenta un mínimo de error. Además el *error relativo de los cuadrados* es de **61.072%**, el cual es muy elevado en comparación al primer experimento.

En el tercer experimento solo se modificó la *función de evaluación*, que se escogió la opción '*Use training set*' para que evalúa el clasificador sobre el mismo conjunto sobre el que se está construyendo el modelo predictivo para determinar el error.

```

11:50:41 - trees.J48
Number of Leaves :    85
Size of the tree :    149

Time taken to build model: 0.31 seconds
=== Evaluation on training set ===
=== Summary ===
Correctly Classified Instances 10737      77.3894 %
Incorrectly Classified Instances 3137      22.6106 %
Kappa statistic 0.7047
Mean absolute error 0.0967
Root mean squared error 0.2198
Relative absolute error 37.1549 %
Root relative squared error 60.9557 %
Total Number of Instances 13874

=== Detailed Accuracy By Class ===
          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0.958    0.14    0.798      0.958    0.871    0.96    Redes y Sistemas Distribuidos
          0.736     0      1      0.736    0.848    0.989    Sistemas de Control
          0.581    0.02    0.87      0.581    0.696    0.949    Control de Calidad
          0.308    0.044   0.767    0.308    0.832    0.386    Modelamiento de Datos
          0.564    0.025   0.757    0.564    0.647    0.969    Sistemas Basados en el Conocimiento
          0.552    0.067   0.463    0.552    0.503    0.929    Administraci3n de Herramientas CASE
Weighted Avg. 0.774    0.07    0.789    0.774    0.769    0.962

=== Confusion Matrix ===
  a  b  c  d  e  f  <-- classified as
4862  0  0  0  9 202 | a = Redes y Sistemas Distribuidos
 285 969  0  0  0  63 | b = Sistemas de Control
 267  0 1477 159 185 456 | c = Control de Calidad
  0  0  20 1739 114  43 | d = Modelamiento de Datos
 146  0 201 311 960  83 | e = Sistemas Basados en el Conocimiento
 536  0  0  57  0 730 | f = Administraci3n de Herramientas CASE

```

Fig. 15. Experimentación con el algoritmo J-48, con *Use Training Set*

Fuente: Ruiz, F., (2014).

Como se puede notar con respecto a los experimentos anteriores, conforme va bajando el valor de confianza se permiten más operaciones de poda para obtener árboles cada vez más pequeños, la función de evaluación se la configuró con varias instancias para determinar resultados más eficientes, y con un porcentaje de error más bajo. En el primer caso se pudo observar que el valor del factor de confianza y la semilla, no afectan el resultado, ya que presenta el mismo resultado en los tres casos. En el segundo experimento se dejó el mismo factor de confianza del modelo, obteniendo resultados más óptimos, ya que al aumentar o disminuir el factor de confianza, también varía respectivamente el valor de error y las instancias mal clasificadas.

Fuente: Ruiz, F., (2014).

<i>Algoritmo J48</i>	(C = 0.25)	(C = 0.10)	(C = 0.25)
<i>Función de evaluación</i>	cursos	herramientas	páginas
Número total de instancias	6	6	6
Modo de evaluación	Cross-validation 10	Use training set	Cross-validation 4
Número de hojas	1002	118	85
Tamaño del Árbol	1003	185	149
Tiempo	0.36 seg.	0.3 seg.	0.31 seg.
Instancias bien clasificadas	13618	10734	10737
Porcentaje	98.15%	77.37%	77.39%
Instancias mal clasificadas	256	3140	3137
Porcentaje	1.85%	22.63%	22.61%
Error estadístico	0.98%	0.70%	0.70%
Error relativo absoluto	2.74%	36.89%	37.15%

Tabla 20. Análisis de los resultados con el algoritmo J-48.

Como podemos observar en la **Tabla 20**, los resultados más efectivos son los del primer experimento, **C=0.25**, **M=2**, **modo de evaluación Cross-validation Folds = 10**, y la **función de evaluación** fue el atributo **cursos**. Otra forma de variar el tamaño del árbol es a través de un parámetro que especifica el mínimo número de instancias por nodo, si bien es menos distinguido puesto que depende del número absoluto de instancias en el conjunto de partida.

Análisis e interpretación de resultados de la experimentación

Los resultados obtenidos de la implementación del **Algoritmo J48**, los cuales se pueden apreciar en la **Fig. 13**, del primer experimento que fue el que arrojó los mejores resultados, es que el algoritmo los clasificó por cursos, y a partir de ellos se desprenden las rutas de navegación del estudiante según las peticiones realizadas a las páginas del EVA en cada una de las herramientas a las que frecuentan los estudiantes. En los resultados se puede ver que la mayor parte de los estudiantes, interactúan solo con algunas de las herramientas dentro del sistema, y a continuación se interpretan los resultados obtenidos de la matriz de confusión.

a = Redes y Sistemas Distribuidos: Según el algoritmo éste es el curso con mayor índice de ingresos e interacción por parte de los estudiantes. Los valores en la diagonal son los aciertos, y el resto de valores son los errores, por lo tanto de los **13874** interacciones realizadas por los estudiantes en el EVA, **5073** accesos son correctamente clasificados y 0 errores.

b = Sistemas de Control: Éste es el segundo curso con mayor índice de ingresos por parte de los estudiantes, con un total de **1317** accesos, **1285** están correctamente clasificados y 32 son errores.

c = Control de Calidad: Con un total de **2544** ingresos, **2443** son accesos a las actividades de dicho curso correctamente clasificadas, y **101** son errores.

d = Modelamiento de Datos: El total de accesos por parte de los estudiantes es de **1916**, de los cuales **1890** son correctamente clasificados, y **26** son errores. Este clase es la que tiene el más bajo índice de errores en la clasificación.

e = Modelamiento de Datos: En éste curso hay un total de **1701** datos clasificados, de los cuales **1643** son instancias bien clasificadas por el algoritmo, mientras que **58** son errores.

f = Modelamiento de Datos: Por último está éste curso, con el menor índice de interacciones por parte de los estudiantes. Con un total de **1323** datos clasificados, **1284** fueron correctamente clasificados y los **39** restantes son errores de clasificación.

La **Fig. 13**, presenta el resultado del primer experimento, el cual presenta los mejores resultados obtenidos. Se pueden apreciar en la experimentación realizada con los datos del EVA implementados con el algoritmo **J48** de **Weka**, que al igual que en la mayoría de los algoritmos implementados la mayor parte de los estudiantes acceden a la **página de las actividades** (*foros, cuestionarios, tareas*), con un porcentaje menor al anterior, los estudiantes acceden a las páginas de **apoyo** (*recursos, descargas, links, REA*), y con un menor porcentaje los estudiantes ingresan a las páginas sobre **información general** (*perfil de usuario, calendario, noticias, eventos, etc.*). Por último con un porcentaje mínimo los estudiantes ingresan a páginas de **comunicación y colaboración** (*twitter, chat, mensajería, blog, etc.*), y poca navegación por las páginas de contenido informativo.

En resumen, de acuerdo a los resultados obtenidos con la técnica de clasificación, permitió determinar las rutas de navegación (características más comunes de navegación) que tienen los estudiantes con el sistema y las herramientas con las que más interactúan y usan para su aprendizaje y comunicación con el profesor/tutor responsable de los cursos dentro del sistema.

5.6.4. Resumen de los resultados obtenidos en la minería de los datos

Dentro del pre-procesamiento que se realizó en los datos, se puede sacar en resumen todas las interacciones realizadas por los estudiantes durante la navegación dentro de la plataforma del EVA, se detalla las herramientas a las cuales acceden con mayor frecuencia y las acciones que realizan dentro de cada una de ellas para interactuar con el docente y sus compañeros.

Según los resultados obtenidos se puede decir que los estudiantes más interactúan con las páginas de las **actividades** (*foros, tareas, cuestionarios*), las páginas de **apoyo** (*recursos, utpl,*

subir de archivos), y con un menor porcentaje de interacción están las páginas de **colaboración y comunicación** (*chat, blog, mensajería, twitter*), y con un índice más bajo de interacción están las páginas sobre **información general** (*perfil de usuario, calendario, noticias, eventos, etc.*) dentro de la plataforma del EVA, y realizan las siguientes acciones dentro de los cursos en los que se encuentran enrolados.

Para la interpretación de los resultados obtenidos de la implementación de las técnicas, a continuación se presenta los datos cualitativos y cuantitativos que se determinaron:

- El curso con mayor índice de interacción por parte de los estudiantes es la de “**Sistemas y Redes Distribuidos**”.
- Las herramientas a las que acceden con mayor frecuencia los estudiantes y las acciones que realizan dentro de cada una de las mismas, en los diferentes cursos son los **foros** (*revisar foros, revisar discusiones*), **tareas** (*revisar, subir tareas*) y **recursos** (*revisar*) y al **perfil de usuario** y **quiz** (*revisar*).
- En su mayoría los estudiantes en los cursos, prefieren iniciar según el tema planteado por el docente, realizar sus propias conversaciones a postear en los foros de otros compañeros del curso.
- Pocos son los estudiantes que se remitan a dar respuesta a temas dentro de los foros iniciados.
- Los estudiantes sólo acceden a las herramientas donde el docente ingresa para dejar alguna actividad como: *foros, cuestionarios, recursos y muy pocos al blog*.
- La mayor parte de los estudiantes usan las mismas herramientas.
- Todos los estudiantes revisan los mensajes ingresados en el curso por el tutor.
- Casi en su totalidad los estudiantes no prestan interés en los mensajes comentados por los compañeros.
- Todos los estudiantes una vez que ingresan en los cursos revisan sólo las herramientas en las que el tutor sube o deja alguna tarea, como foros, recursos, tareas y exámenes.
- La interacción más activa de los estudiantes dentro de las herramientas del curso son con la herramienta de los foros.

Los datos obtenidos de cada una de las técnicas con las cuales se realizó el proceso de minería se presenta en la siguiente tabla comparativa:

Accesos	Resultados con la técnica de Asociación	Resultados con la técnica de Clustering	Resultados con la técnica de Clasificación
Páginas de información general	914	562	682
<i>Perfil de usuario (user)</i>	914	562	682
Páginas de Actividades	13178	3127	3127
<i>Foros</i>	6282	2240	2240
<i>Cuestionarios</i>	901	524	524
<i>Tareas</i>	780	363	363
Páginas de Colaboración	4	3	4
<i>Chat</i>	0	0	0
<i>Blog</i>	4	3	4
Páginas de Apoyo	959	600	956
<i>Recursos</i>	629	491	600
<i>Subida de archivos (upload)</i>	198	90	192
<i>Utpl</i>	132	19	128

Tabla 21. Datos comparativos de los resultados obtenidos de la minería.

Fuente: Ruiz, F., (2014).

La razón por la cual se aplicaron técnicas de minería de datos para obtener estos resultados, en lugar de realizarlo directamente desde el EVA mediante consultas en lenguaje SQL, es debido a que los algoritmo permiten eliminar aquellos atributos que son menos representativos para el proceso de análisis, además permite mediante la selección de atributos minimizar el nivel de error en los datos hasta conseguir un modelo de patrón eficiente, y con ello devolver datos más precisos y fáciles de interpretar por el usuario.

Como se puede observar en la **Tabla 21**, las tres técnicas de minería nos presentan en los resultados que los estudiantes navegan más por las páginas de **actividades** (*foros, tareas, cuestionarios*) y por las páginas de **Apoyo** (*recursos, upload y utpl*) y con un nivel bajo de interacción en las páginas de **comunicación y colaboración** (*twitter, blog y foros*).

5.7. Gephi para visualización de la interacción de los usuarios en el EVA.

Además se puede apreciar que los resultados obtenidos de las técnicas de minería no varían en mucho los valores obtenidos. Los datos que varían entre las técnicas con un porcentaje más alto son las técnicas de Asociación, la herramienta con mayor número de interacción son las tareas, y en la técnica de Clustering los foros; además varían las materias; en la técnica de Clasificación interactúan más en el curso de **Sistemas y Redes Distribuidos** (1141 accesos), en la técnica de Clustering en **Control de Calidad** (1033 accesos) y **Modelamiento de Datos** (830 accesos) y

Sistemas basados en el conocimiento (707 accesos), y la técnica de Asociación en **Sistemas de Control** tienen mayor interacción por parte de los estudiantes.

Gephi permite evaluar y analizar grandes grafos, en este caso la navegación del estudiante en las páginas de las herramientas del EVA para su aprendizaje mediante la comunicación con el profesor. Para parametrizar las representaciones y extraer los patrones más comunes de los estudiantes, se utilizó uno de los algoritmos que dispone esta herramienta, el algoritmo de *layout*, el mismo que posee una implementación de un algoritmo dirigido por fuerzas '**Force Atlas**', este algoritmo permite realizar detección de comunidades, el mismo que se fundamenta en un método heurístico que se basa en la optimización de la modularidad, que permite encontrar particiones de alta modularidad dentro de grandes grafos en poco tiempo, y desarrolla una completa estructura jerárquica de la comunidad del grafo.

Más adelante se muestra de forma visual mediante grafos la interacción de los estudiantes y docentes en las actividades del EVA, de las materias de noveno ciclo de la UTPL, Modalidad Abierta del periodo académico Abril 2013–Agosto 2013, que es la muestra que se ha tomado de la base del EVA, para la evaluación y desarrollo de la minería de datos con la finalidad de cumplir los objetivos planteados en la presente tesis. Los datos mostrados en la **Fig. 12**, son el resultado de la interacción que tienen los estudiantes en los diferentes cursos, en el cual tanto los estudiantes que ingresan a realizar alguna acción en las herramientas de EVA a las cuales accede, se muestran más adelante en la sección de la minería de datos, además se detalla mediante grafos la interacción que los estudiantes tienen con el sistema, el grafo describe todas las páginas visitadas por parte de los estudiantes cuando ingresan a realizar alguna acción dentro de las herramientas y su interactúan con las mismas.

5.7.1. Interpretación de los datos del grafo.

La interacción de los estudiantes con las herramientas que dispone el EVA, se las clasificó a las páginas por secciones: **páginas de información general** (*perfil de usuario, servicios académicos, correo electrónico, guías, eventos, calendario, etc.*); **páginas de Actividades** (*tareas, foros, cuestionarios*); **páginas de comunicación y colaboración** (*foros, twitter, blog, mensajería, chat*) y **páginas de Apoyo** como (*recursos, links, descargas, recursos educativos abierto*); para de esta manera según las rutas de navegación en las diferentes páginas, determinar cuáles son las herramientas más frecuentadas por los estudiantes para realizar algún aporte; como se puede apreciar en la **Fig. 12**, los estudiantes no acceden a todas herramientas cuando navegan a través del EVA; los estudiantes de los cursos minados, a la cual acceden con mayor frecuencia a las páginas de actividades.

La gráfica muestra que las páginas con mayor número de interacción por parte de los estudiantes, con (4305 accesos), correspondiente a un **43.84 %**, para **revisar** (view), las actividades planteadas por el profesor dentro del curso; la interacción dentro de ellos es de 2600 correspondiente al **26.85%** en la herramienta de los **foros**; con un porcentaje menor a este, esta seguido los **cuestionarios** con un valor de 1214 interacciones, que es el **12.53%**, por parte de los estudiantes. Los **recursos** tienen 565 accesos, es decir, el **5.83%**, y las **tareas** con 553 interacciones, correspondiente al **5.51%**. Los estudiantes ingresan al **perfil del usuario** un total de 523 ingresos, que es el **5.40%** de navegación, y por último están los **blogs** con 3 accesos, el **0.03 %** y al **twitter** con un porcentaje casi nulo.

Los datos obtenidos de la **Fig. 12**, muestran las materias que fueron tomadas para desarrollar el estudio de minería de datos para determinar la interacción de los estudiantes con las herramientas del EVA, las cuales corresponden a noveno ciclo de *Informática* de la *modalidad abierta*, las cuales presentan un mayor índice de interacción con las herramientas del EVA, este ciclo contiene cinco materias con una población de 118 *participantes* enrolados en las diferentes materias; de las cuales 126 son estudiantes y 6 son *docentes*.

Los estudiantes presentan un total de 15206 interacciones a las diferentes herramientas, del periodo académico *Abril 2013 - Agosto 2013*. A continuación en la **Tabla 22**, se detalla las materias de noveno ciclo de *Informática* con el número de interacciones por parte de los estudiantes en el sistema.

Fuente: Ruiz, F., (2014).

CURSOS	NÚMERO	NÚMERO
	ESTUDIANTES	INTERACCIONES
Administración de Herramientas CASE [A]	13	1717
Control de Calidad [A]	31	2544
Modelamiento de Datos [A]	24	2362
Redes y Sistemas Distribuidos [A]	22	5073
Sistemas Basados en el Conocimiento [A]	24	2193
Sistemas de Control [A]	12	1317

Tabla 22. Datos de interacción en las materias de Informática

Según los resultados obtenidos de la gráfica realizada con *Gephi*, se puede decir que las materias con mayor número de interacción son **Redes y Sistemas Distribuidos [A]** con un total de 5073 interacciones, **Control de Calidad [A]** con un total de 2544 interacciones en las actividades dentro del sistema; y el curso **Modelamiento de Datos [A]** con un total de 2362 interacciones por parte de los estudiantes, y con un menor nivel de interacción están, el curso **Sistemas Basados en el Conocimiento [A]**, con 2193 accesos en las páginas del curso; con 1717 accesos esta **Administración de Herramientas CASE [A]**, y por último el curso de **Sistemas de Control [A]** con 1317. Además también están las páginas más accedidas por los estudiantes en el EVA, son las páginas de *Actividades* (foros, cuestionarios y tareas); y las páginas de *Apoyo* (recursos, descargas); éstas páginas los estudiantes navegan para la comunicación con el profesor.

La configuración de la herramienta de visualización de grafos **GEPHI**, se utilizó para la obtención de la gráfica de la interacción de los estudiantes con el EVA, y en la cual previamente se hicieron consultas mediante lenguaje SQL de todas las materias tomadas para el análisis de minería y se lo exporto en un archivo en formato **.csv** para posteriormente subir el archivo y configurarlo, ver **(ANEXO G)** para obtener los datos buscados mediante un grafo que muestre la navegación de los estudiantes durante la interacciones con el EVA.

5.7.2. Obtención de resultados de interacción en los cursos.

Los datos obtenidos de la interacción de los estudiantes que navegan en las diferentes páginas dentro de las herramientas y servicios disponibles en la plataforma del EVA se detallan uno a uno más adelante, los cuales se presentan los resultados de manera estadística con el procesos de minería (Weka) y visual por medio de grafos (Gephi).

Cursos.- Permite al estudiante revisar la actividades propuestas por el profesor ya que aquí se almacena toda la información relacionada a los cursos en los que está enrolado el estudiante en

el sistema. Esta información es útil para tener un registro de información de todos los eventos realizados por los usuarios en las materias y categorías en las herramientas dentro del sistema.

Fuente: Ruiz, F., (2014).

<i>RESULTADOS DE INTERACCIÓN EN LOS CURSOS</i>		
<i>Acción realizada por el estudiante</i>	<i>Número de estudiantes que visitan la página</i>	<i>Número total de veces que visitan las páginas</i>
Revisa páginas del curso Administración de herramientas.	1	1717
Ingresos a revisar el curso.	1	479
Ingresos a reportar usuarios.	1	5
Ingresos para adicionar módulos.	1	9
Ingresos para actualizar los módulos.	1	1
Ingresos para editar las secciones del curso.	1	8
Revisa páginas del curso de Control de calidad.	1	2544
Ingresos a revisar el curso.	1	1033
Ingresos a reportar usuarios.	1	6
Ingresos para adicionar módulos.	1	17
Ingresos para actualizar los módulos.	1	6
Ingresos para editar las secciones del curso.	1	17
Revisa páginas del curso de Modelamiento de datos.	1	2362
Ingresos a revisar el curso.	1	947
Ingresos a reportar usuarios.	0	0
Ingresos para adicionar módulos.	1	6
Ingresos para actualizar los módulos.	1	6
Ingresos para editar las secciones del curso.	1	22
Revisa páginas del curso de Redes y sistemas distribuidos.	1	5073
Ingresos a revisar el curso.	1	1141
Ingresos a reportar usuarios.	1	62
Ingresos para adicionar módulos.	1	13
Ingresos para actualizar los módulos.	1	6
Ingresos para editar las secciones del curso.	1	21
Revisa páginas del curso Sistemas basados en el conocimiento.	2	2193
Ingresos a revisar el curso.	1	884
Ingresos a reportar usuarios.	1	2
Ingresos para adicionar módulos.	1	19
Ingresos para actualizar los módulos.	1	3
Ingresos para editar las secciones del curso.	1	19
Revisa páginas del curso de Sistemas de control.	1	1317
Ingresos a revisar el curso.	1	474
Ingresos a reportar usuarios.	0	0
Ingresos para adicionar módulos.	1	4
Ingresos para actualizar los módulos.	0	0
Ingresos para editar las secciones del curso.	1	3

Tabla 23. Datos de Interacción de los estudiantes en los cursos de noveno ciclo.

La **Tabla 23**, muestra las acciones realizadas por los estudiantes dentro de las páginas del curso, se puede apreciar que la mayor parte de los estudiantes ingresan al sistema a revisar las diferentes actividades propuestas por el profesor dentro de los cursos.

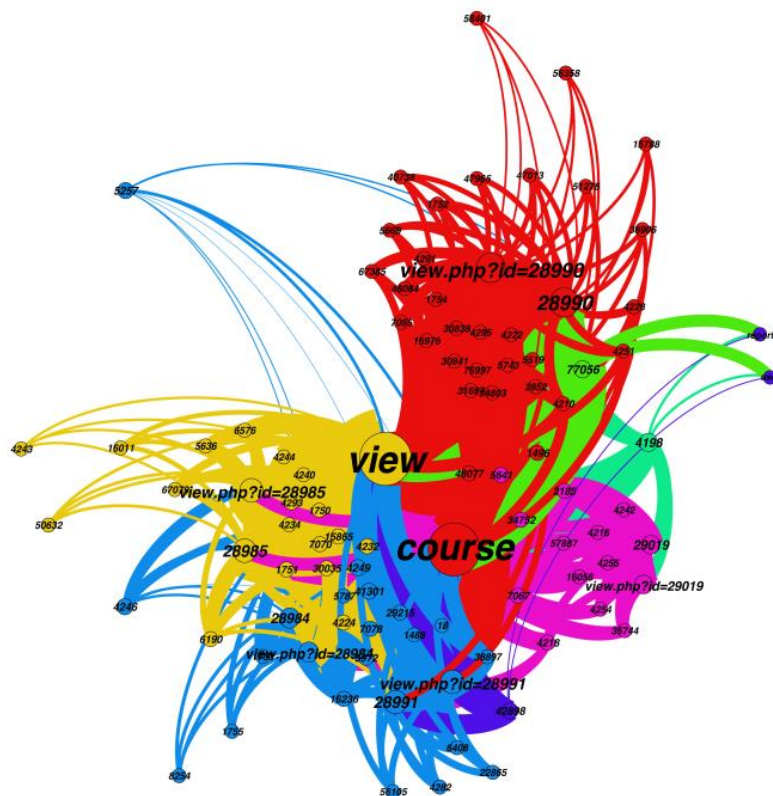


Fig. 13. Resultado de las interacciones realizadas en los cursos (páginas frecuentadas)

Fuente: Ruiz, F., (2014).

Los grafos muestran la interacción con cada una de las herramientas y las páginas a las cuales acceden los estudiantes y las acciones que realizan dentro de ellas, tanto para el aprendizaje del estudiante como para la comunicación con su docente y demás compañeros; además, se puede apreciar dentro de los grafos que cada circunferencia corresponde a los estudiantes enrolados en los curso, y las acciones realizadas por los estudiantes y la interacción que tienen entre ellas.

Como se puede apreciar en la **Fig. 13**, la mayor parte de estudiantes acceden a los diferentes cursos para revisar las actividades propuestas por cada uno de los docentes de los cursos de noveno ciclo de Informática. Uno de los cursos con mayor número de accesos por parte de los estudiantes es el curso de **'Redes y Sistemas Distribuidos [A]'** (color rojo); el curso de **Control de Calidad** (color azul), y **Modelamiento de Datos** (color amarillo), en el orden respectivo; y con un porcentaje menor están los cursos de **'Sistemas basados en el conocimiento [A]'** (color azul); y otro de los cursos al cual los estudiantes acceden a revisar las actividades propuestas por el docente dentro del curso, está el curso de **'Administración de herramientas CASE'** (color verde); y por ultimo con un nivel menor a los anteriores de interacción por parte de los estudiantes está el curso de **'Sistemas de Control'** (color celeste).

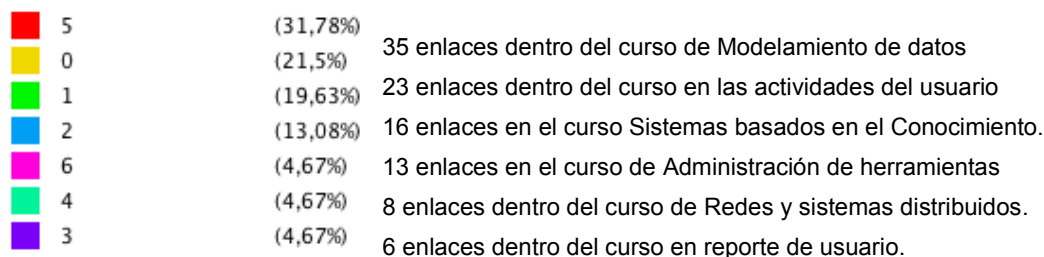


Fig. 14. Valores de rango de interacción de los estudiantes en los cursos.

Fuente: Ruiz, F., (2014).

Como se puede apreciar en la **Fig. 14**, se encontraron 7 comunidades, las cuales representan los grupos con el número de ingresos por parte de los estudiantes en los diferentes cursos, además estas comunidades presentan los valores de rango según el número de accesos que tienen los estudiantes a las páginas dentro de los cursos, el valor porcentual de accesos y el número de enlaces realizados a cada una de las herramientas y acciones dentro del EVA. Esta gráfica se la obtuvo con Gephi, permite observar mediante valores porcentuales las herramientas con mayor número de visitas de manera global en todos los cursos tomados para el proceso de minería. Para obtener estos datos se exportó en formato .csv desde la base del EVA, todos los datos de interacción de los estudiantes de noveno ciclo de la carrera de Informática.

Los resultados obtenidos a través de los datos ingresados fueron determinados a partir del análisis de la entidad de Log de los estudiantes de la UTPL, de noveno ciclo de la carrera de Informática del periodo académico Abril 2013 – Agosto 2013, de la modalidad a Distancia.

5.7.3. Conclusiones:

Como podemos apreciar en los resultados obtenidos con Gephi de cada una de las herramientas utilizadas durante la interacción de los estudiantes el EVA, se puede decir que, la mayoría de los resultados concuerdan con los resultados obtenidos con las técnicas de minería, entre ellas se destacan las siguientes:

Todos los estudiantes ingresan a revisar las actividades propuestas por los profesores/tutores en cada uno de los cursos en los que se encuentran enrolados. Según los resultados obtenidos de la minería, el curso con mayor nivel de interacción en el EVA es la asignatura de **Redes y Sistemas Distribuidos, Control de Calidad y Modelamiento de datos**, a diferencia de los resultados obtenidos con Gephi que presenta es **Sistemas basados en el conocimiento y Sistemas de Control** y la asignatura con el mínimo nivel de interacción es **Administración de herramientas CASE**.

Ambos resultados concuerdan que las páginas más frecuentadas en los cinco cursos por los estudiantes son las páginas de las actividades (*foros, cuestionarios, tareas*) y los recursos; y las

páginas con menor nivel de interacción son las páginas de Colaboración y comunicaron (*twitter, blog, chat*). La herramienta más frecuentada por los estudiantes enrolados en los cursos, en ambos casos es la herramienta de los foros. Los estudiantes que interactúan con la herramienta de los foros primeramente revisan las discusiones de los demás compañeros, antes de agregar una nueva discusión.

5.8. Información de la navegación de los estudiantes en las páginas del EVA.

La interacción que tienen los estudiantes con las herramientas y servicios de la plataforma del EVA cuando navegan por las páginas del sistema, se detallan más adelante, de forma gráfica y estadística. En la gráfica se puede apreciar variedad de colores, de los cuales cada uno de ellos representa la acción realizada en la herramienta y los nodos con mayor número de accesos se resalta con letras más grandes y colores más fuertes, también muestra la ruta de navegación que el estudiante tiene dentro de la herramienta.

La interacción de los estudiantes con las herramientas del EVA, son detalladas en la **Tabla 24**, la cual describe de forma general las páginas más frecuentadas por los estudiantes que fueron minados en el estudio realizado para la presente tesis.

5.8.1. Obtención de resultados de interacción en las páginas del EVA

Los datos de interacción de los estudiantes se detallan más adelante:

Fuente: Ruiz, F., (2014).

<i>RESULTADOS DE INTERACCIÓN Y NAVEGACIÓN EN LAS PÁGINAS DEL EVA</i>				
<i>Nro.</i>	<i>Curso</i>	<i>Tipo de Página</i>	<i>Nº de accesos</i>	<i>Porcentaje</i>
0	REDES Y SISTEMAS DISTRIBUIDOS	Páginas de actividades	4636	46,36%
1	CONTROL DE CALIDAD	Páginas de actividades	2103	21,03%
2	MODELAMIENTO DE DATOS	Páginas de actividades	2089	20,89%
3	REDES Y SISTEMAS DISTRIBUIDOS	Páginas de actividades y apoyo	4758	47,58%
4	CONTROL DE CALIDAD	Páginas de actividades y apoyo	2302	23,02%
5	MODELAMIENTO DE DATOS	Páginas de actividades y apoyo	2170	21,70%
6	REDES Y SISTEMAS DISTRIBUIDOS	Páginas de actividades: Curso	1243	12,43%
7	REDES Y SISTEMAS DISTRIBUIDOS	Páginas de actividades: Foros	3002	30,02%
8	SISTEMAS BASADOS EN EL CONOCIMIENTO	Páginas de apoyo: Recursos	344	3,44%
9	MODELAMIENTO DE DATOS	Páginas de actividades: Cuestionario	223	2,23%
10	REDES Y SISTEMAS DISTRIBUIDOS	Páginas de apoyo: Usuario	289	2,89%
11	REDES Y SISTEMAS DISTRIBUIDOS	Páginas de actividades: Tareas	301	3,01%

Tabla 24. Resultados de la Interacción de los estudiantes en el EVA.

Los datos extraídos de la interacción de los estudiantes con las herramientas del EVA, muestra en porcentaje la navegación por las diferentes páginas dentro del sistema. Como se puede

apreciar en la **Tabla 24**, el Curso con mayor nivel de interacción es **Redes y Sistemas Distribuidos**, y las páginas mas accesos por parte de los estudiantes y los docentes, es en las páginas de **Actividades** (*curso, foros, tareas, cuestionarios*) y las de **Apoyo** (*recursos, upload*) y una herramienta de las páginas del perfil de usuario (*user*).

A continuación se presenta la interacción que tienen los estudiantes con las herramientas y servicios del EVA, a los cuales el estudiante ingresa con mayor frecuencia y realiza algún aporte en los cursos, que proponen los docentes para el proceso de aprendizaje del estudiante y la comunicación con el mismo. Más adelante se presentan tanto resultados visuales (grafos), como resultados estadísticos (SQL).

- ✓ *Foros,*
- ✓ *Cuestionarios,*
- ✓ *Tareas,*
- ✓ *Recursos,*
- ✓ *Perfil del usuario (user) y*
- ✓ *Subida de archivos (upload)*

Y con un nivel menor de frecuencia, por un mínimo de estudiantes interactúan con las siguientes herramientas:

- ✓ *Blog,*
- ✓ *Mensajería,*
- ✓ *Chat...*

5.8.2. Obtención de resultados de interacción en las Tareas

Los datos obtenidos de la interacción de los estudiantes con las páginas dentro de las tareas:

Fuente: Ruiz, F., (2014).

<i>RESULTADOS DE INTERACCIÓN EN LAS TAREAS</i>		
<i>Acción realizada por el estudiante</i>	<i>Número de estudiantes que visitan la página</i>	<i>Número total de veces que visitan las páginas</i>
Ingresos a las páginas de las tareas en los cursos.	126	780
Ingresos para revisar todas las tareas propuestas.	73	207
Ingresos para revisar las tareas del curso.	95	333
Ingresos para subir las tareas al sistema.	73	73
Ingresos para ver la presentación de las tareas.	75	75

Tabla 25. Datos de Interacción de los estudiantes.

La herramienta de las **TAREAS** es una de las actividades con mayor número de ingresos.

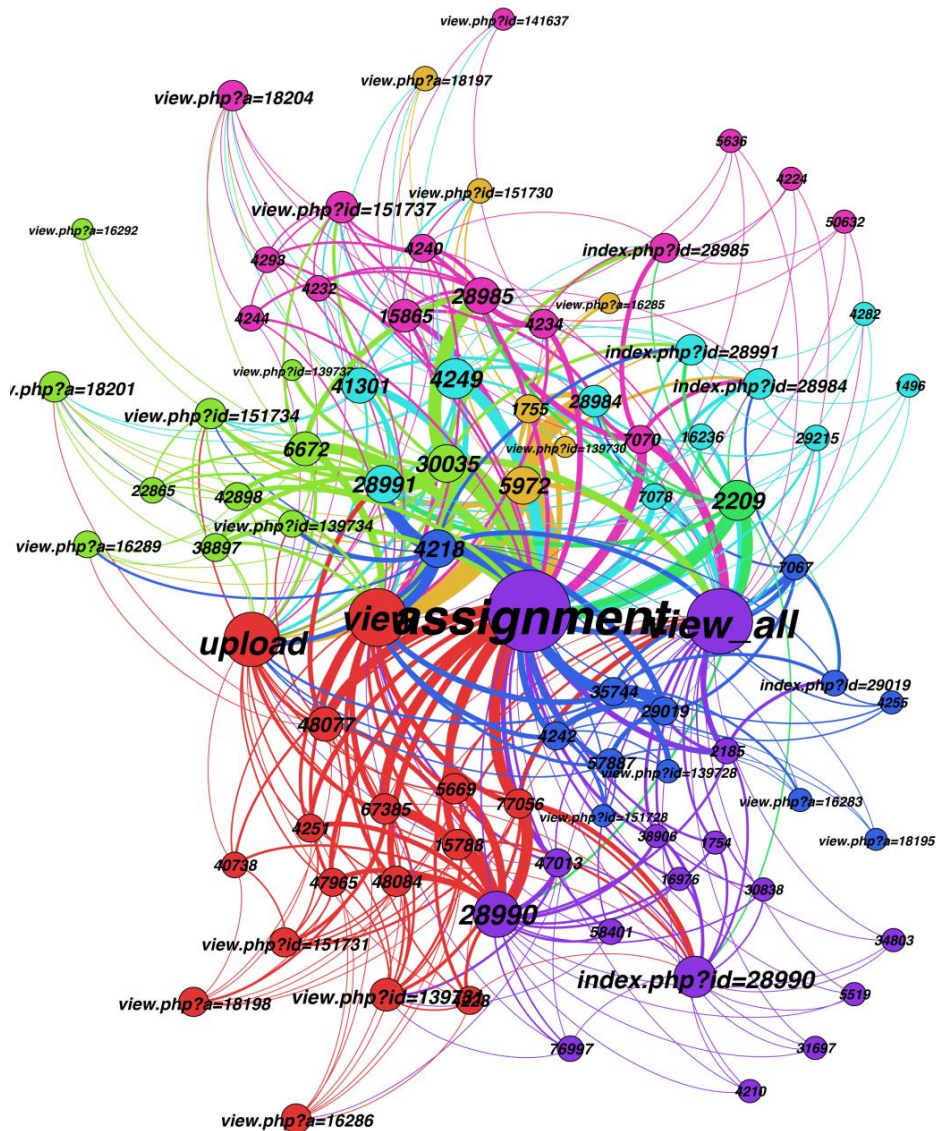


Fig. 15. Resultado de las interacciones realizadas en las tareas (páginas frecuentadas).
 Fuente: Ruiz, F., (2014).

En la **Tabla 25**, se puede ver que la mayor parte de los estudiantes ingresan a *revisar las tareas*, y la mayor parte de ellos la *suben a las tareas*, y pocos revisan la *presentación de sus tareas*.

La gráfica muestra la interacción de los estudiantes en las páginas con mayor número de ingresos en la herramienta **Tareas**, en la **Fig. 15**, podemos observar que en el epicentro se encuentra el nodo *tareas*, y a partir de ese nodo parten los demás nodos que pertenecen a los estudiantes que interactúan con la herramienta y desde allí salen las aristas en conexión con los nodos más pequeños que pertenecen a cada una de las páginas visitadas por los estudiantes durante su navegación por el EVA. Las aristas con mayor intensidad que se conectan con los diferentes nodos de la gráfica son las interacciones más frecuentes por parte de los estudiantes a

los diferentes servicios y páginas del EVA. En la grafica se puede ver que el curso con mayor aporte en la actividad de de las tareas es **Redes y Sistemas Distribuidos**.

5.8.3. Obtención de resultados de interacción en los foros

Los resultados obtenidos de la interacción de los estudiantes con los **Foros** propuestos por el tutor en el EVA, se presentan a continuación:

Fuente: Ruiz, F., (2014).

<i>RESULTADOS DE INTERACCIÓN EN LOS FOROS</i>		
<i>Rutas de navegación del estudiante</i>	<i>Nro. de estudiantes que visitan la página</i>	<i>Nro. total de veces que visitan las páginas</i>
Ingresos a las páginas de los foros.	122	6282
Ingresos a todos los foros dentro de los cursos.	126	769
Ingresos a revisar los foros.	126	2695
Ingresan para adicionar foros.	6	26
Ingresos a revisar los debates en los foros.	124	2247
Ingresos a la suscripción en los foros propuestos	9	9
Ingresan a adicionar debates dentro de los foros	6	125
Ingresan a adicionar debates dentro de los foros.	122	301
Ingresan a adicionar posts dentro de los foros	105	105
Ingresos a eliminar los debates dentro de los foros	3	3
Ingresan a buscar dentro de los foros	3	3
Páginas donde los estudiantes dan de baja el foro	1	1
Ingreso a la página de actualización de posts en los foros	85	89
Ingreso a la página de reportes de usuario en los foros	6	24

Tabla 26. Datos de Interacción de los estudiantes en los foros.

La interacción de los estudiantes con la herramienta de los '**Foros**', como se puede apreciar en la **Tabla 26**, es una de las herramientas con *mayor nivel interaccion y aporte* por parte de los estudiantes y la que más usan los docentes para el proceso de aprendizaje del estudiante. Como se puede ver, la mayoría de los estudiantes ingresan a revisar los *foros* propuestos por el docente del curso; y luego ingresan a revisar las discusiones ingresadas por los estudiantes, lo cual quiere decir, que el estudiante antes de ingresar una discusión, antes revisa las discusiones de sus compañeros. Después ingresan a adicionar una discusión, en un *nivel menor* de interacción ingresan a revisar, actualizar, crear y buscar posts dentro del foro, y muy pocos a la suscripción en los foros. El curso con mayor nivel de interaccion en la actividad de los foros es **Redes y Sistemas Distribuidos**.

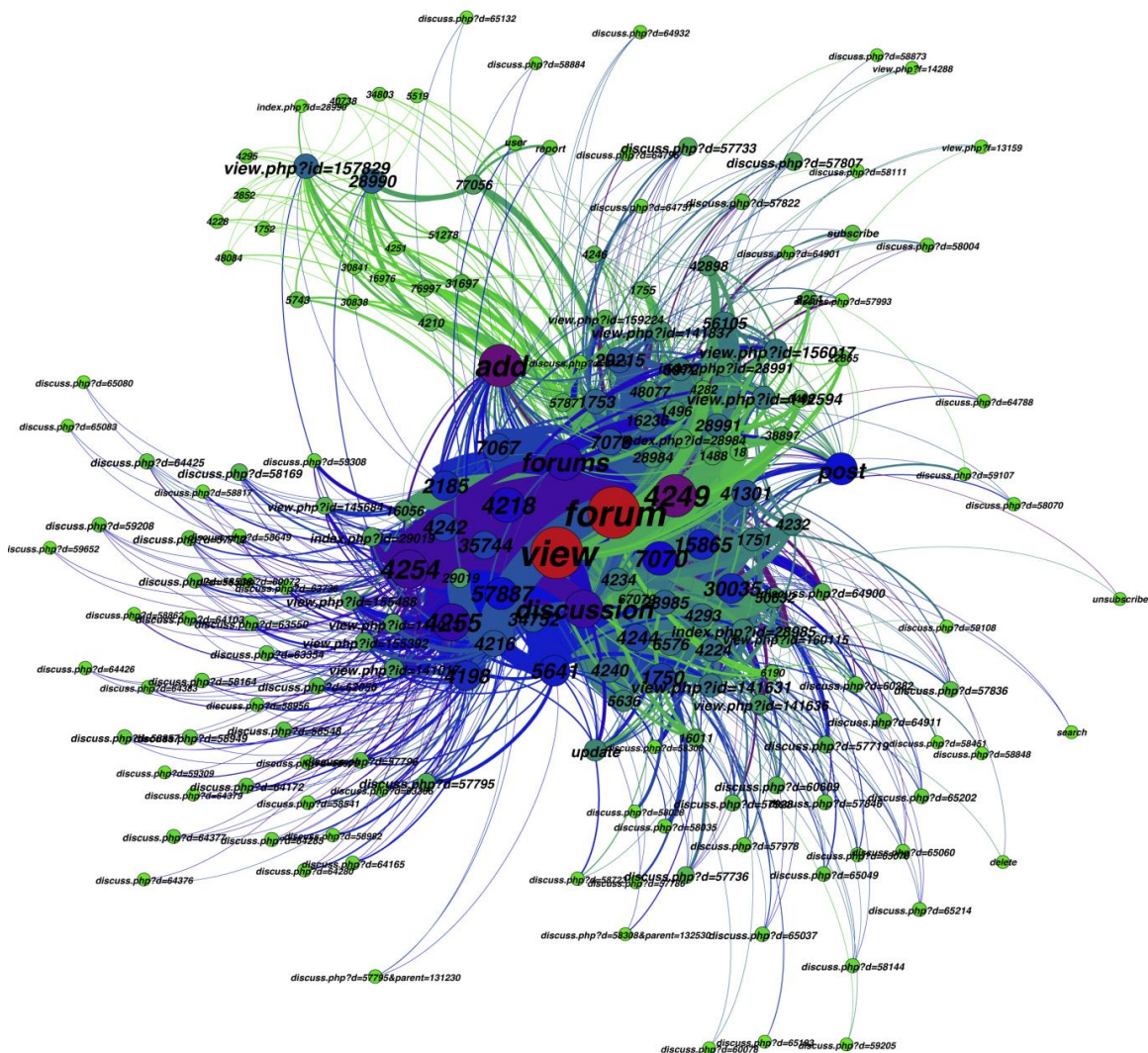


Fig. 16. Resultado de las interacciones realizadas en los foros (páginas frecuentadas)
 Fuente: Ruiz, F., (2014).

5.8.4. Obtención de resultados de interacción en los recursos

Los resultados obtenidos de la interacción de los estudiantes con los **Recursos** dentro de los cursos en la plataforma, se detallan en la **Tabla 27**, a las cuales ingresan los estudiantes para revisar, leer, bajar o hacer uso de los recursos propuestos por los tutores/profesores de cada uno de los cursos de noveno ciclo de Informática para la ayuda del estudiante.

Fuente: Ruiz, F., (2014).

RESULTADOS DE INTERACCIÓN EN LOS RECURSOS		
Rutas de navegación del estudiante	Nro de estudiantes que visitan la página	Nro total de veces que visitan las páginas
Ingresos a los recursos dentro de los cursos.	108	629
Ingresos a revisar a todos los recursos de los cursos.	76	76
Ingresos a revisar los recursos.	108	523
Ingresos para adicionar un recurso (docente).	6	28

Tabla 27. Datos de Interacción de los estudiantes en los recursos.

En los resultados obtenidos de la interacción de los estudiantes con los **'Recursos'** en el EVA, se puede ver en la **Tabla 27**, que la mayoría de los estudiantes ingresan al sistema a revisar los recursos propuestos por el docente, además ingresan a los recursos para leer, bajar o revisar el recurso disponible dentro del curso.

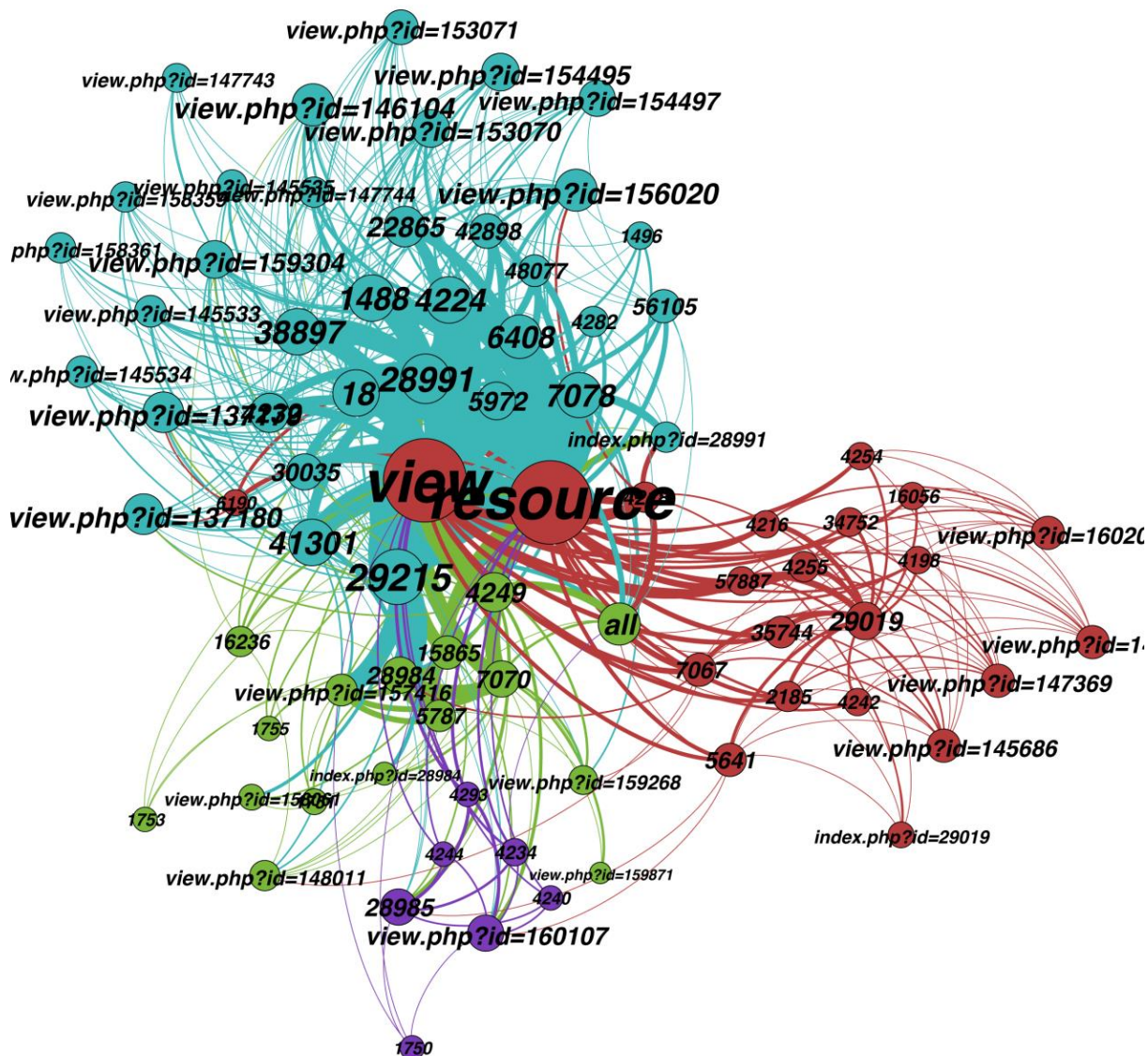


Fig. 17. Resultado de las interacciones realizadas en los recursos (páginas frecuentadas)
 Fuente: Ruiz, F., (2014).

En la **Fig. 17**, se puede apreciar que los estudiantes que acceden a las páginas de los **Recursos** frecuentan más a revisar los recursos ingresados por el docente. Además se puede ver que solo hay cinco de los seis cursos evaluados que interactúan con esta herramienta, y el curso con mayor aporte por parte de los estudiantes es **Sistemas Basados en el Conocimiento**.

5.8.5. Obtención de resultados de interacción en los cuestionarios

Los resultados obtenidos de la interacción de los estudiantes con las tareas propuestas por el docente en los cursos de noveno ciclo de Informática, ver **Tabla 28**, presentada a continuación.

Fuente: Ruiz, F., (2014).

<i>RESULTADOS DE INTERACCIÓN EN LOS CUESTIONARIOS</i>		
<i>Rutas de navegación del estudiante</i>	<i>Nro de estudiantes que visitan la página</i>	<i>Nro total de veces que visitan las páginas</i>
Ingresos a las páginas de los cuestionarios.	126	901
Ingresos a revisar todos los cuestionarios de los cursos.	96	96
Ingresos a revisar los cuestionarios del curso.	86	164
Ingresos para previsualizar los cuestionarios.	15	15
Ingresos para realizar uno de los intentos.	61	61
Ingresos para continuar con el intento en el cuestionario.	14	14
Ingresos para finalizar el intento en el cuestionario.	64	64
Ingresos para revisar los cuestionarios finalizados.	14	14
Ingresos para reportar un cuestionario.	5	5
Ingresos para adicionar un cuestionario (docente).	1	1
Ingresos para editar un cuestionario (docente).	6	462

Tabla 28. Datos de Interacción de los estudiantes en los cuestionarios.

Como se puede apreciar en la **Tabla 28**, los estudiantes navegan en la herramienta de los **Cuestionarios** dentro del sistema, en el cual ingresan para *revisar los cuestionarios* propuestos por el docente dentro de los cursos, y luego proceden a resolver el cuestionario y finalizar y cerrar el cuestionario, y algunos ingresan a previsualizar los cuestionarios resueltos.

En la **Fig. 19**, se puede ver la interacción que tienen los estudiantes con la actividad de los **Cuestionarios**, la cual es una de las herramientas con un nivel de aporte medio-alto con relación a otras herramientas, ya que es la segunda herramienta más utilizada, tanto por el docente como para los estudiantes, para evaluar el grado de aprendizaje en los estudiantes. En la gráfica se puede ver que el curso con mayor índice de interacción con esta actividad es el curso **Administración de Herramientas CASE, Control de Calidad y Sistemas Basados en el Conocimiento**.

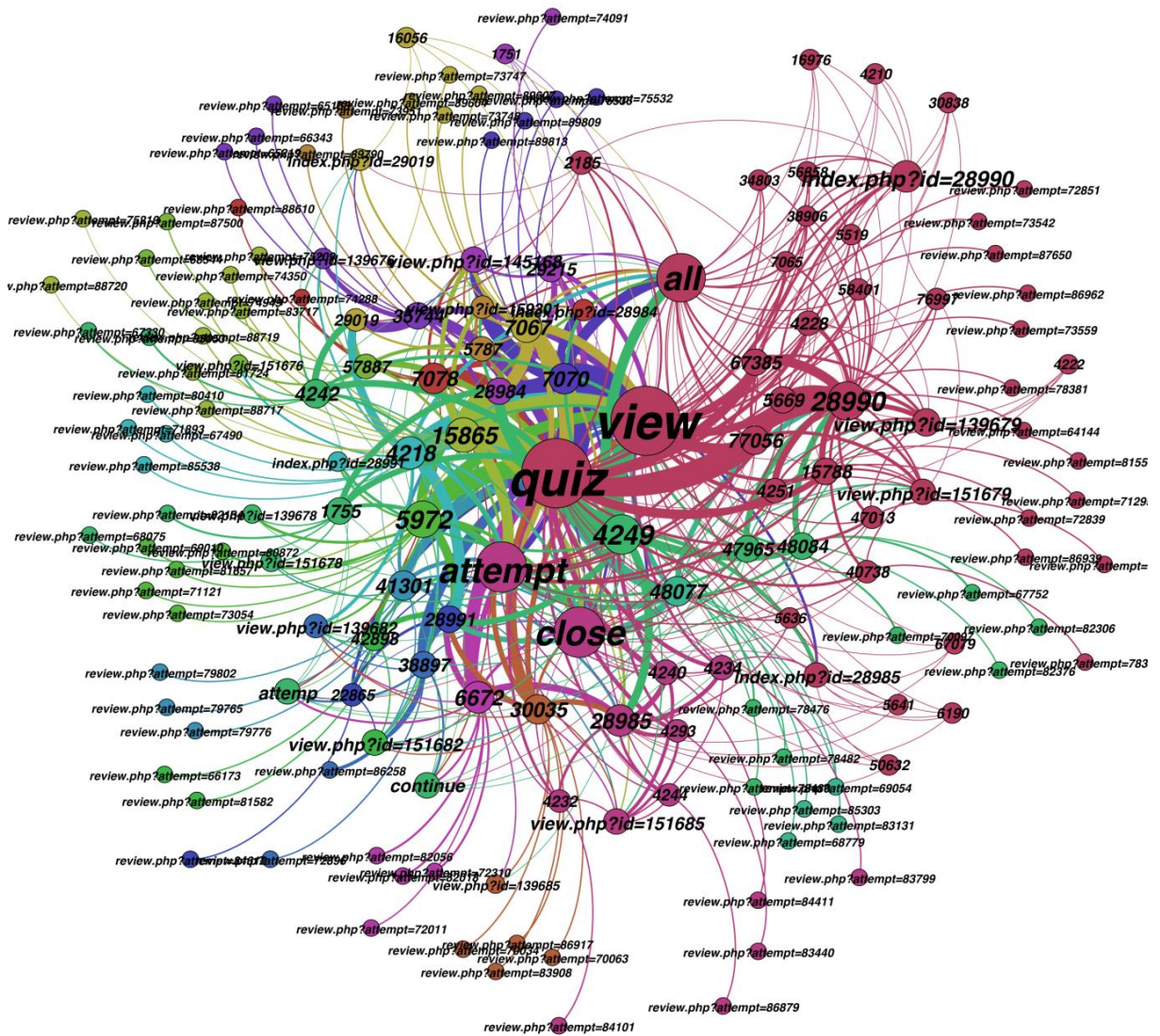


Fig. 18. Resultado de las interacciones realizadas en los cuestionarios (páginas frecuentadas)
 Fuente: Ruiz, F., (2014).

La **Fig. 18**, muestra que los **Cuestionarios (Quiz)** es una de las herramientas con mayor nivel de interacción por parte de estudiantes y docentes de todos los cursos de noveno ciclo de Informática de la Modalidad a distancia de la UTPL.

5.8.6. Obtención de resultados de interacción en el perfil de usuario

Los resultados de la interacción de los estudiantes con las páginas del perfil de usuario en los cursos por parte del estudiante se detallan a continuación en la **Tabla 29**, la cual detalla cada uno de los accesos a las páginas de la plataforma dentro de esta herramienta.

Fuente: Ruiz, F., (2014).

RESULTADOS DE INTERACCIÓN EN EL PERFIL DE USUARIO

Rutas de navegación del estudiante	Nro de estudiantes que visitan la página	Nro total de veces que visitan las páginas
Ingresos a las páginas del Perfil de usuario.	126	449
Ingresos para revisar el perfil del usuario en el curso.	53	479
Ingresos para actualizar el perfil del usuario.	6	6

Tabla 29. Datos de Interacción de los estudiantes dentro del perfil de usuario.

Como se puede apreciar en la **Tabla 29**, cuando los estudiantes navegan por el sistema revisan su perfil y no entran ni realizan ninguna otra acción y en una mínima cantidad los actualizan.

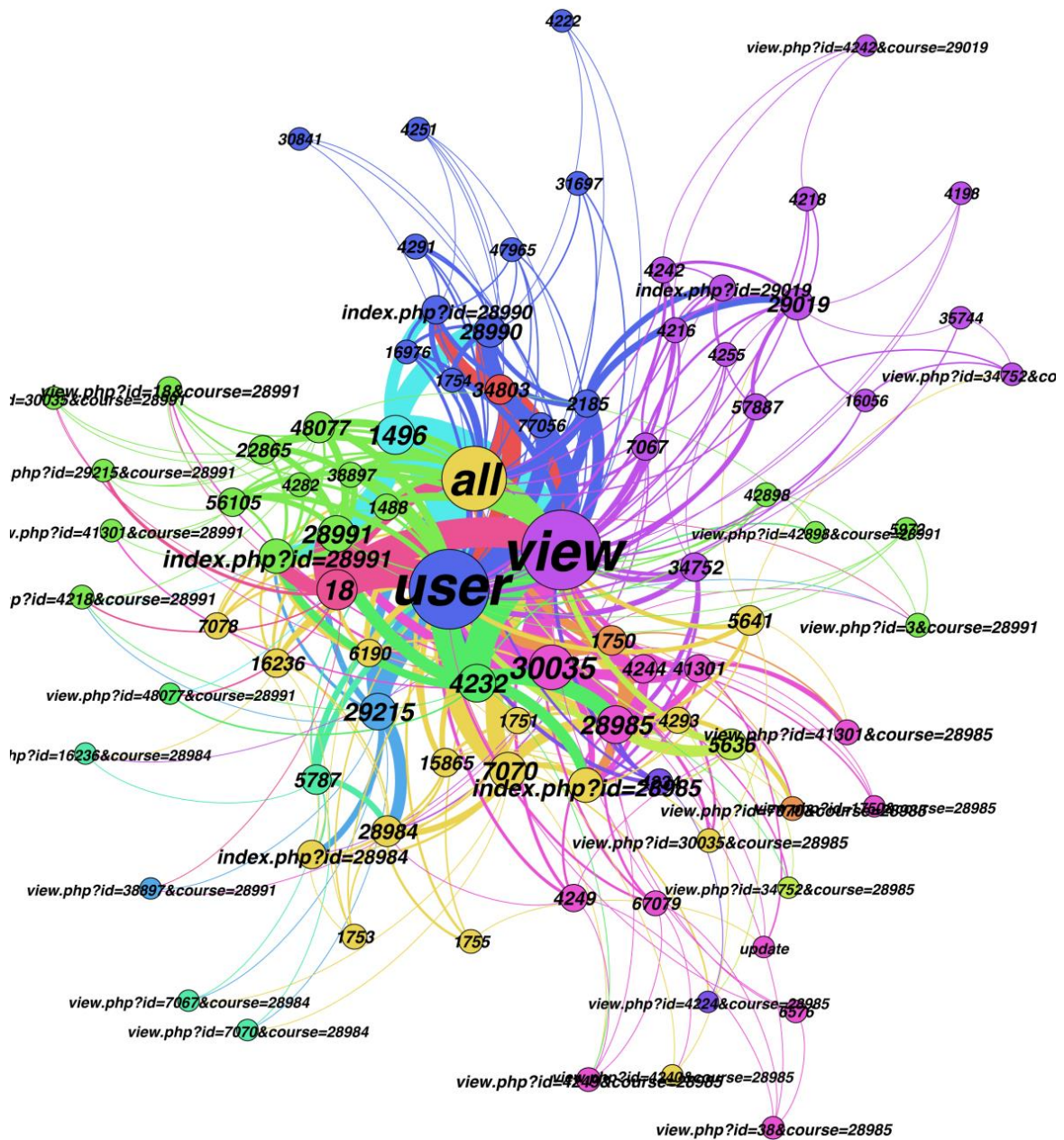


Fig. 19. Resultado de las interacciones realizadas en el perfil del usuario (páginas frecuentadas).

Fuente: Ruiz, F., (2014).

La interacción de los estudiantes enrolados en los cursos de noveno ciclo dentro de su perfil de usuario se puede apreciar en la **Fig. 19**, en la cual los estudiantes ingresan a revisar su perfil en el sistema.

5.8.7. Obtención de resultados de interacción en el Blog

Los resultados obtenidos de la interacción de los estudiantes en la herramienta del **Blog** en el EVA dentro de cada uno de los cursos de noveno ciclo de Informática en los que se encuentran enrolados los estudiantes se detallan en la **Tabla 30**, presentada a continuación.

Fuente: Ruiz, F., (2014).

RESULTADOS DE INTERACCIÓN EN EL BLOG		
Rutas de navegación del estudiante	Nro de estudiantes que visitan la página	Nro total de veces que visitan las páginas
Visitas a la páginas del blog en los cursos.	4	4
Ingresos para revisar los blogs.	4	4

Tabla 30. Datos de Interacción de los estudiantes en el blog del EVA.

Como se puede apreciar en la **Tabla 30**, cuando los estudiantes navegan por el sistema no realizan ningún tipo de acción en el Twitter del EVA, ni ingresan a este tipo de páginas de comunicación y colaboración.

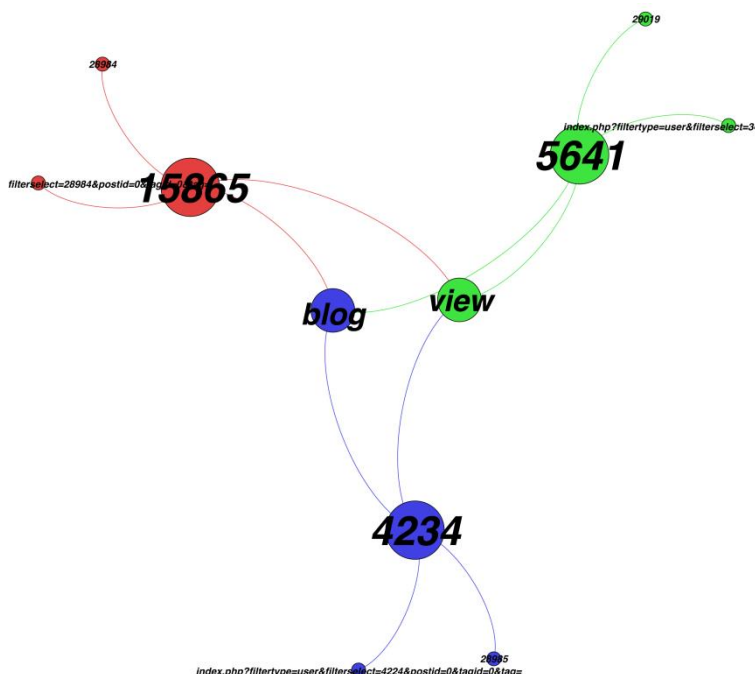


Fig. 20. Resultado de las interacciones realizadas en el blog (% de páginas frecuentadas)

Fuente: Ruiz, F., (2014).

Como se lo puede apreciar en la **Fig. 20**, los estudiantes de noveno ciclo tienen un nivel de interacción muy bajo con la herramienta **Blog** dentro del sistema. Los cursos en los cuales existe

un aporte son en los cursos de *Administración de Herramientas CASE*, *Control de Calidad*, *Modelamiento de datos y Redes y Sistemas Distribuidos*.

5.8.8. Obtención de resultados de interacción en la página UTPL

Los resultados obtenidos de la interacción de los estudiantes con las páginas de la herramienta de *'Utpl'* en las materias de noveno ciclo de Informática.

Fuente: Ruiz, F., (2014).

<i>RESULTADOS DE INTERACCIÓN EN LA PÁGINA UTPL</i>		
<i>Rutas de navegación del estudiante</i>	<i>Nro de estudiantes que visitan la página</i>	<i>Nro total de veces que visitan las páginas</i>
Visita a las páginas de Utpl dentro de los cursos.	122	132
Ingresos para asignar roles en el curso.	3	132

Tabla 32. Datos de Interacción de los estudiantes dentro de la página de Utpl.

La **Tabla 32**, muestra los estudiantes que ingresan al EVA para asignar un rol en su perfil.

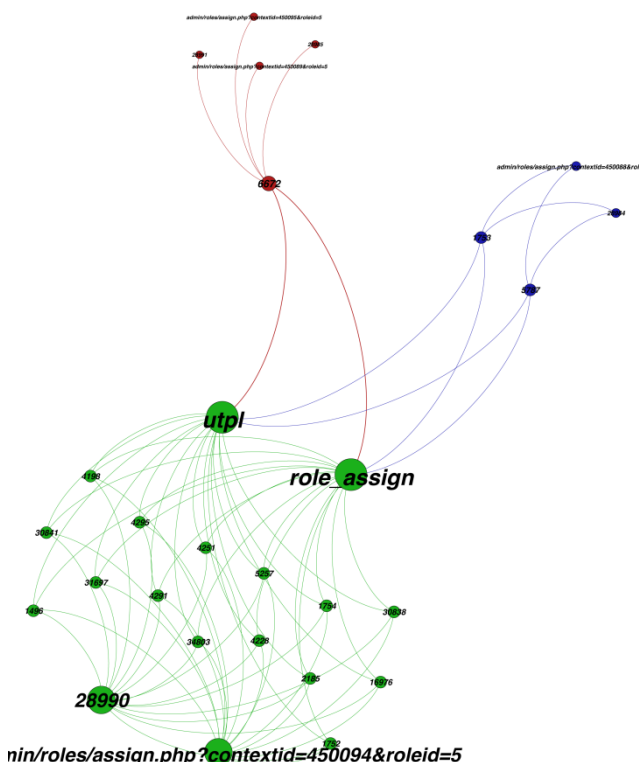


Fig. 22. Resultado de las interacciones realizadas en UTPL (páginas frecuentadas)

Fuente: Ruiz, F., (2014).

Como se puede ver en la **Fig. 22**, son pocos los estudiantes que interactúan con las páginas de esta herramienta, los estudiantes que más interactuaron con estas páginas son los estudiantes enrolados en el curso *Control de calidad*.

5.8.9. Obtención de resultados de interacción en páginas de subir archivos (Upload).

Los resultados obtenidos de la interacción de los estudiantes con la herramienta **Upload**, en las materias de noveno ciclo de Informática de los estudiantes enrolados se detallan en la **Tabla 33**, en la cual ingresaron para subir las tareas de cada asignatura.

Fuente: Ruiz, F., (2014).

RESULTADOS DE INTERACCIÓN EN LA HERRAMIENTA UPLOAD

Rutas de navegación del estudiante	Nro. de estudiantes que visitan la página	Nro. total de veces que visitan las páginas
Ingresos a las páginas de subida de archivos.	22	198
Ingresos a subir archivos en los foros.	24	151
Páginas de las tareas en las que subieron archivos.	47	47

Tabla 33. Datos de Interacción de los estudiantes en la subida de archivos del EVA.

La **Fig. 23**, muestra la interacción del estudiante con la herramienta de descarga de archivos.

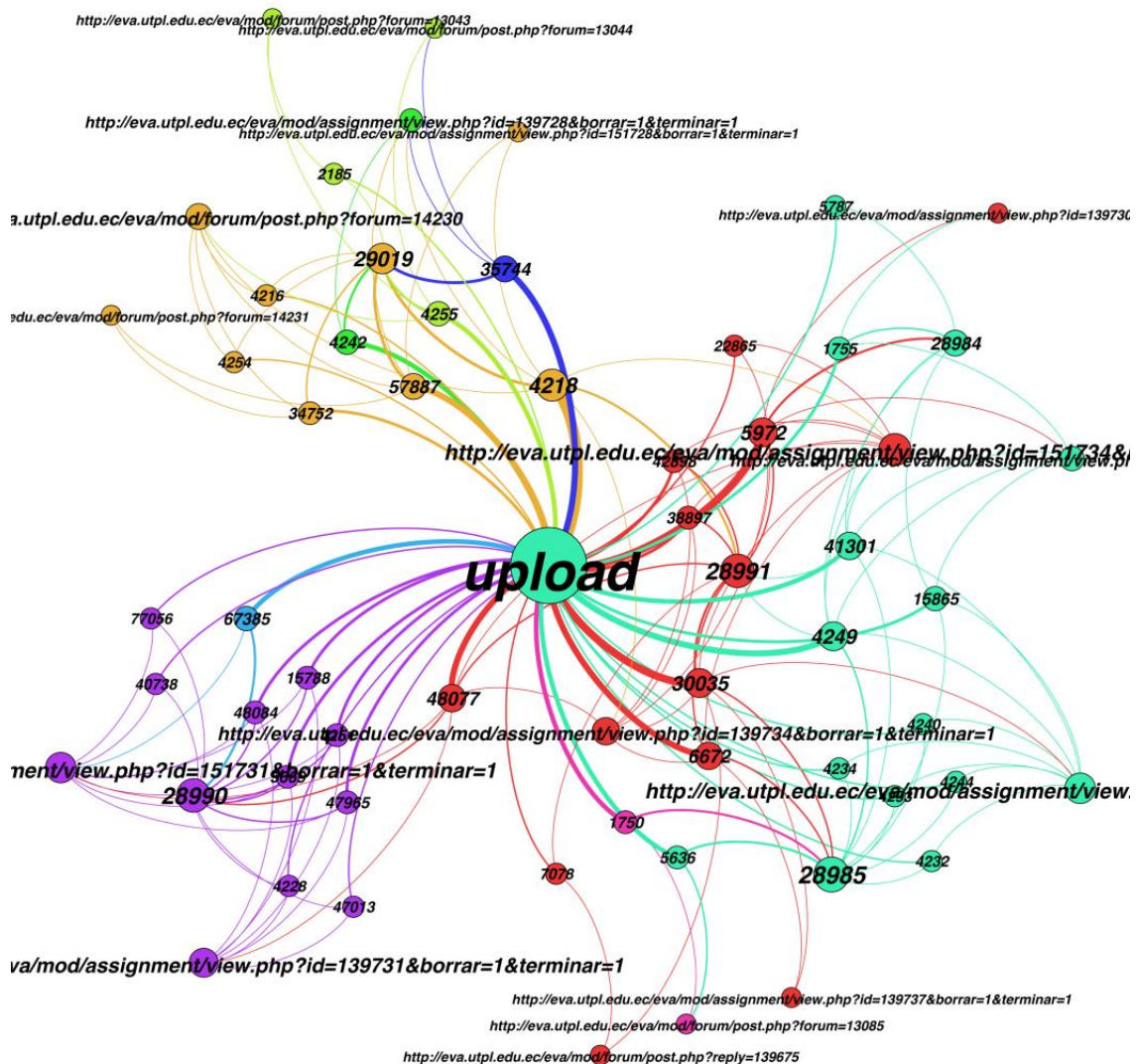


Fig. 23. Resultado de las interacciones realizadas en páginas de subir archivos.

Fuente: Ruiz, F., (2014).

Como se puede apreciar en la **Fig. 23**, la mayor parte de los estudiantes de noveno ciclo interactúan con las páginas de la herramienta de subir archivos dentro de las *tareas y los foros*, que son las herramientas con las que más interactúan los estudiantes dentro del sistema.

5.9. Resultados de interacción de los estudiantes en el EVA

Los resultados obtenidos de la interacción de los estudiantes con las tareas propuestas por el tutor en los cursos de noveno ciclo de Informática se detallan en la **Tabla 34**, presentada a continuación, la cual detalla el número total de accesos a las diferentes páginas disponibles en la plataforma del EVA, como las páginas de Información general, de Actividades, de Apoyo y de las páginas de Comunicación y colaboración por parte de los estudiantes dentro del sistema.

Fuente: Ruiz, F., (2014).

<i>RESULTADOS DE INTERACCIÓN EN LAS PÁGINAS DEL EVA</i>		
<i>Páginas visitadas por el estudiante</i>	<i>Nro de estudiantes que visitan la página</i>	<i>Nro total de veces que visitan las páginas</i>
Página principal del usuario	98	12764
Páginas de información general	4	6
Páginas de actividades	112	4667
Páginas de comunicación y colaboración	4	4

Tabla 34. Indicadores estadísticos de las interacciones de los estudiantes en el EVA.

La información extraída a partir de las actividades que realizan los estudiantes cuando interactúan con el EVA, se la realizó mediante la aplicación de varias técnicas de minería, con la finalidad de buscar los indicadores de navegación de los estudiantes en la plataforma para representar un modelo de datos con los indicadores obtenidos.

Como se puede apreciar en la **Tabla 34**, la mayor parte de los estudiantes de la UTPL de noveno ciclo, acceden con mayor frecuencia a las páginas de las actividades (foros, tareas, recursos); por otra parte le sigue en el nivel de frecuencia de acceso por parte de los estudiantes a las páginas de comunicación y colaboración (Chat, twitter, mensajería, etc.), y por último tenemos el acceso a las páginas informativas como son (inicio, calendario, eventos, noticias, etc.), a las cuales el acceso por parte de los estudiantes es bien escaso.

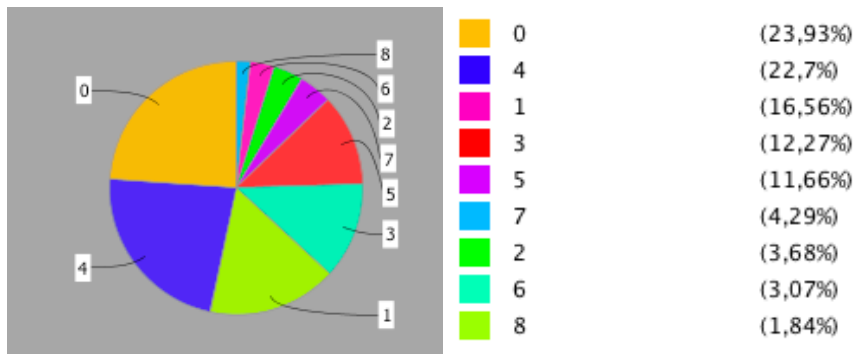


Fig. 24. Resultado de las interacciones realizadas en el EVA (% páginas frecuentadas)

Fuente: Ruiz, F., (2014).

Las páginas con mayor frecuencia por parte de los estudiantes en el EVA, se detallan en la **Fig. 24**, la gráfica muestra en detalle porcentual y mediante un diagrama circular seccionada con las diversas herramientas a las que accede del estudiante en una división porcentual de acceso según su frecuencia de navegación en las diferentes páginas del EVA.

En el análisis realizado con la interacción de los estudiantes de noveno ciclo de Informática de la UTPL en el EVA, se puede apreciar que de las tres actividades que los estudiantes utilizan para la comunicación con su profesor, con la que más interactúan es con los foros, en las tablas detalladas anteriormente se puede observar, que la mayoría de los estudiantes interactúan con el tutor a través de conversaciones en los foros de cada uno de los curso; considerando además de las demás actividades a las que el estudiante accede para su aprendizaje.

5.10. Validación de resultados obtenidos con las técnicas de minería frente a Gephi.

Para validar si los resultados obtenidos con las técnicas de minería se utilizó la herramienta Gephi para observar en que herramientas hay más interacción y corroborar con los datos obtenidos con la aplicación de las técnicas.

En la fase de la minería con las diferentes técnicas aplicadas se pretendió buscar datos de extracción de los patrones más significativos que tienen los estudiantes cuando navegan dentro de la plataforma, y a partir de ello, determinar mediante los atributos encontrados de las rutas más frecuentes de interacción con las páginas de la plataforma del EVA y a los indicadores representarlos en un modelo de datos, para que en estudios posteriores se pueda diseñar un nuevo prototipo del modelo de estudiante con los resultados obtenidos mediante la aplicación de los algoritmos de las diferentes técnicas aplicadas en esta tesis; técnica de Clustering, Reglas de asociación y Clasificación.

La idea de representar en un nuevo modelo de datos con los indicadores de navegación de los estudiantes, son con la finalidad de optimizar el acceso a las herramientas dentro del sistema que ayudan a la comunicación con el tutor y compañeros, y sobre todo al aprendizaje del estudiante de manera eficiente. A tratar de encontrar las características más comunes en los estudiantes cuando interactúan con el EVA en su aprendizaje.

Fuente: Ruiz, F., (2014).

<i>Accesos</i>	Resultados con la técnica de minería	Resultados con Gephi	Resultados consultas SQL
Páginas de Información general	483	479	483
Páginas de Actividades	3125	3127	3127
Páginas de Colaboración	4	4	4
Páginas de Apoyo	600	598	600

Tabla 35. Datos de Interacción de los estudiantes en los cursos.

En la **Tabla 35**, se puede apreciar los resultados obtenidos con las dos metodologías utilizadas para determinar la interacción de los estudiantes con la plataforma del EVA, en la primera metodología, es la de aprendizaje automático mediante técnicas de minería de datos, la cual presenta un mayor nivel de interacción con las herramientas pertenecientes a la categoría de **páginas de Actividades** (*foros, tareas, cuestionarios*), al igual que la metodología gráfica y estadística con un valor de 3125 accesos con las técnicas de minería, y 3127 accesos con Gephi; seguido están las **páginas de información general** (*perfil de usuario*) con un total de 483 accesos obtenidos con las técnicas de minería, y 479 con Gephi; luego está la categoría de las **páginas de Apoyo** (*recursos, utpl y subir archivos*) que presenta un valor de 600 con las técnicas de minería y 598 con Gephi; y finalmente esta la categoría de las **páginas de Colaboración y comunicación** (*twitter y blog*) con un total de 4 accesos en ambas técnicas.

Además se puede apreciar que a pesar de la variación de los datos en los resultados de las técnicas de minería con respecto a los obtenidos con Gephi, son similares en las páginas a las cuales acceden los estudiantes cuando interactúan con el sistema, y se obtuvieron resultados mediante consultas SQL en la base de datos de EVA para la validación de los resultados, en el cual se puede notar que la mayor parte de los valores obtenidos coinciden con los valores obtenidos de la aplicación de las técnicas de minería de datos.

Con la aplicación de las técnicas de aprendizaje automático y la herramienta Gephi para el proceso de minería de datos en los *Logs* de la base del EVA, se obtuvieron los valores mostrados en la **Tabla 35**, que muestra la comparativa de los resultados obtenidos de ambas técnicas.

Con los resultados obtenidos del análisis realizado con el minado mediante técnicas de e-Learning, como el análisis realizado con Gephi, sirven para que el docente puede definir las

herramientas con las que existe mayor aporte por parte de los estudiantes y de esta manera mejorar el rendimiento académico del estudiante explotando más las actividades con las que mejor se adapte el estudiante para su aprendizaje. Además sirve al docente para que determine los puntos más débiles del estudiante en el aprendizaje de los contenidos y prestar mas atención a las temáticas donde el estudiante requiere de retroalimentación; y por parte del estudiante para que preste más atención y refuerce por su parte con lectura o trabajos adicionales que le ayuden a mejorar su nivel de estudio y por ende el rendimiento académico.

CAPITULO VI:

REPRESENTACIÓN DE ATRIBUTOS E INDICADORES DE NAVEGACIÓN DEL USUARIO EN UN MODELO DE DATOS

6. Representación de indicadores y atributos en un modelo de datos

6.1. Introducción

El modelo a desarrollar debería tener en cuenta varios criterios conforme a las necesidades y los tipos de uso educativo que se realizar en la plataforma del EVA. Cinco aspectos deberían estar presentes en un buen entorno: *interacción, introspección, innovación, integración e información*, [Pallof & Pratt, 2001] sugieren tomar en cuenta estos cinco aspectos. La *interacción* tiene que ver con los aditamentos necesarios para establecer procesos comunicacionales sincrónicos y asincrónicos entre todos los participantes. La *introspección* se asocia con la posibilidad de que las herramientas comunicacionales permitan generar el desarrollo de pensamiento crítico y creativo por medio de una serie de recursos, en donde se acompañe de audio, video y texto, junto con espacios para actividades a desarrollar por parte de los estudiantes. La *innovación* permite al profesor generar actividades diferentes a las acostumbradas en una clase convencional, facilitando a los estudiantes participar cada vez más y sacar provecho de su propio estilo de aprendizaje, así como también ampliar la gama de alternativas para llevar adelante procesos de evaluación.

La *integración* tiene que ver con la manera en que los recursos tecnológicos incorporan la información a un proceso de generación de conocimientos, partiendo de una base relacional entre los participantes del curso. La *información* refiere a la posibilidad de que el estudiante cuente con hechos o datos necesarios para el buen desarrollo del curso, sin importar en qué lugar se encuentre esa información: bases de datos especializadas, bibliotecas digitales, etc.

Con los resultados obtenidos de la minería realizada a la base de datos del EVA presentado en el Capítulo VI: '**Minería de datos: Experimentación y Resultados**', se pretende buscar atributos de la interacción de los estudiantes con la plataforma y a partir de ello, obtener indicadores de navegación para representarlos en un modelo de datos del usuario.

6.2. Resultados del comportamiento de navegación de los estudiantes.

Para el proceso de extracción de patrones de interés en los estudiantes con las herramientas del EVA, se realizó el experimento con la materia en la cual los estudiantes interactúan más, la materia de '**Modelamiento de Datos [A]**' de noveno ciclo de la carrera de Informática de la UTPL, la cual tiene un total de 24 estudiantes enrolados, de los cuales todos han interactuado con algunas de las herramientas en el sistema. Los datos para la experimentación fueron obtenidos de la base de datos del EVA, la cual se la obtuvo mediante consultas en lenguaje SQL con la herramienta de *Navicat for MySQL* del periodo académico abril 2013 – agosto 2013.

Todos los estudiantes enrolados en la materia de **'Modelamiento de Datos [A]**, ingresaron al curso e interactuaron con varias de las herramientas dentro de la plataforma y realizaron varias actividades, las cuales se mencionan a continuación:

- ✓ Ingreso al curso en el que se encuentran enrolados, para revisar las Actividades propuestas por el profesor.
- ✓ Interacción con las herramientas y servicios disponibles en el curso dentro del EVA: foros, cuestionario, tareas, recursos, rol, descargas, perfil de usuario, utpl y blog.
- ✓ Los estudiantes ingresan a revisar dentro del curso las actividades que el profesor haya puesto a disposición del estudiante para su desarrollo.
- ✓ Durante el proceso de interacción en la herramienta de los foros los estudiantes revisaron los foros propuestos por el profesoro en el curso dentro de la plataforma.
- ✓ Participación de los estudiantes en la herramienta de los foros propuestos por el tutor.
- ✓ Conforme revisan los foros los estudiantes adicionan una discusión dentro de los foros ingresados por el profesor dentro del curso.
- ✓ Son pocos los estudiantes que postean o responden a las discusiones ingresadas por otros estudiantes en la herramienta de los foros.
- ✓ Los estudiantes antes de adicionar una discusión dentro del foro, primero tienden a revisar las discusiones ingresadas por otros estudiantes dentro del foro.
- ✓ Durante el proceso de interacción en la herramienta de los cuestionarios los estudiantes realizan un sólo intento en el desarrollo de los mismos antes de terminarlos.
- ✓ En base a los recursos subidos por el profesor, los estudiantes desarrollan las actividades.
- ✓ En base a las tareas subidas por el docente del EVA, los estudiantes desarrollan las tareas para luego subidas a la plataforma dentro del curso.

6.3. Interacción de los estudiantes en el EVA

Para determinar los indicadores de navegación de los estudiantes en las herramientas del EVA, se realizó la obtención de atributos a partir de factores como la interacción, el interés y la

colaboración que tienen los estudiantes, los cuales se muestran a continuación en la **Tabla 36**.

Fuente: Ruiz, F., (2014).

ATRIBUTOS	DESCRIPCIÓN
<i>Número de sesiones realizadas</i>	Indicador del número de veces que el estudiante ha ingresado a los cursos dentro del sistema.
<i>Número de accesos a los cursos</i>	Indicador del número de veces que los estudiantes ingresan a los cursos en los que están matriculados dentro del sistema.
<i>Número de accesos al perfil del usuario</i>	Indicador del número de veces que el estudiante ingresa a su perfil dentro del sistema.
<i>Número de accesos en los recursos.</i>	Indicador del número de veces que los estudiantes revisan los recursos ingresados por su tutor del curso dentro del sistema.
<i>Número de visitas a los foros.</i>	Indicador del número de veces que el estudiante ingresa y revisa los foros posteados por el tutor del curso.
<i>Número de posteos en los foros</i>	Indicador del total de veces que el estudiante crea un post en los foros.
<i>Número de veces que el estudiante revisa las tareas.</i>	Indicador del número de veces que el estudiante ingresa al curso a revisar las tareas propuestas por el tutor en el EVA.
<i>Número de veces que el estudiante sube una tarea</i>	Indicador de las veces que el estudiante ingresa a las tareas del curso para subir una tarea en el curso en el que se encuentra enrolado.
<i>Número de conversaciones que realiza en foros.</i>	Indicador del número de veces que el estudiante realiza un tema de discusión en los foros dentro de los cursos del sistema.
<i>Número de ingresos en el twitter</i>	Indicador del número de veces que el estudiante utiliza la herramienta del Twitter para comunicarse con el tutor y compañeros.
<i>Número de ingreso del estudiante en las tareas.</i>	Indicador del número de aportes que el estudiante ha subido al EVA.
<i>Número de visitas en los cuestionarios.</i>	Indicador del número de veces que el estudiante ingresa a revisar los quiz ingresados por el tutor.
<i>Número de intentos en los cuestionarios</i>	Indicador del número de intentos que el estudiante realiza en los exámenes subidos por el tutor en los cursos dentro del sistema.
<i>Número de ingresos al chat.</i>	Indicador de veces que el estudiante accede a enviar mensajes en el chat.
<i>Número de ingresos a la mensajería.</i>	Indicador de las veces que el estudiante ingresa a la página de la mensajería a enviar, leer o escribir mensajes dentro del curso.
<i>Número de ingresos al Blog.</i>	Nivel de ingresos para buscar ayuda en la herramienta del blog.

Tabla 36. Atributos estadísticos de las interacciones de los estudiantes en el EVA.

Estos atributos fueron extraídos de los datos de EVA para determinar a través de la navegación que tienen los estudiantes con las herramientas del EVA, los indicadores con las preferencias de los estudiantes para su aprendizaje.

6.4. Indicadores obtenidos en las actividades de aprendizaje

Los indicadores obtenidos de la minería están relacionados con las actividades académicas a desarrollar (*cursos, carreras, modalidades, comunidades virtuales, herramientas*) además las características de los destinatarios y de las posibilidades y limitaciones técnicas disponibles (*ancho de banda, equipamiento informático y disponibilidad de tiempo*). Existe gran cantidad de enfoques para llevar a cabo esta tarea. La mayoría de ellos están orientados a determinar las

características de los entornos en función de las categorías de análisis como características, requerimientos técnicos, herramientas para el aprendizaje del estudiante, herramientas para la gestión y administración, etc. Las categorías varían de un enfoque a otro, es muy común encontrar cierta recurrencia en las cuestiones centrales [Weiss y Indurkha, 1998].

Con los indicadores de navegación encontrados en el análisis de minado se los representara más adelante y poderlo representar en un modelo de datos del usuario.

Los indicadores obtenidos se los muestra en la **Tabla 37**, con el detalle de los indicadores.

INDICADOR	DESCRIPCIÓN	ATRIBUTOS
Nivel_interacción	Este indicador muestra el nivel de interacción que tienen los estudiantes con las herramientas de la plataforma del EVA.	Número de ingresos a los foro Número de ingresos a los cuestionarios Número de ingresos a los recursos Número de ingresos a los tareas Número de ingresos al perfil de usuario Número de ingresos al upload (subida de archivos) Número de ingresos al twitter Número de ingresos a los blogs.
Nivel_interes	Este indicador presenta el nivel de interés que tiene el estudiante con las materias que esta cursando según el número de accesos a cada una de las herramientas de la plataforma del EVA.	Número de veces que revisan los foros Número de veces que revisan los cuestionarios Número de veces que realizan los cuestionario Número de veces que revisan las tareas Número de veces que suben las tareas Número de veces que revisan los recursos.
Nivel_colaboracion	El indicador mide el nivel de colaboración que tienen los estudiantes en las herramientas con las cuales participa dentro de la plataforma del EVA.	Veces que visitan los foros. Número de blogs revisados. Número de conversaciones que el estudiante inicia. Número de conversaciones que el estudiante participa Número de veces que ingresan al twitter. Veces que el estudiante revisa los mensajes. Número de respuestas del estudiante en mensajes.

Tabla 37. Indicadores obtenidos de las interacciones de los estudiantes en el EVA.

Fuente: Ruiz, F., (2014).

Los indicadores establecidos para el modelado del usuario se los obtuvo a partir de los resultados obtenidos con la minería de la interacción de los estudiantes con las herramientas del EVA. A continuación se detalla los indicadores definidos para el modelado:

6.4.1. Nivel de interés en los cursos.

El nivel de interés que tienen los estudiantes dentro del curso con las herramientas y servicios de la plataforma, ayudan a medir el grado de importancia que tienen las actividades en el aprendizaje del estudiante y para su comunicación y colaboración con el docente, para lo cual se clasificó a los estudiantes en clases de interés (herramientas por las cuales el estudiante le gusta o prefiere interactuar) y por otra parte midiendo el nivel de interacción (herramientas en las cuales el estudiante ingresa a realizar las actividades). Estos indicadores se los obtuvo a partir de las aportaciones, posteos, descargas y cuestionamientos que los estudiantes han realizado en el curso en el que se encuentran enrolados dentro de la plataforma del EVA.

Los niveles planteados para la clasificación de los estudiantes con la interacción de las herramientas y servicios son: alto, medio y bajo.

El nivel de interés que los estudiantes a los cuales se les realizó la minería mediante técnicas de aprendizaje automático, se los obtuvo según los resultados obtenidos de la interacción con las herramientas en cada una de las técnicas de minería aplicadas, las cuales se detallan a continuación. Los niveles de interés que tienen los estudiantes dentro de las diferentes páginas clasificadas en la **Tabla 14**, en la cual se determinaron las páginas de las herramientas a las cuales accede el estudiante con mayor frecuencia, y en cada uno de los atributos en los datos obtenidos, los cuales serán evaluados con los niveles: bajo, medio y alto.

6.4.2. Nivel de interés en páginas de Actividades.

El nivel de interés que los estudiantes presentan en las herramientas dentro de las páginas de Actividades (*foros, cuestionario, tareas*) son las actividades de aprendizaje con mayor número de accesos e interacción, Los niveles de interés diseñados para la evaluación de este atributo en los datos obtenidos son: bajo, medio y alto.

6.4.3. Nivel de interés en páginas de Apoyo.

El nivel de interés que los estudiantes presentan con las herramientas dentro de las páginas de Apoyo (*recursos, subir archivos, utpl*) son las actividades de aprendizaje con un nivel menor de frecuencia de acceso que los anteriores, pero son páginas donde los estudiantes navegan e interactúan con gran frecuencia. Los niveles de interés clasificados para la evaluación de este atributo en los datos obtenidos son: bajo, medio y alto.

6.4.4. Nivel de interacción en los cursos.

El nivel de interacción que los estudiantes tienen dentro de los cursos en el cual se encuentran enrolados, se mide a través del uso e ingreso en las herramientas dentro de la plataforma con las cuales interactúan los estudiantes para su comunicación y aprendizaje; para lo cual se midió el número de accesos dentro del cursos, y las páginas que visito para realizar alguna acción sobre las actividades propuestas por su profesor, así como también el tiempo promedio de navegación en el sistema y el de interacción con las herramientas disponibles en el mismo. Los niveles de interacción estimados para evaluación son: principiante, medio avanzado y experto.

6.4.5. Nivel de colaboración en los cursos.

El nivel de colaboración que los estudiantes tienen dentro de los cursos se determina a través del uso e ingreso en las herramientas de comunicación y colaboración (*foros, chat, twitter, blog, etc.*) del EVA, y con las que interactúan los estudiantes para su comunicación y aprendizaje; se midió el número de accesos en las páginas y las acciones que realizó en cada una de ellas, de acuerdo a los resultados obtenidos de las técnicas aplicadas. Se consideraron los siguientes niveles de interacción estimados para evaluación: principiante, medio avanzado y experto.

6.5. Metodología Entidad – Relación.

Existe una gran variedad de técnicas para la representación o diseño de modelos de datos, pero por su sencillez, uno de los más utilizados y su popularidad se ha escogido para la representación del modelo de datos del usuario el modelo Entidad – Relación, el cual se basa principalmente en la percepción de objetos que se asocian entre ellos, los objetos tienen características llamados atributos. En el modelo E-R se presentaran los indicadores obtenidos del análisis realizado del proceso de minería, mediante un conjunto de tablas (relaciones) de la interacción del estudiante con las herramientas del EVA, para su aprendizaje y comunicación.

Por lo tanto, para diseñar de una manera adecuada los beneficios de las diferentes técnicas de modelaje, es necesario considerar dos factores sumamente importantes para conseguir este fin, y son *definir una metodología compleja y dificultosa la cual causará una actitud negativa en los usuarios* y el otro factor es *determinar el ambiente bajo el cual se utiliza dicha metodología*, ya que depende mucho el lugar donde se desarrollara un Modelo de Datos, es necesario que se encuentre bajo un ambiente informal con poca participación por parte de los usuarios y modeladores, y con mayor razón si los participantes no asumen responsabilidades concretas con el producto final que será el modelo de Datos.

Por lo tanto, para aplicar un modelo de datos se debe considerar primero los beneficios considerados anteriormente, y estos serán limitados si no se da una situación idónea tal como la describimos a continuación:

“Una metodología sencilla para el Modelo de Datos con reglas claras para su aplicación dentro de un marco de trabajo con participación y responsabilidades bien definidas”

Para desarrollar una metodología no muy compleja y de manera sencilla se parte de una consideración esencial:

“Los Modelos de Datos deben considerar única y exclusivamente aquellos elementos que

sean indispensables y su definición se planteará de forma sencilla, clara, no ambigua y entendible” para el desarrollo del modelo de datos se considero el modelo E-R, considerando los puntos mencionados, para lo cual se detalla a continuación una descripción de este modelo y de las fases que hay que tener en cuenta para su elaboración y diseño.

6.5.1. Modelo Entidad-Relación

Todo modelo tiene una representación gráfica, y por lo general se basa en percepciones del mundo real, el modelo más populares es el modelo entidad-relación o diagrama E/R, el cual está formado por objetos básicos, las entidades y estos a su vez poseen atributos y las relaciones entre estos objetos. El principal objetivo del modelado de datos es que permite representar las relaciones entre entidades.

Según [**Sosa O., 2010**] el modelo de datos está compuesto por:

- *Entidades.-* Objeto que existe y se distingue de otros objetos
- *Atributos.-* Característica interesantes sobre una determinada entidad.
- *Relaciones.-* Una asociación entre dos entidades.
- *Cardinalidad.-* El tipo de relaciones que puede establecerse entre entidades.
- *Llaves.-* Atributo o conjunto de atributos que no se repiten. Es única y minima.

Conjuntos de entidades y atributos

- *Entidades: todo lo que existe y es capaz de ser descrito (sustantivo).*
- *Atributos: es una característica (adjetivo) de una entidad que puede hacer 1 de tres cosas:*
 - Identificar
 - Relacionar
 - Describir

6.5.2. Elementos del esquema E-R

Según [**A. Jaime, 2005**] los nombres apropiados de los elementos del esquema Entidad-Relación se detallan de la siguiente manera:

- *Tipos de entidad: nombres en singular*
- *Tipos de vínculo: verbos*
- *Atributos: también son sustantivos*
- *Tipos de entidad: en mayúscula*
- *Tipos de vínculo: en mayúscula*

- *Atributos: primera letra en mayúscula*
- *Roles: en minúscula*
- *Tipos de vínculo: ordenado de izquierda a derecha (o de arriba hacia abajo) de manera que se pueda leer*

El modelo de datos *Entidad – Relación*, permite representar relaciones entre entidades abstraídas del mundo real, además presenta una fácil implantación y manipulación del modelo relacional durante el desarrollo del diseño e implementación de la metodología.

6.6. Modelo de datos del usuario con los indicadores obtenidos

El modelo de datos del usuario el cual se lo detalla más adelante, fue realizado en conjunto con un estudio similar del proyecto puzzle que trata de las acciones de los estudiantes que interactúan con las herramientas del EVA.

El modelo de datos del usuario presentado más adelante en la **Fig. 25**, está estructurada por varias tablas, como son: la tabla del modelo de usuario que es, de la cual se despliegan las demás tablas; las principales son (*mdl_user*, *mdl_course*) que son las tablas del EVA, y a partir de estas dos tablas parten las demás entidades del modelo, además dentro de la tabla del modelo de usuario se adiciona las tablas con los indicadores y en ellos cada uno de los atributos extraídos de la navegación de los estudiantes en el sistema obtenidos del análisis ya mencionado anteriormente.

Se realizaron dos tablas con los indicadores seleccionados; la primera es la tabla (*interés_curso*), la cual presenta los atributos de interés en las diferentes herramientas y actividades de los estudiantes durante su interacción con la plataforma, y la segunda tabla (*interacción_curso*), la cual describe los atributos del aporte que lo estudiantes realizan en las herramientas y servicios dentro de los cursos, y el nivel de interacción que tiene el estudiante con las herramientas del EVA, además la tabla (*colaboración_curso*), la misma que presenta los atributos del análisis de los resultados obtenidos de la aplicación de varias técnicas de minería, que son el nivel de participación que tienen los estudiantes con las herramientas de colaboración (foros, twitter y el blog). Estos indicadores y atributos fueron obtenidos a partir de la navegación de los estudiantes y accesos a las páginas del sistema.

Además, están las tablas (*utilización_herramientas* y *participación_curso*), que son indicadores obtenidos de un estudio paralelo que se realizó con un tema similar de tesis del proyecto puzzle; la cual describe las tablas (*acciones*, *objetos*, *estilos de aprendizaje*) y determinan el estilo de

aprendizaje de los estudiantes según las acciones realizadas en las páginas del EVA.

6.6.1. Modelo de usuario

Para el diseño del modelado de los datos del usuario, se consideró representar los indicadores y atributos de las rutas de navegación de los estudiantes, las cuales se las estableció según el nivel de interacción, interés y colaboración que presenta el estudiante en cada una de las actividades a realizar dentro de las herramientas del EVA, el cual se lo representa de la siguiente manera:

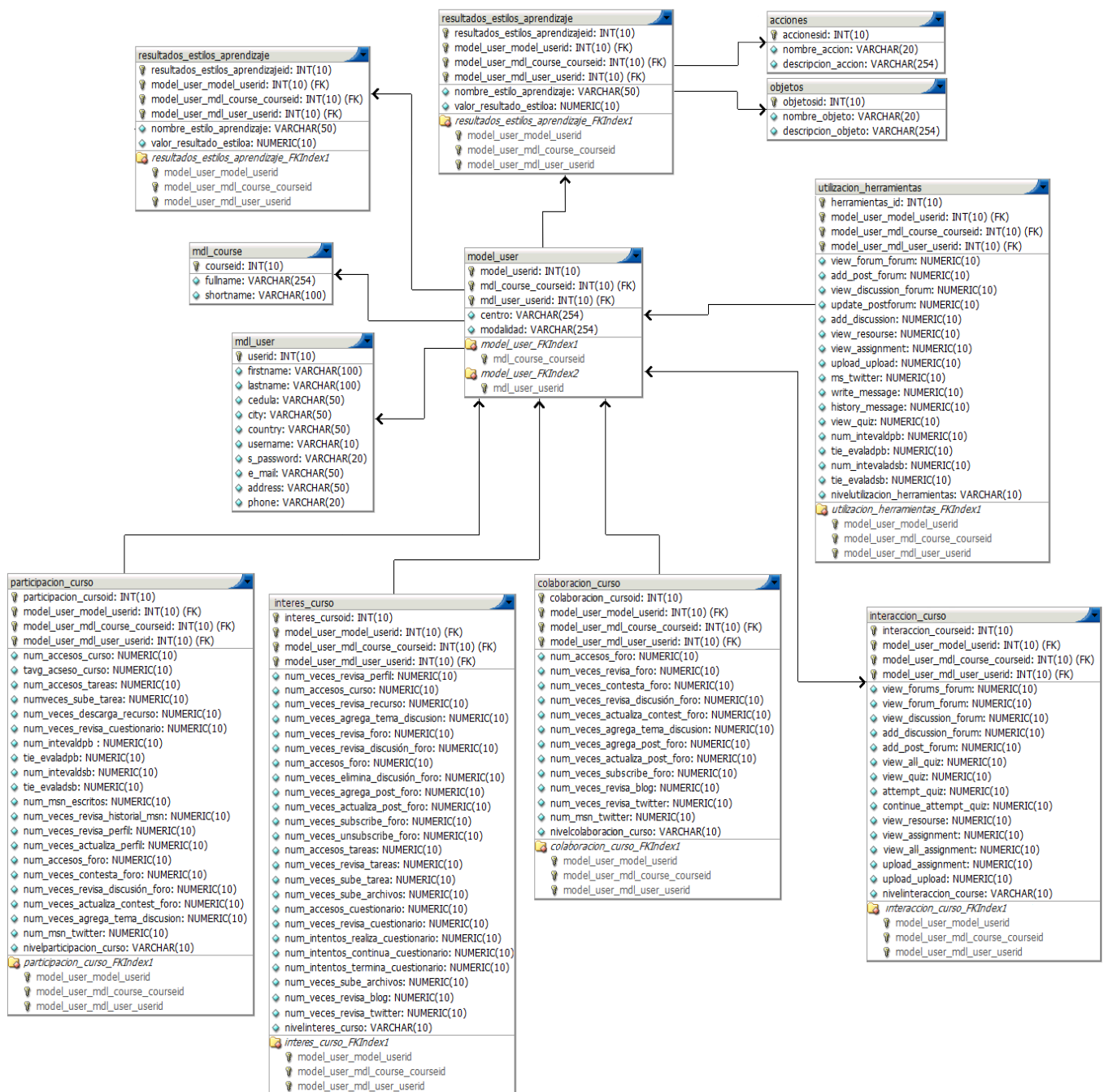


Fig. 25. Modelo del usuario con los indicadores de los estudiantes en el EVA.

Fuente: Ruiz, F., (2014).

El modelado del estudiante encontrado con los resultados obtenidos del minado de los datos, presentado en la **Fig. 25**, muestra el nivel de interacción, participación y colaboración que tienen los estudiantes con las herramientas del EVA, a partir de los niveles de participación se determinó el *nivel de interés y colaboración* que tiene en cada una; y a partir del nivel de ingresos se determinó el *nivel de interés* que tiene el estudiante.

DISCUSIÓN:

En el desarrollo de esta tesis se ha podido demostrar la gran utilidad que tiene la minería de datos al aplicarla a un caso como el que fue presentado en este trabajo. Entre varias de las ventajas que tiene la minería de datos, ayuda a determinar las rutas de navegación en entornos virtuales, ya que es de suma ayuda para extraer aquella información que le es de utilidad al estudiante para su aprendizaje y de gran utilidad a los docentes para la toma de decisiones en cuanto a los indicadores obtenidos de las herramientas de interés en el EVA, saber sus preferencias y formas de aprender mediante plataformas de aprendizaje virtual; además sirve de ayuda para que el profesor puede utilizar esta información de los resultados obtenidos del análisis realizado, con la finalidad de agrupar a los estudiantes según sus características y preferencias. Además contribuye para el aprendizaje del estudiante que el profesor determine la mejor forma de trabajo de los estudiantes con las actividades en el sistema, y de esta manera seleccionar el conjunto de actividades, ya sean de conocimiento (foros, tareas, cuestionarios, etc.), *Colaboración* (chat, mensajería, twitter, etc.) de refuerzo o apoyo (recursos, descargas, videoconferencias, tutorías, etc.) o *Informativas* (noticias, eventos, calendario de evaluaciones, etc.) y de acuerdo a estos resultados determinar según la función de sus características las actividades con las que el estudiante se desenvuelve y desarrolla un mejor nivel de aprendizaje.

La identificación de patrones de navegación en los estudiantes que interactúan con las herramientas disponibles en el EVA, ayuda a descubrir las preferencias que los estudiantes tienen con respecto a las actividades de desarrollo que el profesor pone a disposición para su aprendizaje. Para obtener estos patrones que permitan descubrir la información oculta dentro de cada uno de los estudiantes que hacen uso de las herramientas de la plataforma y navegan con a través de la misma, existe una gran variedad de técnicas que permiten identificar y clasificarlos según sus características de navegación o preferencias, es por lo cual se experimento con varias técnicas de minería para determinar de los resultados obtenidos de cada uno de ellos, cual es la que más se adapta a nuestras necesidades.

Luego de una investigación profunda de cada una de las técnicas que existen de minería de datos para la obtención de patrones de navegación se vio conveniente trabajar con tres técnicas de minería (dos algoritmos por cada técnica), cuatro fueron con algoritmos supervisados **Reglas de asociación** (*A priori*, *GSP*) y **Clustering**, (*Simple KMeans*, *EM*) y un algoritmo supervisado **Clasificación** (Árboles de decisión: *J48*); además se realizó tres experimentos con cada uno de los algoritmos propuestos, con la finalidad de según los resultados obtenidos determinar cual de ellos es más preciso y eficiente para este tipo de estudio.

Una vez realizada la extracción de los datos dentro de la base de datos de la plataforma del EVA, con la ayuda de la herramienta de Navicat for MySQL, se procedió con el minado con cada uno de los algoritmos propuestos con la herramienta Weka la cual dispone de todos los algoritmos mencionados anteriormente dentro de su conjunto de técnicas, se puede decir que los algoritmos más aptos para este tipo de estudio son las reglas de asociación y los árboles de decisión. Las reglas de asociación prácticamente lo que hace es encontrar asociaciones dentro de los datos de entrada y luego genera todos los ítems-sets con un elemento, y estos los genera con los de dos elementos y luego a ese lo usa para generar los de tres elementos y así hasta que queden sólo las posibles combinaciones que cumplan con las medidas mínimas del valor de soporte, para luego generar estas reglas revisando que cumplan con el criterio mínimo de confianza. Los algoritmos de clustering permiten clasificar a los estudiantes según las características y similitudes entre ellos con la finalidad de agruparlos por sus preferencias dentro de las herramientas dentro de la plataforma del EVA.

- La mayor parte de los estudiantes que navegan en el EVA, interactúan principalmente con las páginas de los foros, cuestionarios, tareas e ingresan a revisar los recursos que el tutor facilita al estudiante dentro del curso.
- Si el estudiante ingresa a revisar los foros propuestos por el tutor/profesor, por lo general, tiende a revisar las respuestas ingresadas de los foros o posts de otros compañeros dentro de los foros.
- Cuando los estudiantes ingresan a las páginas de las respuestas elaboradas por otros estudiantes dentro de los foros, tienden a dar respuesta al foro en el cual se encuentra inscripto, no todos se inscriben en el foro al cual van a dar respuesta.
- Los estudiantes que ingresan a resolver los cuestionarios propuestos por el profesor en el curso, en su mayoría sólo realizan un sólo intento y luego terminan el cuestionario.
- Si el estudiante sube una tarea al EVA dentro del curso en el que se encuentra enrolado, en su mayoría primero ingresan a revisar las tareas propuestas por el tutor/profesor.
- La mayor parte de los estudiantes revisan, leen o descargan las actividades ingresadas dentro de los recursos por su profesor antes de ingresar a subir algún archivo dentro de las tareas.
- Los foros son la primera opción con la cual los estudiantes interactúan cuando navegan por el sistema.
- En cuatro de los cinco cursos los estudiantes navegan con mayor frecuencia en las páginas que tienen que ver con la revisión, inscripción y posteo dentro de los foros.

- Las herramientas y servicios con las que hay casi una nula interacción de parte de los estudiantes son en los blogs, el chat, la mensajería y el twitter.
- Las páginas con mayor navegación de los estudiantes dentro de las herramientas y servicios son las páginas de las actividades (*foros, cuestionarios, tareas*) en el orden respectivo y páginas de apoyo (*recursos, subir archivos*) respectivamente.
- Las páginas con mínima o nula navegación de los estudiantes dentro del sistema son las páginas de información (*noticias, eventos, calendario, etc.*) y de colaboración (*chat, mensajería y twitter*).

Como una evaluación general del minado realizado con los algoritmos, que de todos los cursos de noveno ciclo de Informática, modalidad Abierta, en su mayoría los estudiantes acceden a las páginas donde los profesores navegan e interactúan, es decir, los estudiantes interactúan sólo con las herramientas en las que el profesor propone alguna actividad que realizar. Además, no interactúan con las herramientas de colaboración y comunicación con su profesor y demás compañeros como son (*chat, twitter, mensajería, etc.*)

CONCLUSIONES

Las conclusiones obtenidas del análisis de los resultados a partir de los experimentos realizados para determinar las rutas de navegación de los estudiantes y la interacción con las herramientas del EVA, y hace uso el estudiante para su aprendizaje y comunicación con el profesor.

- La implementación de técnicas de minería de datos aplicados a la educación, los cuales se relacionen con la navegación ayudan a descubrir los intereses que tienen los estudiantes con las diferentes herramientas y servicios dentro de la plataforma del EVA, con la finalidad de saber cuáles son las preferencias de los estudiantes para su aprendizaje y colaboración con el profesor y además permite determinar el nivel de aporte que el estudiante tiene con cada una de las herramientas del sistema.
- Una vez realizados los experimentos de comparación con varios de los algoritmos aplicados de las diferentes técnicas de minería, se puede decir que cada una de las técnicas tienen tanto ventajas como desventajas con respecto al dominio de aplicación. Es decir, en el caso de la técnica de *Clasificación*, con el algoritmo de árboles de decisión (J-48) se obtuvo una excelente capacidad de clasificación de los datos de entrada en comparación a otros algoritmos, como son los de reglas de asociación.
- El algoritmo con resultados más óptimos y eficientes para realizar este tipo de análisis es el Algoritmo **GSP** y **J48**, ya que son algoritmos para determinar las rutas de navegación que tienen los estudiantes durante la interacción con las herramientas del EVA, además son algoritmos que determinan el número total de accesos que el estudiante realiza en cada una de las páginas de la plataforma.
- Con el algoritmo de árboles **J-48**, se obtuvieron reglas de clasificación según las acciones que el estudiante realiza en las herramientas dentro del conjunto de entrenamiento (archivos log); entre los cuales existen varios elementos en común, pero con variación en los tiempos de ejecución y la clasificación de los conjuntos. Este algoritmo es muy sensible en cuestión de los cambios que se realizan en los datos del conjunto de entrenamiento, debido a lo complicado que puede ser determinar, a partir de la “integración” de modelos de minería de datos correspondientes a fuentes de datos homogéneas, un modelo de minería de datos válido para el conjunto formado por la unión de todas las fuentes de datos.
- En los resultados obtenidos de la implementación de la técnica de Clustering, se obtuvieron datos similares entre grupos, con la variación de un clúster y el tiempo de ejecución; en el

resultado de los algoritmos de las Reglas de asociación los resultados si variaron un poco más, a diferencia del algoritmo GSP, el cual devuelve todas las secuencias de ingreso por parte de los estudiantes en determinadas páginas, lo cual no pasa con el algoritmo de A-priori, ya que este algoritmo devuelve sólo la acción realizada pero no el total de accesos que el estudiante ha realizado en determinada página.

- Con el algoritmo A priori las reglas que se obtuvieron a partir de los conjuntos de entrenamiento disjuntos se los realizó con valores de soporte y confianza diferentes, para buscar los resultados más óptimos, en los experimentos realizados se obtuvieron similares características en los resultados obtenidos en el conjunto de datos centralizados. En estos experimentos es posible determinar un modelo de minería de datos válido para el conjunto formado por la unión de todas las reglas obtenidas.
- La técnica de Clustering ayuda a agrupar a los estudiantes con la finalidad de encontrar características similares en el comportamiento que tienen cuando interactúan con las herramientas del sistema, pero no ayuda a determinar las rutas de navegación que tienen los estudiantes dentro de la plataforma.
- El resultado obtenido con las técnicas de minería de datos y la herramienta Gephi en relación a los accesos, son similares en su mayoría, ambas metodologías arrojaron que el curso con mayor nivel de interacción es en la asignatura 'Modelamiento de datos [A]', además de determinar que las páginas con mayor nivel de frecuencia o accesos durante la navegación del estudiante en el sistema es en la categoría de páginas de Actividades correspondientes a los foros, cuestionarios y tareas.
- Los estudiantes con mayor número de interacciones en las herramientas de la plataforma son los enrolados en las materias de: Modelamiento de datos, Sistemas basados en el conocimiento y Teoría de colas.
- Los cursos con menor número de accesos e interacción con las herramientas de la plataforma son los estudiantes pertenecientes a las materia de Administración de herramientas CASE y Planeación estratégica.
- Las herramientas más utilizadas por los estudiantes cuando interactúan con la plataforma son: Foros, Cuestionarios, Recursos, Tareas, y Subida de archivos.

- Con la implementación de técnicas de minería Web en los datos de interacción de los estudiantes en la plataforma se puede determinar el nivel de aporte que brindan las herramientas y servicios en el aprendizaje, la comunicación y relación del estudiante con el profesor dentro de los cursos en los que se encuentran enrolados.
- Las técnicas visuales mediante grafos de la interacción que tienen los estudiantes con las herramientas de la plataforma del EVA, ayudan en la interpretación y análisis de los resultados obtenidos dentro de cada una de las herramientas para determinar el nivel de aporte que tienen los estudiantes en los cursos.
- Todos los algoritmos con los que se experimentó determinan que la herramienta con la cual los estudiantes tienen un nivel mayor de interacción y contribución es en los Foros, la cual corresponde a las páginas de actividades con un porcentaje total del 72%, seguido esta los cuestionarios (quiz) con un total del 27%, y luego esta la herramienta Recursos con un total del 15%, y finalmente están las tareas con un 8%, el blog con 0,7% y el twitter con el 0,3%.

RECOMENDACIONES

Las recomendaciones que se presentan a continuación, se las obtuvo a partir del análisis realizado en la presente tesis y según los resultados obtenidos de cada uno de los algoritmos implementados para el minado de los cursos.

- Las reglas de asociación clásicas son recomendadas para trabajarse en bases de datos transaccionales, y si es posible con atributos de tipo booleanas; sin embargo, los atributos de tipo categóricos o numéricos, en la minería Web presentan buenos resultados en cuestión de análisis de rutas de navegación o para determinar patrones de navegación o comportamiento de los estudiantes dentro de la plataforma, y a partir de ello si es necesario se puede reestructurar el contenido del sistema de acuerdo a la información obtenida del análisis.
- Del trabajo realizado se puede determinar que no es muy conveniente la aplicación de los algoritmos de reglas de asociación en bases de datos de gran tamaño; por lo menos en este tipo de estudios, o lo que se tenga que ver con búsqueda de patrones (navegación, acciones, interés, interacción, etc.), ya que esto puede resultar complejo y complicado en términos de tiempo y espacio; la obtención de un conjunto de reglas puede ser demasiado amplio y la interpretación de las reglas obtenidas para establecer su utilidad sería una labor muy dispendiosa inclusive para un experto en el tema, además que las reglas obtenidas no pueden ser muy claras y precisas.
- Las reglas de asociación difusas presentan algoritmos con los que se pueden alcanzar mejores representaciones, para obtener patrones de comportamiento de los usuarios durante su interacción con la plataforma del EVA.
- Se recomienda que el profesor interactúe con las diferentes herramientas y servicios de la plataforma del EVA, ya que se determinó que según el análisis de los resultados obtenidos los estudiantes acceden a las páginas e interactúan con las herramientas con las cuales el tutor trabaja; por lo cual se recomienda que el tutor trabaje con la mayor parte de la servicios disponibles para que el estudiante tenga un mayor nivel de aporte con las mismas en la plataforma.
- Según los resultados obtenidos con los algoritmos de Clustering (**SimpleKMeans y EM**) se pudo determinar que para realizar minería de datos aplicados a la educación es conveniente realizarlo por asignaturas, actividades o herramientas; ya que de esta manera se pueden obtener resultados más eficientes, debido a que el algoritmo implementado va a

agrupar a los estudiantes con mayor consistencia y similitud entre sus características e intereses; ya que al realizarlo por cursos no presentara datos muy consistentes debido a que no todos los estudiantes tienen un comportamiento de navegación similar, por ello el modelo presentara resultados poco confiables.

- Se recomienda trabajar con algoritmos de clustering (***SimpleKMeans y EM***) para agrupar a los estudiantes con características e intereses similares y con ello iniciar un aprendizaje colaborativo basado en grupos dentro de un determinado curso, ya que esta técnica permite descubrir patrones en los estudiantes que muestran comportamientos análogos.
- Ya que las técnicas de clasificación (***J-48***) permiten extraer reglas del árbol de decisión, y agrupar estudiantes con similares características, se recomienda utilizar este modelo para predecir el rendimiento de los estudiantes según la interacción con las herramientas y servicios disponibles dentro del curso, y con ello guiar la actividad del estudiante.

TRABAJOS FUTUROS

El análisis del presente trabajo se puede considerar en trabajos futuros en el área de aplicación de minería de datos en entornos e-learning, ya que son pautas de interés para los profesores, tutores y estudiantes involucrados en el aprendizaje-enseñanza:

- ✓ En base al modelo de datos con los indicadores de comportamiento en la navegación del estudiante propuesto en este trabajo, el profesor podría tomar esta información como referencia que le permita predecir, en base al uso de las herramientas el interés del estudiante en las actividades de las asignaturas.
- ✓ Anexión de nuevos indicadores de navegación de los estudiantes con las demás herramientas con las que dispone el entorno del EVA, que no se hayan considerado en el este trabajo, para con ello, obtener indicadores de rendimiento en los estudiantes.
- ✓ Realizar un estudio con los indicadores obtenidos para saber en qué medida influye la navegación de los estudiantes en el EVA para su aprendizaje, y por ende en el rendimiento académico.
- ✓ Se pueden realizar un análisis comparativo propuesto en otros estudios similares, con algoritmos iguales o más eficientes en cuanto a tiempo de ejecución y resultados, para determinar cuál es el mejor candidato para mostrar reglas de asociación con datos relativos en entornos e-learning.

REFERENCIAS BIBLIOGRÁFICAS:

[Arteaga C.; Fabregat R.; Mérida D., s. a.] Carlos Arteaga, Ramón Fabregat, David Mérida, 'Soporte Adaptativo al Aprendizaje Colaborativo e Individual (ASCIL): Un concepto Integral'. Instituto de Informática y Aplicaciones, Universitat de Girona.

[Alcívar, Idróvo & Macas, 2007] Alcívar Zambrano Patricio X., Idróvo Chiriboga Fanny E., Macas Pizarro Víctor H. "Sistema de análisis de patrones de navegación utilizando minería Web", Tesis de grado. Guayaquil – Ecuador.

[Agrawal, R. & R. Srikant]. Agrawal, R. and R. Srikant. *Fast Algorithms for Mining Association Rules in Large Databases*. In *Proceedings of the 20th International Conference on Very Large Data Bases*. 1994. Santiago de Chile, Chile: Morgan Kaufmann.

[Berlanga, 2006] Adriana José Berlanga Flores, "Diseños instructivos adaptativos: Formación personalizada y reutilizable en Entornos educativos", Tesis Doctoral. Universidad de Salamanca - DEPARTAMENTO DE INFORMÁTICA Y AUTOMÁTICA.

[Brusilovsky, 1996] Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User Adapted Interaction*, 6(2-3), 87-89.

[Brusilovsky, 1998] Pedro Brusilovsky y Pesin Leónidas. *Apoyo a la navegación de adaptación de la Educación Hipermedia: Una evaluación de la ISIS-tutor* "en Revista de Informática y Tecnologías de la Información, 6 (1) ,27-38, 1998. 1998.

[Castaño A., 2009] Castaño P. Andrés P. 'MINERÍA DE USO WEB PARA LA IDENTIFICACIÓN DE PATRONES'. Ingeniero electricista, docente Universidad de Caldas. Estudiante del doctorado en Ingeniería Informática de la Universidad Pontificia de Salamanca. Edición Vector, Volumen 4, Enero - Diciembre 2009

[Castells 2001] Castells P., Macías, J. A.: *An Adaptive Hypermedia Presentation Modeling System for Custom Knowledge Representations*. Actas World Conference on the WWW and Internet (WebNet'2001). Orlando, Florida, USA, 2001, pp. 148-153.

[De Bra, 1997] De Bra Pablo y Calvi Licia. "Creación de Hiperdocumentos adaptativa y en la Web" en Actas de la Conferencia WebNet AACE, Toronto, 149-155, 1997.

[Fuentes S. & Ruiz M., 2007] Lic. Sady C. Fuentes Reyes. e Ing. Marina Ruiz Lobaina, 'Minería Web: un recurso insoslayable para el profesional de la información' ,Instituto de Información Científica y Tecnológica (IDICT). Capitolio de La Habana.

[García M. & Álvarez A., s.a] María García Jiménez, Aránzazu Álvarez Sierra. Artículo: “*Análisis de datos en Weka – pruebas de Selectividad*” Universidad Carlos III - Ingeniería de Telecomunicación, s.a.

[Gaudioso, 2002] E. Gaudioso, *Contribuciones al Modelado del Usuario en Entornos Adaptativos de Aprendizaje y Colaboración a través de Internet mediante técnicas de aprendizaje Automático*. PhD thesis, Universidad Nacional de Educación a Distancia. Madrid, 2002.

[Gaudioso E. & Boticario J., 1997] Elena Gaudioso and Jesús G. ‘*Towards web-based adaptive learning communities*’ Boticario Artificial Intelligence Department Facultad de Ciencias, UNED c/ Senda del Rey, 9 28040 Madrid, Spain.

[Hernández, 2004] Hernández Orallo J.; Ramírez Quintana M. José y Ferri Ramírez C. (2004) *Introducción a la minería de Datos*. Editorial Pearson, 2004. ISBN: 84 205 4091 9

[Hernández & Ferri,] *Minería de Datos y Extracción de Conocimiento de Bases de Datos. , Extracción Automática de Conocimiento en Bases de Datos e Ingeniería del Software*

[Jaideep, 2000] Jaideep Srivastava, Robert Cooley, Mukund Deshpande, Pang-Ning Tan. “*Minería del uso de la Web: Discovery and Applications of Usage Patterns from web Data*”, 2000.

[Londoño, 2011] Iván Londoño R. (11 de 02 de 2011). Introducción a la minería de datos. Recuperado el 06 de 25 de 2009, de crónicas de Sofía: <http://cronicasdesofia.blogspot.com/2009/06/que-es-la-mineria-de-datos.html>

[Moreno M.; Miguel L.; García F.; Polo M., s.a] María N. Moreno García*, Luis A. Miguel Quintales, Francisco J. García Peñalvo y M. José Polo Martín, ‘*APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS EN LA CONSTRUCCIÓN Y VALIDACIÓN DE MODELOS PREDICTIVOS Y ASOCIATIVOS A PARTIR DE ESPECIFICACIONES DE REQUISITOS DE SOFTWARE*. Universidad de Salamanca. Departamento de Informática y Automática

[Koch, 2000] Koch, N. (2000). *Software Engineering for Adaptive Hypermedia Systems. Reference Model, Modeling Techniques and Development Process*. Unpublished PhD Thesis, Ludwig-Maximilians- Universität München, Munich.

[Pallof & Pratt, 2001] Pallof R. M. and Pratt K. *Lessons from the cyberspace classroom*. San Francisco: Jossey-Bass. 2001.

[Romero, Ventura & de Castro, 2005] Romero C.; Ventura S.; de Castro C. y García E. (2005, Julio - Diciembre). *Algoritmos Evolutivos para descubrimiento de Reglas de Predicción en la*

Mejora de más Educativos Adaptativos Basados en Web. ADIE, Asociación para el Desarrollo de la Informática Educativa ISSN: 1699-4574, pp 47-60.

[Romero, Ventura & Hervás, 2009] Romero Morales C.; Ventura Soto S. y Hervás Martínez C. (CICA 2009) *Estado actual de la aplicación de la minería de datos a los más de enseñanza basada en web*. Disponible en URL:

http://www.investigacion.frc.utn.edu.ar/labsis/Publicaciones/congresos_labsis/cynthia/CICA_2009_Aplicación_Minería_Datos_basada_ense%C3%B1anza_web.pdf [consulta 11 de Febrero de 2011].

Biblioteca TechNet *Algoritmos de Minería de Datos (Analysis Services: Minería de Datos)* Disponible en URL: <http://technet.microsoft.com/es-es/library/ms175595.aspx>

[Rice, W. H., 2006] Rice, W.H. (2006) Moodle E-learning Course Development. *A complete guide to succesful learning using Moodle*. Packt publishing.

[Sánchez, 2008] Sánchez Enríquez, Heider Ysaías, *Aplicación en Minería de Datos, Web Mining*, 19 de marzo de 2008.

[Valenzuela R., 2006] Ricardo Valenzuela Gaete “*Aplicaciones del Soft Computing al análisis de ficheros log de sitios Web*” Universidad de Granada, Editorial de la Universidad de Granada, 25 de Octubre de 2006.

[Weiss y Indurkha, 1998] Weiss, S.M. y Indurkha, N. “*Predicitve Data Mining. A Practical Guide*”, Morgan Kaufmann Publishers, San Francisco, 1998.

[Witten; Frank; Trigg; Hall; Holmes & Cunningham, 1999] Ian H. Witten; Eibe Frank, Len Trigg, Mark Hall, Geoffrey Holmes, and Sally Jo Cunningham (1999). «*Weka: Practical Machine Learning Tools and Techniques with Java Implementations*». Proceedings of the ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems págs. 192-196. Consultado el 26-06-2007.

ANEXOS

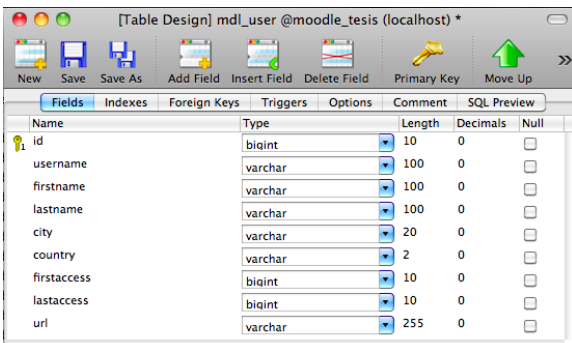
ANEXO A

TABLAS MÁS IMPORTANTES DE MOODLE PREVIA A LA MINERÍA DE DATOS

Se detallan las tablas de la base de datos del EVA más importantes que fueron seleccionadas para la minería de datos, las mismas que contienen la información del modelo de estudiante respecto a las actividades y acciones que el estudiante realiza cuando navega por el EVA.

Tabla user

Esta tabla contiene todos los datos de los estudiantes, que se encuentran matriculados en el periodo lectivo que va a ser analizado, los datos considerados para la minería son: Id del usuario, primer acceso al EVA y el último acceso; estos datos permiten determinar la secuencia de entrada y salida del estudiante durante la interacción con el EVA.



Name	Type	Length	Decimals	Null
id	bigint	10	0	<input type="checkbox"/>
username	varchar	100	0	<input type="checkbox"/>
firstname	varchar	100	0	<input type="checkbox"/>
lastname	varchar	100	0	<input type="checkbox"/>
city	varchar	20	0	<input type="checkbox"/>
country	varchar	2	0	<input type="checkbox"/>
firstaccess	bigint	10	0	<input type="checkbox"/>
lastaccess	bigint	10	0	<input type="checkbox"/>
url	varchar	255	0	<input type="checkbox"/>

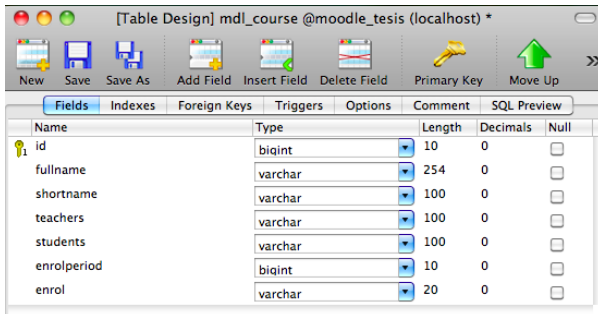
Fig. 26. Tabla del curso con los datos de la entidad mdl_user.

Fuente: Ruiz, F., (2014).

Sobre esta tabla se realizarán consultas SQL, de tal manera que me permite determinar el Id y nombres de los usuarios matriculados en los cursos que serán analizados previo a la minería de datos.

Tabla courses

Esta tabla contiene los datos de los cursos que existen en el EVA, los cuales, permiten analizar determinadas materias que los estudiantes están cursando durante el periodo de estudio, para el análisis de interacción de los estudiantes con esas materias. La información que se seleccionó de estos cursos son: el id del curso, la categoría a la que pertenece dicho curso, y las personas que la toman a esta materia, en este caso los estudiantes que cursan las materias que serán analizadas.



Name	Type	Length	Decimals	Null
id	bigint	10	0	<input type="checkbox"/>
fullname	varchar	254	0	<input type="checkbox"/>
shortname	varchar	100	0	<input type="checkbox"/>
teachers	varchar	100	0	<input type="checkbox"/>
students	varchar	100	0	<input type="checkbox"/>
enrolperiod	bigint	10	0	<input type="checkbox"/>
enrol	varchar	20	0	<input type="checkbox"/>

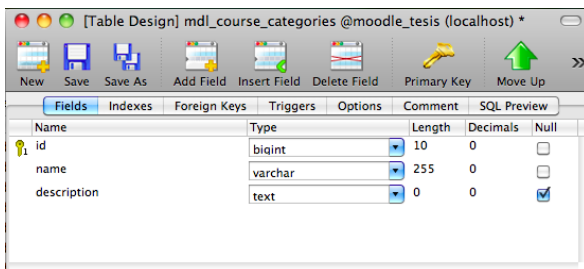
Fig. 27. Tabla del curso con los datos de la entidad mdl_course.

Fuente: Ruiz, F., (2014).

Sobre esta tabla se realizarán consultas SQL, de tal manera que me permite determinar el Id y nombre completo del curso que será analizado previo a la minería de datos.

Tabla de course_categories

Los datos almacenados en esta tabla, permiten determinar la categoría a la que pertenecen los cursos que los estudiantes están cursando, y el Número de cursos que el estudiante esta tomando en este periodo de estudio.



Name	Type	Length	Decimals	Null
id	bigint	10	0	<input type="checkbox"/>
name	varchar	255	0	<input type="checkbox"/>
description	text	0	0	<input checked="" type="checkbox"/>

Fig. 28. Tabla del curso con los datos de la entidad mdl_course_categories.

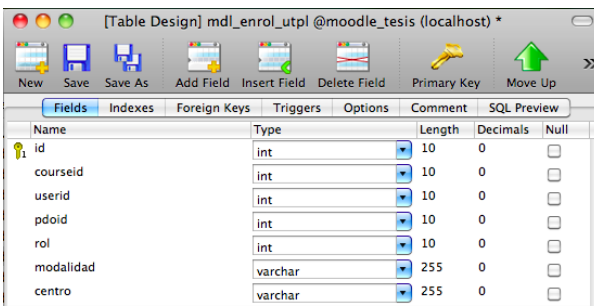
Fuente: Ruiz, F., (2014).

Sobre esta tabla se realizarán consultas SQL, de tal manera que me permite filtrar de todos los cursos existentes en esta tabla los datos del curso que se analizara, y determinar el Id de la modalidad del curso que será analizado, para mediante este dato poder realizar la búsqueda del curso que le queremos realizar la minería con los datos de interés del mismo.

Tabla enrol_utpl

En esta tabla se almacena la información del id del curso, el id del estudiante, el id de la modalidad a la que pertenece el estudiante, el rol que, es decir (número 5) estudiante, el nombre de la modalidad y el centro en el que esta estudiando (Loja). Esta información permite discernir sólo los datos del estudiante que serán analizados, ya que en la base de

datos del EVA, existe una gran cantidad de infamación, ya sea modalidades, centros, estudiantes, cursos etc.



Name	Type	Length	Decimals	Null
id	int	10	0	<input checked="" type="checkbox"/>
courseid	int	10	0	<input type="checkbox"/>
userid	int	10	0	<input type="checkbox"/>
pdoid	int	10	0	<input type="checkbox"/>
rol	int	10	0	<input type="checkbox"/>
modalidad	varchar	255	0	<input type="checkbox"/>
centro	varchar	255	0	<input type="checkbox"/>

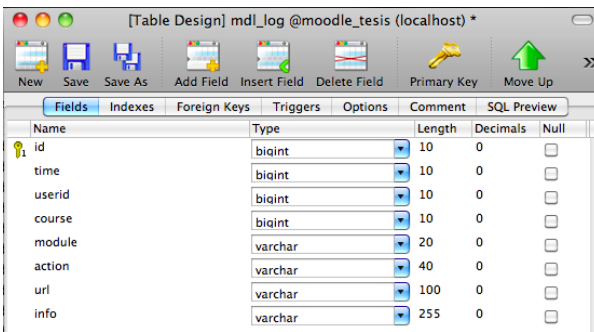
Fig. 29. Tabla del curso con los datos de la entidad mdl_enrol_utpl.

Fuente: Ruiz, F., (2014).

Sobre ésta tabla se realizarán varias consultas SQL, de tal manera que permite asociar tablas con información en común, y así poder filtrar sólo los datos de interés para el presente trabajo, es decir, información de estudiantes que navegan dentro de los cursos, y a partir de ello determinar el Id del estudiante combinados de las tablas en común, el centro el rol y la modalidad del curso en la que se encuentra enrolado.

Tabla Log

La tabla de Logs, es la tabla más importante para el análisis, ya que contiene de una manera resumida todos los eventos y actividades realizadas por los estudiantes cuando interactúan con las herramientas del EVA, los datos que se consideraron para el desarrollo de minería son: el tiempo (fecha que el estudiante realiza las acciones sobre el EVA), el id del usuario, el curso en el que esta enrolado, la herramienta a la que ingresó y por último la acción que realizó.



Name	Type	Length	Decimals	Null
id	bigint	10	0	<input checked="" type="checkbox"/>
time	bigint	10	0	<input type="checkbox"/>
userid	bigint	10	0	<input type="checkbox"/>
course	bigint	10	0	<input type="checkbox"/>
module	varchar	20	0	<input type="checkbox"/>
action	varchar	40	0	<input type="checkbox"/>
url	varchar	100	0	<input type="checkbox"/>
info	varchar	255	0	<input type="checkbox"/>

Fig. 30. Tabla del curso con los datos de la entidad mdl_log.

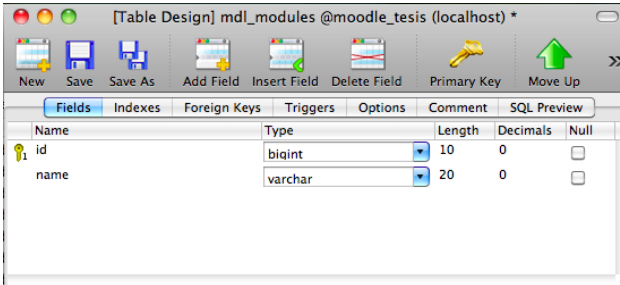
Fuente: Ruiz, F., (2014).

Sobre esta tabla se realizarán consultas SQL, que permitan determinar los estudiantes que navegan por el EVA y las acciones que realizan en las materias que están matriculados, y

asociar tablas afines para que no se repitan los datos de la consulta de cada uno de los usuarios matriculados en los cursos que serán analizados previo a la minería.

Tabla modules

En esta tabla se almacenan datos de las herramientas de las que disponen tanto los estudiantes como los tutores que manejan el EVA.



The image shows a screenshot of a database table design tool. The window title is "[Table Design] mdl_modules @moodle_tesis (localhost)". The interface includes a toolbar with icons for New, Save, Save As, Add Field, Insert Field, Delete Field, Primary Key, and Move Up. Below the toolbar are tabs for Fields, Indexes, Foreign Keys, Triggers, Options, Comment, and SQL Preview. The main area displays a table with two columns: 'id' and 'name'. The 'id' column has a type of 'bigint', a length of 10, 0 decimals, and is marked as the primary key. The 'name' column has a type of 'varchar', a length of 20, 0 decimals, and is not marked as the primary key.

Name	Type	Length	Decimals	Null
id	bigint	10	0	<input type="checkbox"/>
name	varchar	20	0	<input type="checkbox"/>

Fig. 31. Tabla del curso con los datos de la entidad mdl_modules.

Fuente: Ruiz, F., (2014).

Los datos tomados de esta tabla son: el id, y el nombre, de la cual disponen los estudiantes para interactuar con los demás estudiantes y su tutor. Estos datos permiten saber las herramientas que más utiliza el estudiante durante su navegación en el EVA, para su aprendizaje.

ANEXO B

EXTRACCIÓN DE DATOS DEL EVA, PARA EL PROCESO DE MINERÍA.

Debido a la gran cantidad de información dentro de la base de datos del EVA; se tomaron varios datos de las misma, los cuales sean representativos para el análisis que deseamos realizar dentro de la misma, por medio de **consultas SQL**, estas tablas son previas a la minería de datos para la obtención de los patrones de navegación de los estudiantes que interactúan con el EVA, mediante las herramientas de las cuales dispone con su tutor y estudiantes.

PERIODO LECTIVO

SELECT

```
mdl_periodo_utpl.id, mdl_periodo_utpl.pdoid, mdl_periodo_utpl.name
```

FROM

```
mdl_periodo_utpl
```

WHERE

```
mdl_periodo_utpl.pdoid = '17'
```

Esta consulta permite determinar el nombre y e Id del periodo lectivo en el cual están almacenados los datos de las materias a las cuales se les realizará la minería, ya que dentro de la base de datos hay varios datos de distintos periodos lectivos.

ACCIONES QUE LOS ESTUDIANTES PUEDEN REALIZAR EN EL EVA

SELECT

```
mdl_log.userid, mdl_log.course, mdl_log.module, mdl_log.action, mdl_log.info, mdl_enrol_utpl.rol
```

```
FROM mdl_log
```

```
INNER JOIN mdl_enrol_utpl ON mdl_enrol_utpl.courseid = mdl_log.course
```

```
WHERE mdl_log.course = 51750
```

```
AND mdl_enrol_utpl.rol = 5
```

action
add
add comment
add contact
add discussion
add entry
add message
add mod
add post
add_planutpl
assign
attempt
block contact
category_add
change password
close attempt
comment
continue attemp
course_category_s
course_new
delete
delete attempt
delete discussion
delete entry
delete mod
delete post
derive
duplicate
edit
editquestions
editquestion

Fig. 32 Tabla con las acciones que puede realizar el estudiante y tutor en el EVA.

Fuente: Ruiz, F., (2014).

Esta tabla muestra las acciones que los estudiantes realizan sobre el EVA, en las diferentes herramientas a las que acceden cuando navegan en los cursos que se encuentran enrolados e interactúan dentro del EVA; de los cuales se tomarán sólo las acciones más concurrentes por parte de los estudiantes y tutor para su comunicación y colaboración con las herramientas que dispone la plataforma del EVA.

MATERIAS ESCOGIDAS PARA LA MINERÍA

Las materias escogidas para realizar la minería de datos es noveno ciclo de Informática de la Modalidad a Distancia del periodo académico *Abril 2013 – Agosto 2013*.

A continuación se detalla cada una de las consultas SQL realizadas a cada una de las materias para filtrar la información de los estudiantes que cursan estas asignaturas y las acciones que realizaron en el mismo, durante su interacción con el EVA.

ADMINISTRACIÓN DE HERRAMIENTAS CASE

Para la extracción de los datos de las materias a las cuales se les realizará la minería se los realizó de forma manual mediante consultas en el lenguaje SQL, como se detalla a continuación.

Los datos del curso se obtienen mediante la siguiente consulta SQL a la base de datos:

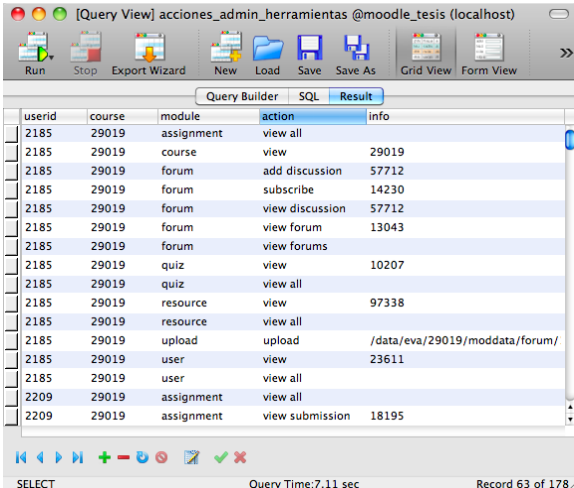
SELECT DISTINCT

```
mdl_course.id, mdl_course.fullname, mdl_enrol_utpl.pdoid, mdl_enrol_utpl.modalidad
FROM mdl_course
INNER JOIN mdl_enrol_utpl ON mdl_enrol_utpl.courseid = mdl_course.id
WHERE mdl_course.id = 51749
```

Los datos de los estudiantes enrolados en este curso y las acciones que realizan dentro del mismo, se obtienen mediante la siguiente consulta SQL a la base de datos:

SELECT

```
mdl_log.userid, mdl_log.course, mdl_log.module, mdl_log.action, mdl_log.info
FROM mdl_course
INNER JOIN mdl_log ON mdl_log.course = mdl_course.id
WHERE mdl_log.course = 51749
GROUP BY mdl_log.userid, mdl_log.module, mdl_log.action ASC
```



The screenshot shows a database query result window titled "[Query View] acciones_admin_herramientas @moodle_tesis (localhost)". The window displays a table with the following columns: userid, course, module, action, and info. The table contains 20 rows of data, showing interactions performed by users 2185, 2209, and 2209 on course 29019. The actions include view all, view, add discussion, subscribe, view discussion, view forum, view forums, view, view all, view, view all, upload, view, view all, view all, and view submission. The status bar at the bottom indicates "SELECT Query Time:7.11 sec Record 63 of 178".

userid	course	module	action	info
2185	29019	assignment	view all	
2185	29019	course	view	29019
2185	29019	forum	add discussion	57712
2185	29019	forum	subscribe	14230
2185	29019	forum	view discussion	57712
2185	29019	forum	view forum	13043
2185	29019	forum	view forums	
2185	29019	quiz	view	10207
2185	29019	quiz	view all	
2185	29019	resource	view	97338
2185	29019	resource	view all	
2185	29019	upload	upload	/data/eva/29019/moddata/forum/
2185	29019	user	view	23611
2185	29019	user	view all	
2209	29019	assignment	view all	
2209	29019	assignment	view submission	18195

Fig. 33 Tabla con los datos de interacción de estudiantes y tutor que cursan la materia.

Fuente: Ruiz, F., (2014).

Esta consulta permite determinar la secuencia de navegación de los estudiantes cuando interactúan con la materia de administración de herramientas CASE, es decir, buscar los datos de los estudiantes y tutor que han navegado por esta materia, y las herramientas que han utilizado mientras han interactuado en el EVA., la Fig., presenta una vista general de las acciones realizadas por los estudiantes en las herramientas del EVA, del curso correspondiente dentro del sistema.

CONTROL DE CALIDAD [A].

Los datos del curso se obtienen mediante la siguiente consulta SQL a la base de datos del EVA:

SELECT DISTINCT

```
mdl_course.id, mdl_course.fullname, mdl_enrol_utpl.pdoid, mdl_enrol_utpl.modalidad
FROM mdl_course
INNER JOIN mdl_enrol_utpl ON mdl_enrol_utpl.courseid = mdl_course.id
WHERE mdl_course.id = 51751
```

Los datos de los estudiantes enrolados en este curso y las acciones que realizan dentro del mismo, se obtienen mediante la siguiente consulta SQL a la base de datos:

SELECT

```
mdl_log.userid, mdl_log.course, mdl_log.module, mdl_log.action, mdl_log.info
FROM mdl_course
INNER JOIN mdl_log ON mdl_log.course = mdl_course.id
WHERE mdl_log.course = 51751
GROUP BY mdl_log.userid, mdl_log.module, mdl_log.action ASC
```

userid	course	module	action	info
38	28985	assignment	add	16759
38	28985	assignment	update	16759
38	28985	assignment	update grades	6672
38	28985	assignment	view	16759
38	28985	assignment	view all	
38	28985	assignment	view submission	16292
38	28985	course	add mod	quiz 10634
38	28985	course	delete mod	quiz 10634
38	28985	course	editsection	1
38	28985	course	update mod	forum 13085
38	28985	course	view	28985
38	28985	forum	add	13085
38	28985	forum	update	13085
38	28985	forum	view discussion	57719
38	28985	forum	view forum	13085

Fig. 34 Tabla con los datos de interacción de estudiantes y tutor que cursan la materia.

Fuente: Ruiz, F., (2014).

Esta consulta permite determinar la secuencia de navegación de los estudiantes cuando interactúan con la materia de *Modelamiento de datos*, es decir, buscar los datos de los estudiantes que han navegado por esta materia, y las herramientas que han utilizado mientras han interactuado en el EVA., la Fig., presenta una vista general de las acciones realizadas por los estudiantes en las herramientas del EVA, del curso correspondiente dentro del sistema.

MODELAMIENTO DE DATOS [A]

Para la extracción de los datos de las materias a las cuales se les realizará la minería se los realizó de forma manual mediante consultas en el lenguaje SQL, como se detalla a continuación.

Los datos del curso se obtienen mediante la siguiente consulta SQL a la base de datos:

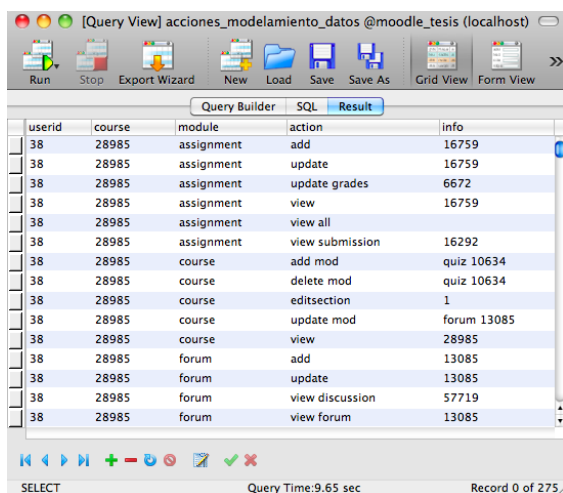
SELECT DISTINCT

```
mdl_course.id, mdl_course.fullname, mdl_enrol_utpl.pdoid, mdl_enrol_utpl.modalidad
FROM mdl_course
INNER JOIN mdl_enrol_utpl ON mdl_enrol_utpl.courseid = mdl_course.id
WHERE mdl_course.id = 51763
```

Los datos de los estudiantes enrolados en este curso y las acciones que realizan dentro del mismo, se obtienen mediante la siguiente consulta SQL a la base de datos:

SELECT

```
mdl_log.userid, mdl_log.course, mdl_log.module, mdl_log.action, mdl_log.info
FROM mdl_course
INNER JOIN mdl_log ON mdl_log.course = mdl_course.id
WHERE mdl_log.course = 51763
GROUP BY mdl_log.userid, mdl_log.module, mdl_log.action ASC
```



The screenshot shows a database query result window titled "[Query View] acciones_modelamiento_datos @moodle_tesis (localhost)". The window contains a table with the following data:

userid	course	module	action	info
38	28985	assignment	add	16759
38	28985	assignment	update	16759
38	28985	assignment	update grades	6672
38	28985	assignment	view	16759
38	28985	assignment	view all	
38	28985	assignment	view submission	16292
38	28985	course	add mod	quiz 10634
38	28985	course	delete mod	quiz 10634
38	28985	course	editsection	1
38	28985	course	update mod	forum 13085
38	28985	course	view	28985
38	28985	forum	add	13085
38	28985	forum	update	13085
38	28985	forum	view discussion	57719
38	28985	forum	view forum	13085

Fig. 34 Tabla con los datos de interacción de estudiantes y tutor que cursan la materia.

Fuente: Ruiz, F., (2014).

Esta consulta permite determinar la secuencia de navegación de los estudiantes cuando interactúan con la materia de *Modelamiento de datos*, es decir, buscar los datos de los estudiantes que han navegado por esta materia, y las herramientas que han utilizado mientras han interactuado en el EVA., la Fig., presenta una vista general de las acciones realizadas por los estudiantes en las herramientas del EVA, del curso correspondiente dentro del sistema.

REDES Y SISTEMAS DISTRIBUIDOS

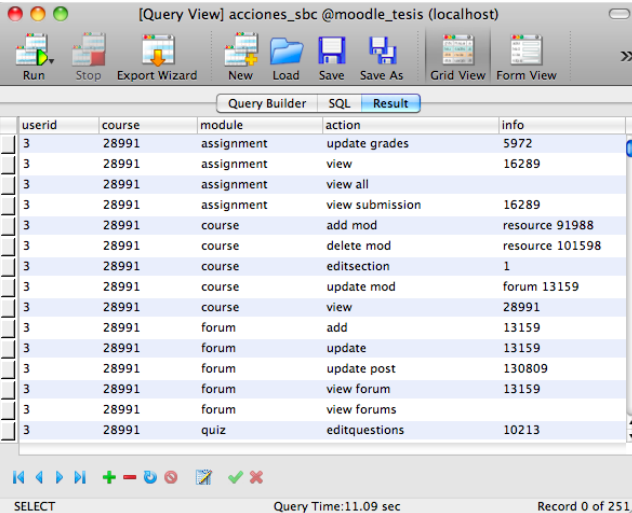
Para la extracción de los datos de las materias a las cuales se les realizará la minería se los realizó de forma manual mediante consultas en el lenguaje SQL, como se detalla a continuación.

Los datos del curso se obtienen mediante la siguiente consulta SQL a la base de datos:

```
SELECT DISTINCT mdl_course.id, mdl_course.fullname, mdl_enrol_utpl.pdoid,  
mdl_enrol_utpl.modalidad  
FROM mdl_course  
INNER JOIN mdl_enrol_utpl ON mdl_enrol_utpl.courseid = mdl_course.id  
WHERE mdl_course.id = 51755
```

Los datos de los estudiantes enrolados en este curso y las acciones que realizan dentro del mismo, se obtienen mediante la siguiente consulta SQL a la base de datos:

```
SELECT  
mdl_log.userid, mdl_log.course, mdl_log.module, mdl_log.action, mdl_log.info  
FROM mdl_course  
INNER JOIN mdl_log ON mdl_log.course = mdl_course.id  
WHERE mdl_log.course = 51755  
GROUP BY mdl_log.userid, mdl_log.module, mdl_log.action  
ASC
```



The screenshot shows a database query result window titled "[Query View] acciones_sbc @moodle_tesis (localhost)". The window displays a table with the following data:

userid	course	module	action	info
3	28991	assignment	update grades	5972
3	28991	assignment	view	16289
3	28991	assignment	view all	
3	28991	assignment	view submission	16289
3	28991	course	add mod	resource 91988
3	28991	course	delete mod	resource 101598
3	28991	course	editsection	1
3	28991	course	update mod	forum 13159
3	28991	course	view	28991
3	28991	forum	add	13159
3	28991	forum	update	13159
3	28991	forum	update post	130809
3	28991	forum	view forum	13159
3	28991	forum	view forums	
3	28991	quiz	editquestions	10213

Fig. 35 Tabla con los datos de interacción de estudiantes y tutor que cursan la materia.

Fuente: Ruiz, F., (2014).

Esta consulta permite determinar la secuencia de navegación de los estudiantes cuando interactúan con la materia de Sistemas basados en el conocimiento, es decir, buscar los datos de los estudiantes que han navegado por esta materia, y las herramientas que han utilizado en el EVA., la Fig., presenta una vista general de las acciones realizadas por los estudiantes en las herramientas del EVA, del curso correspondiente dentro del sistema.

SISTEMAS BASADOS EN EL CONOCIMIENTO

Para la extracción de los datos de las materias a las cuales se les realizará la minería se los realizó de forma manual mediante consultas en el lenguaje SQL, como se detalla a continuación. Los datos del curso se obtienen mediante la siguiente consulta SQL a la base de datos:

```
SELECT DISTINCT mdl_course.id, mdl_course.fullname, mdl_enrol_utpl.pdoid,  
mdl_enrol_utpl.modalidad  
FROM mdl_course  
INNER JOIN mdl_enrol_utpl ON mdl_enrol_utpl.courseid = mdl_course.id  
WHERE mdl_course.id 51750
```

Los datos de los estudiantes enrolados en este curso y las acciones que realizan dentro del mismo, se obtienen mediante la siguiente consulta SQL a la base de datos:

```
SELECT  
mdl_log.userid, mdl_log.course, mdl_log.module, mdl_log.action, mdl_log.info  
FROM mdl_course  
INNER JOIN mdl_log ON mdl_log.course = mdl_course.id  
WHERE mdl_log.course = 51750  
GROUP BY mdl_log.userid, mdl_log.module, mdl_log.action  
ASC
```

userid	course	module	action	info
3	28991	assignment	update grades	5972
3	28991	assignment	view	16289
3	28991	assignment	view all	
3	28991	assignment	view submission	16289
3	28991	course	add mod	resource 91988
3	28991	course	delete mod	resource 101598
3	28991	course	editsection	1
3	28991	course	update mod	forum 13159
3	28991	course	view	28991
3	28991	forum	add	13159
3	28991	forum	update	13159
3	28991	forum	update post	130809
3	28991	forum	view forum	13159
3	28991	forum	view forums	
3	28991	quiz	editquestions	10213

Fig. 35 Tabla con los datos de interacción de estudiantes y tutor que cursan la materia.

Fuente: Ruiz, F., (2014).

Esta consulta permite determinar la secuencia de navegación de los estudiantes cuando interactúan con la materia de Sistemas basados en el conocimiento, es decir, buscar los datos de los estudiantes que han navegado por esta materia, y las herramientas que han utilizado en el EVA., la Fig., presenta una vista general de las acciones realizadas por los estudiantes en las herramientas del EVA, del curso correspondiente dentro del sistema.

SISTEMAS DE CONTROL [A]

Para la extracción de los datos de las materias a las cuales se les realizará la minería se los realizó de forma manual mediante consultas en el lenguaje SQL, como se detalla a continuación. Los datos del curso se obtienen mediante la siguiente consulta SQL a la base de datos:

SELECT DISTINCT

mdl_course.id, mdl_course.fullname, mdl_enrol_utpl.pdoid, mdl_enrol_utpl.modalidad

FROM mdl_course

INNER JOIN mdl_enrol_utpl ON mdl_enrol_utpl.courseid = mdl_course.id

WHERE mdl_course.id = 51746

Los datos de los estudiantes enrolados en este curso y las acciones que realizan dentro del mismo, se obtienen mediante la siguiente consulta SQL a la base de datos:

SELECT mdl_log.userid, mdl_log.course, mdl_log.module, mdl_log.action, mdl_log.info

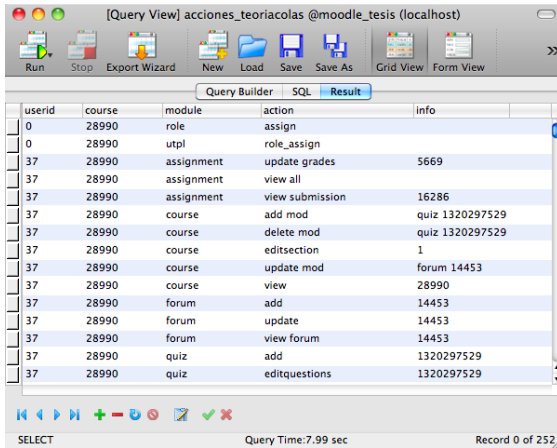
FROM mdl_course

INNER JOIN mdl_log ON mdl_log.course = mdl_course.id

WHERE mdl_log.course = 51746

GROUP BY mdl_log.userid, mdl_log.module, mdl_log.action

ASC



The screenshot shows a database query result window titled "[Query View] acciones_teoriacolas @moodle_tesis (localhost)". The window contains a table with the following data:

userid	course	module	action	info
0	28990	role	assign	
0	28990	utpl	role_assign	
37	28990	assignment	update grades	5669
37	28990	assignment	view all	
37	28990	assignment	view submission	16286
37	28990	course	add mod	quiz 1320297529
37	28990	course	delete mod	quiz 1320297529
37	28990	course	editsection	1
37	28990	course	update mod	forum 14453
37	28990	course	view	28990
37	28990	forum	add	14453
37	28990	forum	update	14453
37	28990	forum	view forum	14453
37	28990	quiz	add	1320297529
37	28990	quiz	editquestions	1320297529

Fig. 36. Tabla con los datos de interacción de estudiantes y tutor que cursan la materia.

Fuente: Ruiz, F., (2014).

Esta consulta permite determinar la secuencia de navegación de los estudiantes que interactúan con la materia de *Teoría de colas*, es decir, buscar los datos de los estudiantes que han interactuado en las herramientas del EVA., ver la **Fig. 36**, que presenta una vista general de las acciones realizadas por los estudiantes en el EVA, del curso correspondiente dentro del sistema.

ANEXO C

TRAZAS DE NAVEGACIÓN DE LOS ESTUDIANTES OBTENIDOS DEL EVA DE LA UTPL, MEDIANTE CONSULTAS SQL.

Para determinar las trazas de navegación de los estudiantes que interactúan con el EVA, se realizaron varias consultas SQL en el EVA, los cuales permitan determinar las rutas de navegación y las acciones que los usuarios realizan dentro de la plataforma de la UTPL, los datos que se consideraron como información fundamental para la presente investigación son los archivos Log; ya que estos datos son los sucesos ocurridos en el sistema por cada uno de los usuarios que interactúan con los cursos. Estas trazas de navegación son las que permiten medir el nivel de interacción del usuario en el sistema, la facilidad de manejo y colaboración con las herramientas disponibles en el sistema, y la comunicación con los demás estudiantes y tutores.

Las consultas que permiten determinar las rutas de navegación de los estudiantes en el sistema, se realizaron principalmente con la entidad de los Logs (*mdl_log*), el mismo que esta compuesto por los campos detallados en la **Fig. 19** y otras entidades más que se relacionan directamente con la entidad de los Log, con la finalidad de conocer las rutas de navegación de los estudiantes en el EVA y poder con estos datos realizar el experimento, antes de realizar la minería de los mismos.

Fuente: Ruiz, F., (2014).

Atributo	Tipo	Tamaño	Descripción
<i>Id</i>	bigint	10	Identificación del evento o actividad almacenada en la entidad
<i>Userid</i>	Bigint	10	Identificación del usuario que realizó la actividad en el sistema
<i>Course</i>	Bigint	10	Curso en el cual el estudiante realizó las actividades dentro del EVA
<i>Module</i>	Varchar	20	Herramientas/Actividades con las que el estudiante interactúa en el EVA
<i>Action</i>	Varchar	40	Acciones que el estudiante realiza cuando interactúa con las herramientas del EVA.
<i>Info</i>	Varchar	255	Recursos a los cuales ingresa el estudiante cuando navega por el EVA.

Tabla 38. Atributos de la entidad mdl_log.

Estos atributos representan las acciones realizadas por los estudiantes cuando interactúan dentro del sistema, la cual permitirá descubrir los patrones de navegación que tienen los estudiantes cuando interactúan con el sistema.

B. TRAZAS DE NAVEGACIÓN DE LOS ESTUDIANTES EN LAS HERRAMIENTAS DEL EVA

La razón por la que se partió con el análisis de los archivos Log, es debido a estudios ya realizados como se lo detallo en trabajos relacionados, los cuales permiten encontrar patrones de navegación en los usuarios que acceden al sistema. Para la obtención de estos datos se aplicó la técnica mediante consultas de lenguaje SQL que permiten la agrupación y la suma total de datos

relacionados a la interacción de las tablas, la cual permitirá diferir las trazas de navegación de los estudiantes dentro de los cursos.

Los datos de navegación de los estudiantes que interactúan con las herramientas disponibles en el sistema, mediante consultas en el lenguaje SQL se presentan según la **Tabla 38**, la cual se detalla en el “Análisis del Entorno Virtual de Moodle, (EVA) de la UTPL”; y serán analizadas cada una de las acciones presentadas en dicha tabla.

C.B.1 Número de acceso a los cursos de Informática de la UTPL

El Número de accesos al curso por los estudiantes, permite identificar el número de veces que el estudiante a accedido a un curso, y a realizado algún tipo de acción (vistas a las páginas, ingreso a recursos, vistas a foros, subió tareas, etc.) en las herramientas de las cuales disponen los cursos, como lo detalla la consulta SQL en la **Fig. C.B.1**.

SELECT

COUNT (mdl_log.userid) AS USUARIOS,

COUNT (mdl_log.course) AS CURSOS,

COUNT (mdl_log.module) AS HERRAMIENTAS,

COUNT (mdl_log.action) AS ACCIONES,

COUNT (mdl_log.info) AS RECURSOS

FROM mdl_log

WHERE action = 'view' AND (module = 'course' || module = 'resource' || module = 'blog' || module = 'quiz' || module = 'assignment' || module = 'twitter' || module = 'chat' || module = 'forum' || module = 'message') AND course = 51749 || course = 51751 || course = 51763 || course = 51755 || course = 51750 || course = 51746

USUARIOS	CURSOS	HERRAMIENTAS	ACCIONES	RECURSOS
8993	8993	8993	8993	8993

Fig. C.B.1.1 Número de accesos de los estudiantes al curso en el EVA

Fuente: Ruiz, F., (2014).

C.B.3 Número de aportes al curso

El número de aportes dentro de los cursos permite conocer cuáles son las herramientas con las que los estudiantes interactúan con mayor frecuencia, es decir, el número de veces que un estudiante ha realizado una determinada acción sobre las herramientas del curso en el que está navegando, este parámetro permite saber según las acciones realizadas a través de las herramientas dentro del curso, las trazas de navegación más frecuentes del estudiante, de tal manera que estos datos permitan determinar indicadores de navegación de los estudiantes, cuando adquieren los conocimientos durante su interacción con el EVA.

El número total de aportes realizados dentro de las herramientas del curso de *Administración de Herramientas CASE*, se presenta en la siguiente consulta SQL.

SELECT

COUNT (mdl_log.userid) **AS** COUNT_USUARIO,

COUNT (mdl_log.ip) **AS** COUNT_IPs,

COUNT (mdl_log.course) **AS** COUNT_CURSO,

COUNT (mdl_log.module) **AS** COUNT_HERRAMIENTAS,

COUNT (mdl_log.action) **AS** COUNT_ACCIONES,

COUNT (mdl_log.url) **AS** COUNT_URL,

COUNT (mdl_log.info) **AS** COUNT_RECURSOS

FROM mdl_log

WHERE ((module = 'assignment' AND action = 'wiev') || (module = 'assignment' AND action = 'wiev all') || (module = 'assignment' AND action = 'view submission') || (module = 'assignment' AND action = 'upload') || (module = 'assignment' AND action = 'add') || (module = 'assignment' AND action = 'update') || (module = 'assignment' AND action = 'update grades') || (module = 'blog' AND action = 'wiev') || (module = 'course' AND action = 'wiev') || (module = 'course' AND action = 'user report') || (module = 'course' AND action = 'add mod') || (module = 'course' AND action = 'editsection') || (module = 'course' AND action = 'update mod') || (module = 'course' AND action = 'delete mod') || (module = 'forum' AND action = 'add discussion') || (module = 'forum' AND action = 'view discussion') || (module = 'forum' AND action = 'subscribe') || (module = 'forum' AND action = 'view forum') || (module = 'forum' AND action = 'view forums') || (module = 'forum' AND action = 'update post') || (module = 'forum' AND action = 'subscribe') || (module = 'forum' AND action = 'search') || (module = 'forum' AND action = 'add') || (module = 'forum' AND action = 'add post') || (module = 'forum' AND action = 'subscribe') || (module = 'forum' AND action = 'unsubscribe') || (module = 'quiz' AND action = 'wiev') || (module = 'quiz' AND action = 'wiev all') || (module = 'quiz' AND action = 'attempt') || (module = 'quiz' AND action = 'close attempt') || (module = 'quiz' AND action = 'continue attemp') || (module = 'quiz' AND action = 'editquestions') || (module = 'quiz' AND

action = 'preview') || (module = 'quiz' AND action = 'report') || (module = 'quiz' AND action = 'report_quiz') || (module = 'quiz' AND action = 'review') || (module = 'quiz' AND action = 'add') || (module = 'quiz' AND action = 'update') || (module = 'quiz' AND action = 'wiev all') || (module = 'resource' AND action = 'wiev') || (module = 'resource' AND action = 'wiev all') || (module = 'resource' AND action = 'add') || (module = 'resource' AND action = 'update')) **AND** course = 51749

COUNT_USUARIO	COUNT_IPs	COUNT_CURSO	COUNT_HERRAMIENTAS	COUNT_ACCIONES	COUNT_URL	COUNT_RECURSOS
984	984	984	984	984	984	984

Fig. C.B.2.2 Número de aportes de los estudiantes al curso en el EVA.

Fuente: Ruiz, F., (2014).

La **Fig. C.B.3.1** presenta los datos de la consulta realizada, la misma que permite visualizar las acciones de navegación de los estudiantes que interactúan con las herramientas del curso **Administración de Herramientas CASE**, y un detalle del total de aportes que han realizado al curso mediante su navegación por el mismo.

C.B.4. Número de respuestas en las actividades del curso

Las respuestas realizadas por los estudiantes dentro de las herramientas presentadas por cada uno de los cursos que el estudiante esta expidiendo, permiten encontrar el nivel de interacción más frecuente de los estudiantes con las diferentes herramientas o recursos de los cuales dispone para la comunicación he interacción con lo compañeros y tutor del curso, el cual representa el número de veces que el estudiante interactúo con estas herramienta.

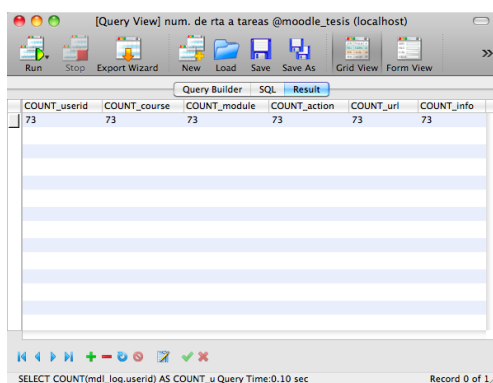
C.B.4.1. Número de respuestas en las Tareas

Estos datos obtenido mediante consultas SQL, permite encontrar las acciones más frecuentes de los estudiantes dentro de la herramienta de **tareas**, la cual ayudara a medir el nivel de interacción del estudiante con el tutor y compañeros dentro del curso de **Administración de Herramientas CASE** y las acciones realizadas, tales como: si ha revisando, respondido, cargado o subido al sistema una tarea, enviando una tarea, etc., cualquiera de las acciones que el estudiante haya realizado dentro de la herramienta **Assignment** 'Tarea', dentro del curso.

El Número total de respuestas realizadas dentro de la herramienta de tareas '**Assignment**', se presenta en la siguiente consulta SQL.

SELECT

```
COUNT (mdl_log.userid) AS COUNT_userid,  
COUNT (mdl_log.course) AS COUNT_course,  
COUNT( mdl_log.module) AS COUNT_module,  
COUNT (mdl_log.action) AS COUNT_action,  
COUNT (mdl_log.url) AS COUNT_url,  
COUNT (mdl_log.info) AS COUNT_info  
FROM mdl_log  
WHERE course = 51749 AND (module = 'assignment' AND (action = 'view' || action = 'view all' ||  
action = 'view submission' || action = 'upload' || action = 'update grades' || action = 'add'))
```



The screenshot shows a window titled "[Query View] num. de rta a tareas @moodle_tesis (localhost)". The window contains a table with the following data:

COUNT_userid	COUNT_course	COUNT_module	COUNT_action	COUNT_url	COUNT_info
73	73	73	73	73	73

The status bar at the bottom of the window displays the SQL query: "SELECT COUNT(mdl_log.userid) AS COUNT_u... Query Time: 0.10 sec" and "Record 0 of 1".

Fig. C.B.4.1.2 Número total de respuestas de los estudiantes en tareas dentro del curso en el EVA

Fuente: Ruiz, F., (2014).

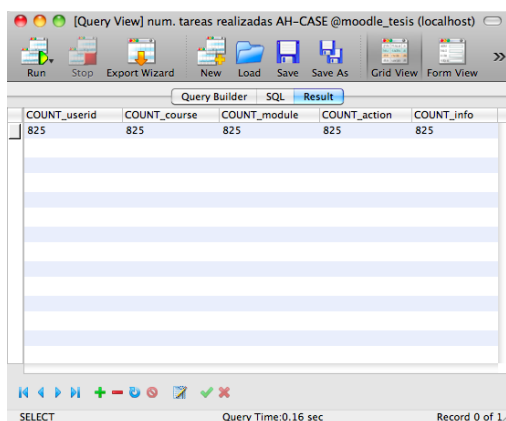
La **Fig. C.B.4.1.1** presenta los datos de la consulta realizada, la misma que permite visualizar las acciones de navegación de los estudiantes que interactúan con la herramienta de tareas '**Assignment**' del curso de **Administración de Herramientas CASE**, y un detalle del total de acciones realizadas por el estudiante mediante su navegación por el curso.

C.B.5.1. Número de Tareas realizadas

Este dato obtenido mediante consultas en lenguaje SQL, permite encontrar el Número de tareas enviadas de los estudiantes dentro de la herramienta de **Assignment** la cual ayudara a medir el nivel de interacción del estudiante con el tutor y compañeros dentro del curso de **Administración de Herramientas CASE** y las acciones realizadas en el mismo, tales como: si ha revisando, respondido, ingresado o cargado un trabajo, etc., según la cantidad de tareas enviadas por el tutor, ya sea, cualquiera de las acciones que el estudiante haya realizado dentro de la herramienta **Quiz** 'exámenes', dentro del curso.

El Número total de respuestas realizadas dentro de la herramienta de tareas '**Assignment**', se presenta en la siguiente consulta SQL.

```
SELECT  
COUNT (mdl_log.userid) AS COUNT_userid,  
COUNT (mdl_log.course) AS COUNT_course,  
COUNT (mdl_log.module) AS COUNT_module,  
COUNT (mdl_log.action) AS COUNT_action,  
COUNT (mdl_log.info) AS COUNT_info  
FROM mdl_role_assignments  
INNER JOIN mdl_log ON mdl_log.userid = mdl_role_assignments.userid  
WHERE roleid = 5 AND (action like 'upload' || action like 'add') AND course = 51749
```



The screenshot shows a window titled "[Query View] num. tareas realizadas AH-CASE @moodle_tesis (localhost)". The window contains a table with the following data:

COUNT_userid	COUNT_course	COUNT_module	COUNT_action	COUNT_info
825	825	825	825	825

At the bottom of the window, it displays "SELECT" and "Query Time: 0.16 sec" and "Record 0 of 1".

Fig. C.B.5.1.2 Número total de tareas realizadas de los estudiantes dentro del curso en el EVA.

Fuente: Ruiz, F., (2014).

La **Fig. C.B.5.1.1** presenta los datos de la consulta realizada, la misma que permite visualizar un detalle del total de las tareas realizadas durante su navegación en el EVA, y la interacción de con el tutor mediante la herramienta '**Assignment**' del curso de **Administración de Herramientas CASE**.

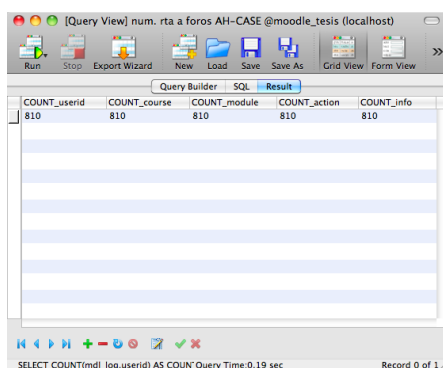
C.B.6.1. Número de respuestas en los Foros

Este dato obtenido mediante consultas en lenguaje SQL, permite encontrar las acciones más frecuentes de los estudiantes dentro de la herramienta de **foros**, la cual ayudara a medir el nivel de interacción del estudiante con el tutor y compañeros dentro del curso de **Administración de Herramientas CASE** y las acciones realizadas en el mismo, tales como: si ha revisando, respondido, agregado, o asignado un foro, etc., cualquiera de las acciones que el estudiante haya realizado dentro de la herramienta **Forum** 'foros', dentro del curso.

El Número total de respuestas realizadas en la herramienta foros '*Forum*', se presenta en la siguiente consulta SQL.

SELECT

```
COUNT (mdl_log.userid) AS COUNT_userid,  
COUNT (mdl_log.course) AS COUNT_course,  
COUNT (mdl_log.module) AS COUNT_module,  
COUNT (mdl_log.action) AS COUNT_action,  
COUNT (mdl_log.info) AS COUNT_info  
FROM mdl_log  
WHERE course = 51749 AND (module = 'forum' AND (action = 'view forum' || action = 'view  
forums' || action = 'view discussion' || action = 'subscribe' || action = 'add discussion' || action =  
'add' || action = 'add post' || action = 'update' || action = 'update post' || action = 'user report'))
```



The screenshot shows a window titled "[Query View] num. rta a foros AH-CASE @moodle_tesis (localhost)". The window contains a table with the following data:

COUNT_userid	COUNT_course	COUNT_module	COUNT_action	COUNT_info
810	810	810	810	810

At the bottom of the window, the SQL query is visible: "SELECT COUNT(mdl_log.userid) AS COUNT_course". The status bar indicates "Record 0 of 1".

Fig. C.B.6.1.2 Número total de respuestas de los estudiantes en los foros dentro del curso en el EVA
Fuente: Ruiz, F., (2014).

La **Fig. C.B.6.1.1** presenta los datos de la consulta realizada, la misma que permite visualizar las acciones de navegación de los estudiantes que interactúan con la herramienta de los foros '*Forum*' del curso de **Administración de Herramientas CASE**, y un detalle del total de acciones realizadas por el estudiante mediante su navegación por el curso.

C.B.7.1. Número de respuestas en los Recursos

Este dato obtenido mediante consultas en lenguaje SQL, permite encontrar las acciones más frecuentes de los estudiantes dentro de la herramienta de **recursos** la cual ayudara a medir el nivel de interacción del estudiante con el tutor y compañeros dentro del curso de *Administración de Herramientas CASE* y las acciones realizadas en el mismo, tales como: si ha revisando, abierto o ingresado a un recurso, etc., cualquiera de las acciones que el estudiante haya realizado dentro de la herramienta **Resource** 'recursos', dentro del curso.

SELECT

mdl_log.userid, mdl_log.course, mdl_log.module, mdl_log.action, mdl_log.info

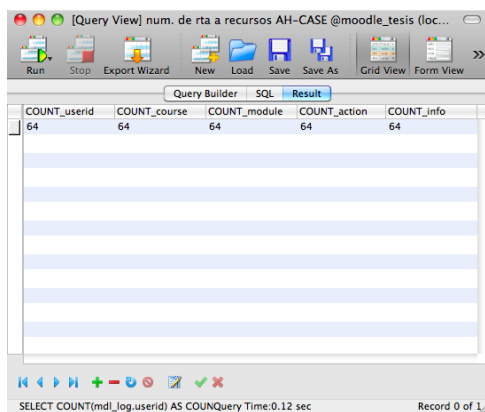
FROM

mdl_log

WHERE course = 51749

AND (module = 'resource' AND (action = 'view' || action = 'view all' || action = 'add' || action = 'update'))

El Número total de respuestas realizadas en la herramienta de recursos '**Resource**', se presenta en la siguiente consulta SQL.



The screenshot shows a 'Query View' window with a table of results. The table has five columns: COUNT_userid, COUNT_course, COUNT_module, COUNT_action, and COUNT_info. The first row contains the value '64' for each of these columns. The window title is '[Query View] num. de rta a recursos AH-CASE @moodle_tesis (loc...'. The SQL query at the bottom is 'SELECT COUNT(mdl_log.userid) AS COUNQuery Time:0.12 sec Record 0 of 1'.

COUNT_userid	COUNT_course	COUNT_module	COUNT_action	COUNT_info
64	64	64	64	64

Fig. C.B.7.1.2 Número total de respuestas de los estudiantes en los recursos del curso en el EVA.

Fuente: Ruiz, F., (2014).

La **Fig. C.B.7.1.1** presenta los datos de la consulta realizada, la misma que permite visualizar las acciones de navegación de los estudiantes que interactúan con la herramienta de los recursos '**Resource**' del curso de **Administración de Herramientas CASE**, y un detalle del total de acciones realizadas por el estudiante mediante su navegación por el curso.

C.B.8.1. Número de respuestas en los Exámenes

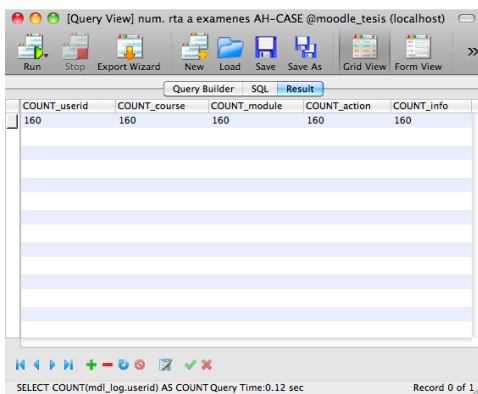
Este dato obtenido mediante consultas en lenguaje SQL, permite encontrar las acciones más frecuentes de los estudiantes dentro de la herramienta de **exámenes** la cual ayudara a medir el nivel de interacción del estudiante con el tutor y compañeros dentro del curso de **Administración de Herramientas CASE** y las acciones realizadas en el mismo, tales como: si ha revisando, respondido o ingresado a un examen, etc., cualquiera de las acciones que el estudiante haya realizado dentro de la herramienta **Quiz** 'exámenes', dentro del curso.

El Número total de respuestas realizadas en el recurso '**Quiz**', se presenta en la siguiente consulta SQL.

SELECT

```
COUNT (mdl_log.userid) AS COUNT_userid,  
COUNT (mdl_log.course) AS COUNT_course,  
COUNT (mdl_log.module) AS COUNT_module,  
COUNT (mdl_log.action) AS COUNT_action,  
COUNT (mdl_log.info) AS COUNT_info  
FROM mdl_log
```

```
WHERE course = 29019 AND (module = 'quiz' AND (action = 'view' || action = 'view all' || action =  
'attempt' || action = 'continue attempt' || action = 'close attempt' || action = 'preview' || action =  
'report' || action = 'report_quiz' || action = 'review' || action = 'add' || action = 'editquestions' || action  
= 'update'))
```



The screenshot shows a database query result window titled "[Query View] num. rta a exámenes AH-CASE @moodle_tesis (localhost)". The window has a menu bar with options: Run, Stop, Export Wizard, New, Load, Save, Save As, Grid View, and Form View. Below the menu bar, there are tabs for "Query Builder", "SQL", and "Result". The "Result" tab is active, displaying a table with the following data:

COUNT_userid	COUNT_course	COUNT_module	COUNT_action	COUNT_info
160	160	160	160	160

At the bottom of the window, there is a status bar with navigation icons and the text: "SELECT COUNT(mdl_log.userid) AS COUNT Query Time:0.12 sec Record 0 of 1".

Fig. C.B.8.1.2 Número total de respuestas de los estudiantes en los exámenes del curso en el EVA.
Fuente: Ruiz, F., (2014).

La **Fig. C.B.8.1.1** presenta los datos de la consulta realizada, la misma que permite visualizar las acciones de navegación de los estudiantes que interactúan con la herramienta de los recursos 'Quiz' del curso de **Administración de Herramientas CASE**, y un detalle del total de acciones realizadas por el estudiante mediante su navegación por el curso.

ANEXO D

HERRAMIENTAS POR LAS QUE NAVEGA EL ESTUDIANTE Y LAS ACCIONES QUE REALIZA

Las acciones almacenadas en los archivos Log, de los estudiantes de noveno ciclo de la carrera de Informática que navegan por el EVA e interactúan con las herramientas del sistema, se muestra en la **Tabla 40**, que contiene los cursos a los que el estudiante y tutor acceden para su comunicación y colaboración en el EVA.

SELECT

mdl_log.userid, mdl_log.course, mdl_log.module, mdl_log.action, mdl_log.url,

FROM mdl_log

WHERE (module = 'course' AND (action = 'view')) **AND** course = 51749

Se utilizó el mismo código SQL para todas las materias, en algunos casos sólo varía el id del curso y el modulo y la acción, según la navegación del estudiante en las herramientas del EVA.

Fuente: Ruiz, F., (2014).

MATERIAS	SERVICIOS	ACCIONES REALIZADAS	PÁGINA VISITADA POR LOS ESTUDIANTES	NÚMERO INTERACCIÓN
Administración de Herramientas CASE	Course	view	view.php?id=29019	479
	Course	user report	user.php?id=29019&user=4198&mode=outline	5
	Assignment	view	view.php?id=139728	21
	Assignment	view all	index.php?id=29019	27
	Assignment	upload	view.php?a=16283	9
	Quiz	view	view.php?id=139676	16
	Quiz	view	view.php?id=151676	9
	Quiz	view all	index.php?id=29019	10
	Quiz	attempt	review.php?attempt=65168	1
	Quiz	attempt	review.php?attempt=65219	1
	Quiz	attempt	review.php?attempt=66343	1
	Quiz	attempt	review.php?attempt=67330	1
	Quiz	attempt	review.php?attempt=68544	1
	Quiz	attempt	review.php?attempt=71893	1
	Quiz	attempt	review.php?attempt=83053	1
	Quiz	attempt	review.php?attempt=85538	1
	Quiz	attempt	review.php?attempt=87500	1
	Quiz	continue attemp	review.php?attempt=83053	1
	Quiz	close attempt	review.php?attempt=62805	1
	Quiz	close attempt	review.php?attempt=65168	1
Quiz	close attempt	review.php?attempt=65219	1	
Quiz	close attempt	review.php?attempt=66343	1	
Quiz	close attempt	review.php?attempt=67330	1	

Quiz	close attempt	review.php?attempt=68544	1
Quiz	close attempt	review.php?attempt=71893	1
Quiz	close attempt	review.php?attempt=76930	1
Quiz	close attempt	review.php?attempt=76932	1
Quiz	close attempt	review.php?attempt=83053	1
Quiz	close attempt	review.php?attempt=85538	1
Quiz	close attempt	review.php?attempt=87500	1
Forum	view forum	view.php?id=141013	100
Forum	view forum	view.php?id=141017	80
Forum	view forum	view.php?id=145684	38
Forum	view forum	view.php?id=155488	76
Forum	view forum	view.php?id=155392	74
Forum	view forums	index.php?id=29019	93
Forum	view discussion	discuss.php?d=57712	8
Forum	view discussion	discuss.php?d=57795	26
Forum	view discussion	discuss.php?d=57796	15
Forum	view discussion	discuss.php?d=58164	6
Forum	view discussion	discuss.php?d=58169	13
Forum	view discussion	discuss.php?d=57795&parent=131230	2
Forum	view discussion	discuss.php?d=58538	8
Forum	view discussion	discuss.php?d=58541	6
Forum	view discussion	discuss.php?d=58548	11
Forum	view discussion	discuss.php?d=58649	7
Forum	view discussion	discuss.php?d=58817	4
Forum	view discussion	discuss.php?d=58949	6
Forum	view discussion	discuss.php?d=58857	4
Forum	view discussion	discuss.php?d=58862	3
Forum	view discussion	discuss.php?d=58982	5
Forum	view discussion	discuss.php?d=58979	8
Forum	view discussion	discuss.php?d=58956	7
Forum	view discussion	discuss.php?d=59208	4
Forum	view discussion	discuss.php?d=59308	10
Forum	view discussion	discuss.php?d=59309	2
Forum	view discussion	discuss.php?d=59652	1
Forum	view discussion	discuss.php?d=59932	4
Forum	view discussion	discuss.php?d=60072	7
Forum	view discussion	discuss.php?d=60332	1
Forum	view discussion	discuss.php?d=60736	1
Forum	view discussion	discuss.php?d=62056	19
Forum	view discussion	discuss.php?d=63354	10
Forum	view discussion	discuss.php?d=63550	12
Forum	view discussion	discuss.php?d=63366	10
Forum	view discussion	discuss.php?d=63736	11
Forum	view discussion	discuss.php?d=64172	7
Forum	view discussion	discuss.php?d=64103	6
Forum	view discussion	discuss.php?d=64379	2
Forum	view discussion	discuss.php?d=64165	5
Forum	view discussion	discuss.php?d=64285	5
Forum	view discussion	discuss.php?d=64377	3

Forum	view discussion	discuss.php?d=64383	3
Forum	view discussion	discuss.php?d=64376	1
Forum	view discussion	discuss.php?d=64280	4
Forum	view discussion	discuss.php?d=64425	6
Forum	view discussion	discuss.php?d=64426	2
Forum	view discussion	discuss.php?d=65080	1
Forum	view discussion	discuss.php?d=65083	1
Forum	subscribe	view.php?f=13043	1
Forum	subscribe	view.php?f=13044	1
Forum	subscribe	view.php?f=14230	1
Forum	view subscribers	-----	0
Forum	add discussion	discuss.php?d=57712	1
Forum	add discussion	discuss.php?d=57795	1
Forum	add discussion	discuss.php?d=57796	1
Forum	add discussion	discuss.php?d=58164	1
Forum	add discussion	discuss.php?d=58169	1
Forum	add discussion	discuss.php?d=58538	1
Forum	add discussion	discuss.php?d=58541	1
Forum	add discussion	discuss.php?d=58548	1
Forum	add discussion	discuss.php?d=58649	1
Forum	add discussion	discuss.php?d=58817	1
Forum	add discussion	discuss.php?d=58857	1
Forum	add discussion	discuss.php?d=58862	1
Forum	add discussion	discuss.php?d=58949	1
Forum	add discussion	discuss.php?d=58956	1
Forum	add discussion	discuss.php?d=58979	1
Forum	add discussion	discuss.php?d=58982	1
Forum	add discussion	discuss.php?d=59208	1
Forum	add discussion	discuss.php?d=59308	1
Forum	add discussion	discuss.php?d=59309	1
Forum	add discussion	discuss.php?d=59652	1
Forum	add discussion	discuss.php?d=59672	1
Forum	add discussion	discuss.php?d=59932	1
Forum	add discussion	discuss.php?d=60072	1
Forum	add discussion	discuss.php?d=60332	1
Forum	add discussion	discuss.php?d=60736	1
Forum	add discussion	discuss.php?d=62056	1
Forum	add discussion	discuss.php?d=63354	1
Forum	add discussion	discuss.php?d=63366	1
Forum	add discussion	discuss.php?d=63550	1
Forum	add discussion	discuss.php?d=63736	1
Forum	add discussion	discuss.php?d=64103	1
Forum	add discussion	discuss.php?d=64165	1
Forum	add discussion	discuss.php?d=64172	1
Forum	add discussion	discuss.php?d=64280	1
Forum	add discussion	discuss.php?d=64285	1
Forum	add discussion	discuss.php?d=64376	1
Forum	add discussion	discuss.php?d=64377	1
Forum	add discussion	discuss.php?d=64379	1

	Forum	add discussion	discuss.php?d=64383	1
	Forum	add discussion	discuss.php?d=64425	1
	Forum	add discussion	discuss.php?d=64426	1
	Forum	add discussion	discuss.php?d=65080	1
	Forum	add discussion	discuss.php?d=65083	1
	Forum	add discussion	discuss.php?d=65213	1
	Forum	add post	discuss.php?d=57796&parent=131219	1
	Forum	add post	discuss.php?d=57795&parent=131230	1
	Forum	delete discussion	-----	0
	Forum	delete post	-----	0
	Forum	user report	user.php?id=29019&user=34752&mode=posts	2
	Forum	unsubscribe	-----	0
	Forum	update post	discuss.php?d=57795#p131230&parent=131230	1
	Forum	update post	discuss.php?d=58169#p131275&parent=131275	1
	Forum	update post	discuss.php?d=58538#p132012&parent=132012	1
	Forum	update post	discuss.php?d=58649#p132224&parent=132224	1
	Forum	update post	discuss.php?d=58817#p132529&parent=132529	1
	Forum	update post	discuss.php?d=58956#p132790&parent=132790	1
	Forum	update post	discuss.php?d=58979#p132825&parent=132825	1
	Forum	update post	discuss.php?d=58982#p132830&parent=132830	1
	Forum	update post	discuss.php?d=59932#p135195&parent=135195	2
	Forum	update post	discuss.php?d=60072#p135477&parent=135477	1
	Forum	update post	discuss.php?d=63354#p141625&parent=141625	1
	Forum	update post	discuss.php?d=63736#p142158&parent=142158	1
	Forum	update post	discuss.php?d=64103#p142653&parent=142653	1
	Forum	update post	discuss.php?d=64285#p142981&parent=142981	1
	Forum	search	-----	0
	Blog	view	index.php?filtertype=user&filterselect=34752&postid=0&tagid=0&tag=	1
	Resource	view	view.php?id=145686	15
	Resource	view	view.php?id=145692	12
	Resource	view	view.php?id=147369	15
	Resource	view	view.php?id=160205	12
	Resource	view all	index.php?id=29019	6
	User	view	view.php?id=4218&course=29019	4
	User	view	view.php?id=23611&course=29019	1
	User	view	view.php?id=2185&course=29019	1
	User	view	view.php?id=34752&course=29019	6
	User	view	view.php?id=4242&course=29019	4
	User	view	view.php?id=4198&course=29019	1
	User	view	view.php?id=4254&course=29019	1
	User	view all	index.php?id=29019	27
	User	update	-----	0
	Utpl	role_assign	admin/roles/assign.php?contextid=450123&roleid=3	1
	Utpl	role_assign	admin/roles/assign.php?contextid=450123&roleid=5	13
	Total			1522
Modelamiento de Modelamiento de Datos	Course	view	view.php?id=28985	947
	Course	user report	-----	0
	Assignment	view	view.php?id=139737	8

Assignment	view	view.php?id=141637	4
Assignment	view	view.php?id=151737	27
Assignment	view all	index.php?id=28985	26
Assignment	upload	view.php?a=16292	2
Assignment	upload	view.php?a=16759	1
Assignment	upload	view.php?a=18204	10
Quiz	view	view.php?id=139685	7
Quiz	view	view.php?id=151685	25
Quiz	view all	index.php?id=28985	25
Quiz	attempt	review.php?attempt=70034	1
Quiz	attempt	review.php?attempt=72310	1
Quiz	attempt	review.php?attempt=78476	1
Quiz	attempt	review.php?attempt=79802	1
Quiz	attempt	review.php?attempt=81724	1
Quiz	attempt	review.php?attempt=82018	1
Quiz	attempt	review.php?attempt=83440	1
Quiz	attempt	review.php?attempt=83799	1
Quiz	attempt	review.php?attempt=83908	1
Quiz	attempt	review.php?attempt=84101	1
Quiz	attempt	review.php?attempt=84411	1
Quiz	attempt	review.php?attempt=86879	1
Quiz	continue attemp	review.php?attempt=72310	1
Quiz	continue attemp	review.php?attempt=78476	1
Quiz	continue attemp	review.php?attempt=82018	1
Quiz	close attempt	review.php?attempt=70034	1
Quiz	close attempt	review.php?attempt=72310	1
Quiz	close attempt	review.php?attempt=78476	1
Quiz	close attempt	review.php?attempt=79802	1
Quiz	close attempt	review.php?attempt=81724	1
Quiz	close attempt	review.php?attempt=82018	1
Quiz	close attempt	review.php?attempt=83440	1
Quiz	close attempt	review.php?attempt=83799	1
Quiz	close attempt	review.php?attempt=83908	1
Quiz	close attempt	review.php?attempt=84101	1
Quiz	close attempt	review.php?attempt=84411	1
Quiz	close attempt	review.php?attempt=86879	1
Forum	view forum	view.php?id=141631	175
Forum	view forum	view.php?id=141636	58
Forum	view forum	view.php?id=160115	70
Forum	view forums	index.php?id=28985	129
Forum	view discussion	discuss.php?d=57719	21
Forum	view discussion	discuss.php?d=57736	13
Forum	view discussion	discuss.php?d=57830	1
Forum	view discussion	discuss.php?d=57836	9
Forum	view discussion	discuss.php?d=57846	10

Forum	view discussion	discuss.php?d=57786	4
Forum	view discussion	discuss.php?d=57928	17
Forum	view discussion	discuss.php?d=57978	10
Forum	view discussion	discuss.php?d=58028	6
Forum	view discussion	discuss.php?d=58035	11
Forum	view discussion	discuss.php?d=58144	3
Forum	view discussion	discuss.php?d=58308	13
Forum	view discussion	discuss.php?d=58451	3
Forum	view discussion	discuss.php?d=58720	1
Forum	view discussion	discuss.php?d=58723	5
Forum	view discussion	discuss.php?d=58847	2
Forum	view discussion	discuss.php?d=58848	2
Forum	view discussion	discuss.php?d=58853	1
Forum	view discussion	discuss.php?d=58882	5
Forum	view discussion	discuss.php?d=58886	2
Forum	view discussion	discuss.php?d=58895	1
Forum	view discussion	discuss.php?d=59108	4
Forum	view discussion	discuss.php?d=58308&parent=132530	2
Forum	view discussion	discuss.php?d=60080	2
Forum	view discussion	discuss.php?d=60282	15
Forum	view discussion	discuss.php?d=60609	17
Forum	view discussion	discuss.php?d=60078	1
Forum	view discussion	discuss.php?d=64900	36
Forum	view discussion	discuss.php?d=64911	12
Forum	view discussion	discuss.php?d=64900&parent=144338	14
Forum	view discussion	discuss.php?d=65037	5
Forum	view discussion	discuss.php?d=65049	6
Forum	view discussion	discuss.php?d=65060	4
Forum	view discussion	discuss.php?d=65070	5
Forum	view discussion	discuss.php?d=65193	2
Forum	view discussion	discuss.php?d=65202	6
Forum	view discussion	discuss.php?d=65214	3
Forum	view discussion	discuss.php?d=59205	1
Forum	subscribe	-----	0
Forum	view subscribers	-----	0
Forum	add discussion	discuss.php?d=57719	1
Forum	add discussion	discuss.php?d=57736	1
Forum	add discussion	discuss.php?d=57786	1
Forum	add discussion	discuss.php?d=57830	1
Forum	add discussion	discuss.php?d=57836	1
Forum	add discussion	discuss.php?d=57846	1
Forum	add discussion	discuss.php?d=57928	1
Forum	add discussion	discuss.php?d=57978	1
Forum	add discussion	discuss.php?d=58028	1
Forum	add discussion	discuss.php?d=58035	1
Forum	add discussion	discuss.php?d=58144	1
Forum	add discussion	discuss.php?d=58308	1

Forum	add discussion	discuss.php?d=58451	1
Forum	add discussion	discuss.php?d=58720	1
Forum	add discussion	discuss.php?d=58723	1
Forum	add discussion	discuss.php?d=58847	1
Forum	add discussion	discuss.php?d=58848	1
Forum	add discussion	discuss.php?d=58853	1
Forum	add discussion	discuss.php?d=58863	1
Forum	add discussion	discuss.php?d=58882	1
Forum	add discussion	discuss.php?d=58883	1
Forum	add discussion	discuss.php?d=58886	1
Forum	add discussion	discuss.php?d=58895	1
Forum	add discussion	discuss.php?d=58985	1
Forum	add discussion	discuss.php?d=58986	1
Forum	add discussion	discuss.php?d=59108	1
Forum	add discussion	discuss.php?d=59205	1
Forum	add discussion	discuss.php?d=60078	1
Forum	add discussion	discuss.php?d=60080	1
Forum	add discussion	discuss.php?d=60282	1
Forum	add discussion	discuss.php?d=60609	1
Forum	add discussion	discuss.php?d=64900	1
Forum	add discussion	discuss.php?d=64911	1
Forum	add discussion	discuss.php?d=65037	1
Forum	add discussion	discuss.php?d=65049	1
Forum	add discussion	discuss.php?d=65060	1
Forum	add discussion	discuss.php?d=65070	1
Forum	add discussion	discuss.php?d=65193	1
Forum	add discussion	discuss.php?d=65202	1
Forum	add discussion	discuss.php?d=65214	1
Forum	add post	discuss.php?d=58308&parent=132530	1
Forum	add post	discuss.php?d=58723&parent=132558	1
Forum	add post	discuss.php?d=57928&parent=132563	1
Forum	add post	discuss.php?d=64900&parent=144338	1
Forum	delete discussion	view.php?id=141631	2
Forum	delete post	-----	0
Forum	user report	user.php?id=28985&user=4224&mode=posts	1
Forum	unsubscribe	view.php?f=13085	1
Forum	update post	discuss.php?d=57719#p130214&parent=130214	2
Forum	update post	discuss.php?d=58028#p130942&parent=130942	1
Forum	update post	discuss.php?d=58035#p130971&parent=130971	5
Forum	update post	discuss.php?d=58451#p131861&parent=131861	1
Forum	update post	discuss.php?d=58882#p132668&parent=132668	2
Forum	update post	discuss.php?d=59108#p133094&parent=133094	1
Forum	update post	discuss.php?d=64900#p144133&parent=144133	1
Forum	update post	discuss.php?d=64911#p144160&parent=144160	1
Forum	update post	discuss.php?d=64900#p144338&parent=144338	3

	Forum	search	search.php?id=28985&search=	1
	Blog	view	index.php?filtertype=user&filterselect=4224&postid=0&tagid=0&tag=	1
	Resource	view	view.php?id=160107	22
	Resource	view all	index.php?id=28985	1
	User	view	view.php?id=7070&course=28985	12
	User	view	view.php?id=30035&course=28985	9
	User	view	view.php?id=1751&course=28985	1
	User	view	view.php?id=34752&course=28985	4
	User	view	view.php?id=4249&course=28985	5
	User	view	view.php?id=4224&course=28985	4
	User	view	view.php?id=50632&course=28985	2
	User	view	view.php?id=38&course=28985	4
	User	view	view.php?id=41301&course=28985	9
	User	view	view.php?id=1750&course=28985	4
	User	view	view.php?id=4234&course=28985	2
	User	view	view.php?id=67079&course=28985	1
	User	view	view.php?id=4232&course=28985	5
	User	view	view.php?id=6672&course=28985	4
	User	view	view.php?id=4240&course=28985	3
	User	view	view.php?id=5636&course=28985	1
	User	view	view.php?id=5641&course=28985	1
	User	view	view.php?id=15865&course=28985	1
	User	view	view.php?id=16011&course=28985	1
	User	view	view.php?id=4293&course=28985	1
	User	view	view.php?id=6190&course=28985	1
	User	view	view.php?id=4244&course=28985	1
	User	view all	index.php?id=28985	87
	User	update	view.php?id=30035&course=28985	1
	User	update	view.php?id=41301&course=28985	2
	Utpl	role_assign	admin/roles/assign.php?contextid=450089&roleid=3	1
	Utpl	role_assign	admin/roles/assign.php?contextid=450089&roleid=5	24
	Total			2097
Sistemas Basados en el Conocimiento	Course	view	view.php?id=28991	884
	Course	user report	user.php?id=28991&user=42898&mode=outline	1
	Course	user report	user.php?id=28991&user=42898&mode=complete	1
	Assignment	view	view.php?id=139734	21
	Assignment	view	view.php?id=151734	23
	Assignment	view all	index.php?id=28991	32
	Assignment	upload	view.php?a=16289	9
	Assignment	upload	view.php?a=18201	10
	Quiz	view	view.php?id=139682	19
	Quiz	view	view.php?id=151682	24
	Quiz	view all	index.php?id=28991	13
	Quiz	attempt	review.php?attempt=66173	1
	Quiz	attempt	review.php?attempt=67490	1

Quiz	attempt	review.php?attempt=68779	1
Quiz	attempt	review.php?attempt=69010	1
Quiz	attempt	review.php?attempt=70063	1
Quiz	attempt	review.php?attempt=72011	1
Quiz	attempt	review.php?attempt=72896	1
Quiz	attempt	review.php?attempt=78482	1
Quiz	attempt	review.php?attempt=79776	1
Quiz	attempt	review.php?attempt=80410	1
Quiz	attempt	review.php?attempt=80872	1
Quiz	attempt	review.php?attempt=81582	1
Quiz	attempt	review.php?attempt=82056	1
Quiz	attempt	review.php?attempt=84612	1
Quiz	attempt	review.php?attempt=85303	1
Quiz	attempt	review.php?attempt=86258	1
Quiz	attempt	review.php?attempt=86917	1
Quiz	continue attemp	review.php?attempt=80872	1
Quiz	continue attemp	review.php?attempt=82056	1
Quiz	continue attemp	review.php?attempt=86258	3
Quiz	close attempt	review.php?attempt=66173	1
Quiz	close attempt	review.php?attempt=67490	1
Quiz	close attempt	review.php?attempt=68779	1
Quiz	close attempt	review.php?attempt=69010	1
Quiz	close attempt	review.php?attempt=70063	1
Quiz	close attempt	review.php?attempt=72011	1
Quiz	close attempt	review.php?attempt=72896	1
Quiz	close attempt	review.php?attempt=78482	1
Quiz	close attempt	review.php?attempt=79776	1
Quiz	close attempt	review.php?attempt=80410	1
Quiz	close attempt	review.php?attempt=80872	1
Quiz	close attempt	review.php?attempt=81582	1
Quiz	close attempt	review.php?attempt=82056	1
Quiz	close attempt	review.php?attempt=84612	1
Quiz	close attempt	review.php?attempt=85303	1
Quiz	close attempt	review.php?attempt=86258	1
Quiz	close attempt	review.php?attempt=86917	1
Forum	view forum	view.php?id=142594	119
Forum	view forum	view.php?id=156017	104
Forum	view forums	index.php?id=28991	84
Forum	view discussion	discuss.php?d=57976&parent=131120	2
Forum	view discussion	discuss.php?d=57976&parent=131406	1
Forum	view discussion	discuss.php?d=57976&parent=131266	1
Forum	view discussion	discuss.php?d=57976&parent=131224	1
Forum	view discussion	discuss.php?d=57976&parent=131424	2
Forum	view discussion	discuss.php?d=57976	1
Forum	view discussion	discuss.php?d=62087&parent=139834	2

Forum	view discussion	discuss.php?d=62087&parent=140033	1
Forum	view discussion	discuss.php?d=62087&parent=140051	1
Forum	view discussion	discuss.php?d=62087&parent=141392	1
Forum	view discussion	discuss.php?d=62087&parent=141305	1
Forum	view discussion	discuss.php?d=62087	1
Forum	view discussion	discuss.php?d=62087&parent=140165	1
Forum	subscribe	view.php?f=13159	3
Forum	subscribe	view.php?f=14288	2
Forum	view subscribers	subscribers.php?id=14288	1
Forum	add discussion	-----	0
Forum	add post	view.php?f=13159&parent=131120	1
Forum	add post	view.php?f=13159&parent=131224	1
Forum	add post	view.php?f=13159&parent=131266	1
Forum	add post	view.php?f=13159&parent=131406	1
Forum	add post	view.php?f=13159&parent=131424	1
Forum	add post	view.php?f=13159&parent=131629	1
Forum	add post	view.php?f=13159&parent=131663	1
Forum	add post	view.php?f=13159&parent=132370	1
Forum	add post	view.php?f=13159&parent=132417	1
Forum	add post	view.php?f=13159&parent=132436	1
Forum	add post	view.php?f=13159&parent=132457	1
Forum	add post	view.php?f=13159&parent=132697	1
Forum	add post	view.php?f=13159&parent=132744	1
Forum	add post	view.php?f=13159&parent=132837	1
Forum	add post	view.php?f=13159&parent=133035	1
Forum	add post	view.php?f=14288&parent=139834	1
Forum	add post	view.php?f=14288&parent=140033	1
Forum	add post	view.php?f=14288&parent=140051	1
Forum	add post	view.php?f=14288&parent=140165	1
Forum	add post	view.php?f=14288&parent=140630	1
Forum	add post	view.php?f=14288&parent=140743	1
Forum	add post	view.php?f=14288&parent=140800	1
Forum	add post	view.php?f=14288&parent=141010	1
Forum	add post	view.php?f=14288&parent=141305	1
Forum	add post	view.php?f=14288&parent=141392	1
Forum	add post	view.php?f=14288&parent=141441	1
Forum	add post	view.php?f=14288&parent=141922	1
Forum	add post	view.php?f=14288&parent=142120	1
Forum	add post	view.php?f=14288&parent=142464	1
Forum	delete discussion	-----	0
Forum	delete post	-----	0
Forum	user report	user.php?id=28991&user=56105&mode=posts	1
Forum	user report	user.php?id=28991&user=3&mode=posts	1
Forum	unsubscribe	-----	0
Forum	update post	view.php?f=13159&parent=130809	1

Forum	update post	view.php?f=13159&parent=131224	1
Forum	update post	view.php?f=13159&parent=131629	1
Forum	update post	view.php?f=13159&parent=132436	1
Forum	update post	view.php?f=14288&parent=139834	3
Forum	update post	view.php?f=14288&parent=140033	3
Forum	update post	view.php?f=14288&parent=141392	2
Forum	search	-----	0
Blog	view	-----	0
Resource	view	view.php?id=137179	26
Resource	view	view.php?id=137180	26
Resource	view	view.php?id=145533	15
Resource	view	view.php?id=145534	13
Resource	view	view.php?id=145535	13
Resource	view	view.php?id=146104	25
Resource	view	view.php?id=147743	9
Resource	view	view.php?id=147744	18
Resource	view	view.php?id=153070	22
Resource	view	view.php?id=153071	15
Resource	view	view.php?id=154495	18
Resource	view	view.php?id=154497	17
Resource	view	view.php?id=156020	35
Resource	view	view.php?id=158359	11
Resource	view	view.php?id=158361	9
Resource	view	view.php?id=159304	23
Resource	view all	index.php?id=28991	30
User	view	view.php?id=29215&course=28991	7
User	view	view.php?id=30035&course=28991	4
User	view	view.php?id=22865&course=28991	1
User	view	view.php?id=1496&course=28991	1
User	view	view.php?id=16236&course=28991	2
User	view	view.php?id=41301&course=28991	5
User	view	view.php?id=18&course=28991	9
User	view	view.php?id=4232&course=28991	2
User	view	view.php?id=7078&course=28991	1
User	view	view.php?id=1488&course=28991	2
User	view	view.php?id=48077&course=28991	5
User	view	view.php?id=4218&course=28991	6
User	view	view.php?id=42898&course=28991	4
User	view	view.php?id=4224&course=28991	1
User	view	view.php?id=38897&course=28991	3
User	view	view.php?id=56105&course=28991	1
User	view	view.php?id=3&course=28991	5
User	view	view.php?id=5972&course=28991	1
User	view	view.php?id=4282&course=28991	1
User	view	view.php?id=6672&course=28991	2

	User	view	view.php?id=4249&course=28991	1
	User	view all	index.php?id=28991	91
	User	update	-----	0
	Utpl	role_assign	admin/roles/assign.php?contextid=450095&roleid=3	1
	Utpl	role_assign	admin/roles/assign.php?contextid=450095&roleid=5	24
	Total			1953
Teoría de Colas	Course	view	view.php?id=28990	942
	Course	user report	user.php?id=28990&user=77056&mode=outline	17
	Course	user report	user.php?id=28990&user=77056&mode=alllogs	1
	Course	user report	user.php?id=28990&user=77056&mode=stats	1
	Course	user report	user.php?id=28990&user=77056&mode=complete	4
	Assignment	view	view.php?id=139731	28
	Assignment	view	view.php?id=151731	21
	Assignment	view all	index.php?id=28990	52
	Assignment	upload	view.php?a=16286	9
	Assignment	upload	view.php?a=18198	11
	Quiz	view	view.php?id=139679	33
	Quiz	view	view.php?id=151679	26
	Quiz	view all	index.php?id=28990	35
	Quiz	attempt	review.php?attempt=64144	1
	Quiz	attempt	review.php?attempt=67752	1
	Quiz	attempt	review.php?attempt=69054	1
	Quiz	attempt	review.php?attempt=70001	1
	Quiz	attempt	review.php?attempt=70097	1
	Quiz	attempt	review.php?attempt=71295	1
	Quiz	attempt	review.php?attempt=72839	1
	Quiz	attempt	review.php?attempt=72851	1
	Quiz	attempt	review.php?attempt=73542	1
	Quiz	attempt	review.php?attempt=73559	1
	Quiz	attempt	review.php?attempt=78331	1
	Quiz	attempt	review.php?attempt=78381	1
	Quiz	attempt	review.php?attempt=81553	1
	Quiz	attempt	review.php?attempt=82306	1
	Quiz	attempt	review.php?attempt=82376	1
	Quiz	attempt	review.php?attempt=83131	1
	Quiz	attempt	review.php?attempt=86939	1
	Quiz	attempt	review.php?attempt=86962	1
	Quiz	attempt	review.php?attempt=87650	1
	Quiz	continue attemp	review.php?attempt=64144	1
	Quiz	continue attemp	review.php?attempt=67752	1
Quiz	continue attemp	review.php?attempt=70097	1	
Quiz	close attempt	review.php?attempt=67752	1	
Quiz	close attempt	review.php?attempt=69054	1	
Quiz	close attempt	review.php?attempt=70001	1	
Quiz	close attempt	review.php?attempt=70097	1	

Quiz	close attempt	review.php?attempt=71295	1
Quiz	close attempt	review.php?attempt=72851	1
Quiz	close attempt	review.php?attempt=72839	1
Quiz	close attempt	review.php?attempt=73542	1
Quiz	close attempt	review.php?attempt=73559	1
Quiz	close attempt	review.php?attempt=78331	1
Quiz	close attempt	review.php?attempt=78381	1
Quiz	close attempt	review.php?attempt=81553	1
Quiz	close attempt	review.php?attempt=82306	1
Quiz	close attempt	review.php?attempt=82376	1
Quiz	close attempt	review.php?attempt=83131	1
Quiz	close attempt	review.php?attempt=86962	1
Quiz	close attempt	review.php?attempt=86939	1
Quiz	close attempt	review.php?attempt=87650	1
Forum	view forum	view.php?id=157829	49
Forum	view forums	index.php?id=28990	4
Forum	view discussion	discuss.php?d=64241&parent=142948	1
Forum	view discussion	discuss.php?d=64241&parent=143112	1
Forum	view discussion	discuss.php?d=64241	1
Forum	subscribe	-----	0
Forum	view subscribers	-----	0
Forum	add discussion	-----	0
Forum	add post	view.php?f=14453&parent=142948	1
Forum	add post	view.php?f=14453&parent=143112	1
Forum	add post	view.php?f=14453&parent=143235	1
Forum	add post	view.php?f=14453&parent=144162	1
Forum	add post	view.php?f=14453&parent=144237	1
Forum	delete discussion	-----	0
Forum	delete post	-----	0
Forum	user report	user.php?id=28990&user=77056&mode=posts	4
Forum	user report	user.php?id=28990&user=77056&mode=discussions	1
Forum	unsubscribe	-----	0
Forum	update post	-----	0
Forum	search	-----	0
Blog	view	-----	0
Resource	view	-----	0
Resource	view all	-----	0
User	view	view.php?id=1496&course=28990	1
User	view	view.php?id=1496&course=28990	1
User	view	view.php?id=30841&course=28990	4
User	view	view.php?id=1752&course=28990	2
User	view	view.php?id=48084&course=28990	1
User	view	view.php?id=48077&course=28990	4
User	view	view.php?id=77056&course=28990	18
User	view	view.php?id=2185&course=28990	1

	User	view	view.php?id=34803&course=28990	1
	User	view	view.php?id=4291&course=28990	2
	User	view	view.php?id=31697&course=28990	1
	User	view	view.php?id=30838&course=28990	1
	User	view	view.php?id=16976&course=28990	1
	User	view	view.php?id=5519&course=28990	1
	User	view	view.php?id=1419&course=28990	1
	User	view	view.php?id=37&course=28990	2
	User	view	view.php?id=76997&course=28990	1
	User	view all	index.php?id=28990	56
	User	update	-----	0
	Utpl	role_assign	admin/roles/assign.php?contextid=450094&roleid=3	2
	Utpl	role_assign	admin/roles/assign.php?contextid=450094&roleid=5	35
	Total			1422
Planeación Estratégica	Course	view	view.php?id=28984	845
	Course	user report	-----	0
	Assignment	view	view.php?id=139730	15
	view	view	view.php?id=151730	10
	Assignment	view all	index.php?id=28984	42
	Assignment	upload	view.php?a=16285	5
	Assignment	upload	view.php?a=18197	5
	Quiz	view	view.php?id=139678	9
	Quiz	view	view.php?id=145168	45
	Quiz	view	view.php?id=151678	10
	Quiz	view	view.php?id=159301	38
	Quiz	view all	index.php?id=28984	30
	Quiz	attempt	review.php?attempt=68075	1
	Quiz	attempt	review.php?attempt=71121	1
	Quiz	attempt	review.php?attempt=73054	1
	Quiz	attempt	review.php?attempt=73747	1
	Quiz	attempt	review.php?attempt=73748	1
	Quiz	attempt	review.php?attempt=73951	1
	Quiz	attempt	review.php?attempt=74091	1
	Quiz	attempt	review.php?attempt=74288	1
	Quiz	attempt	review.php?attempt=74350	1
	Quiz	attempt	review.php?attempt=74949	1
	Quiz	attempt	review.php?attempt=75209	1
	Quiz	attempt	review.php?attempt=75210	1
	Quiz	attempt	review.php?attempt=75532	1
	Quiz	attempt	review.php?attempt=75536	1
	Quiz	attempt	review.php?attempt=78480	1
	Quiz	attempt	review.php?attempt=79765	1
	Quiz	attempt	review.php?attempt=81657	1
	Quiz	attempt	review.php?attempt=82134	1
Quiz	attempt	review.php?attempt=83717	1	

Quiz	attempt	review.php?attempt=88610	1
Quiz	attempt	review.php?attempt=88717	1
Quiz	attempt	review.php?attempt=88719	1
Quiz	attempt	review.php?attempt=88720	1
Quiz	attempt	review.php?attempt=89604	1
Quiz	attempt	review.php?attempt=89607	1
Quiz	attempt	review.php?attempt=89790	1
Quiz	attempt	review.php?attempt=89809	1
Quiz	attempt	review.php?attempt=89813	1
Quiz	continue attemp	review.php?attempt=74288	2
Quiz	continue attemp	review.php?attempt=82134	1
Quiz	close attempt	review.php?attempt=68075	1
Quiz	close attempt	review.php?attempt=71121	1
Quiz	close attempt	review.php?attempt=73054	1
Quiz	close attempt	review.php?attempt=73747	1
Quiz	close attempt	review.php?attempt=73748	1
Quiz	close attempt	review.php?attempt=73951	1
Quiz	close attempt	review.php?attempt=74091	1
Quiz	close attempt	review.php?attempt=74350	1
Quiz	close attempt	review.php?attempt=74949	1
Quiz	close attempt	review.php?attempt=75209	1
Quiz	close attempt	review.php?attempt=75532	1
Quiz	close attempt	review.php?attempt=75536	1
Quiz	close attempt	review.php?attempt=74288	1
Quiz	close attempt	review.php?attempt=78480	1
Quiz	close attempt	review.php?attempt=79765	1
Quiz	close attempt	review.php?attempt=81657	1
Quiz	close attempt	review.php?attempt=82134	1
Quiz	close attempt	review.php?attempt=83717	1
Quiz	close attempt	review.php?attempt=88610	1
Quiz	close attempt	review.php?attempt=88717	1
Quiz	close attempt	review.php?attempt=88719	1
Quiz	close attempt	review.php?attempt=89604	1
Quiz	close attempt	review.php?attempt=89607	1
Quiz	close attempt	review.php?attempt=89790	1
Quiz	close attempt	review.php?attempt=89809	1
Quiz	close attempt	review.php?attempt=89813	1
Forum	view forum	view.php?id=141837	122
Forum	view forum	view.php?id=159224	58
Forum	view forums	index.php?id=28984	126
Forum	view discussion	discuss.php?d=57733	16
Forum	view discussion	discuss.php?d=57807	15
Forum	view discussion	discuss.php?d=57822	15
Forum	view discussion	discuss.php?d=57993	10
Forum	view discussion	discuss.php?d=58004	8

Forum	view discussion	discuss.php?d=58034	7
Forum	view discussion	discuss.php?d=58070	6
Forum	view discussion	discuss.php?d=58111	6
Forum	view discussion	discuss.php?d=58873	4
Forum	view discussion	discuss.php?d=58884	5
Forum	view discussion	discuss.php?d=58993	6
Forum	view discussion	discuss.php?d=59107	12
Forum	view discussion	discuss.php?d=58948	8
Forum	view discussion	discuss.php?d=64734	17
Forum	view discussion	discuss.php?d=64734&parent=143837	4
Forum	view discussion	discuss.php?d=64734&parent=143792	1
Forum	view discussion	discuss.php?d=64757	11
Forum	view discussion	discuss.php?d=64788	6
Forum	view discussion	discuss.php?d=64796	9
Forum	view discussion	discuss.php?d=64901	12
Forum	view discussion	discuss.php?d=64932	9
Forum	view discussion	discuss.php?d=65132	5
Forum	view discussion	discuss.php?d=65234	3
Forum	subscribe	view.php?f=14589	1
Forum	view subscribers	-----	0
Forum	add discussion	discuss.php?d=57733	1
Forum	add discussion	discuss.php?d=57807	1
Forum	add discussion	discuss.php?d=57822	1
Forum	add discussion	discuss.php?d=57993	1
Forum	add discussion	discuss.php?d=58004	1
Forum	add discussion	discuss.php?d=58034	1
Forum	add discussion	discuss.php?d=58070	1
Forum	add discussion	discuss.php?d=58111	1
Forum	add discussion	discuss.php?d=58873	1
Forum	add discussion	discuss.php?d=58884	1
Forum	add discussion	discuss.php?d=58948	1
Forum	add discussion	discuss.php?d=58993	1
Forum	add discussion	discuss.php?d=59107	1
Forum	add discussion	discuss.php?d=64734	1
Forum	add discussion	discuss.php?d=64757	1
Forum	add discussion	discuss.php?d=64788	1
Forum	add discussion	discuss.php?d=64796	1
Forum	add discussion	discuss.php?d=64901	1
Forum	add discussion	discuss.php?d=64932	1
Forum	add discussion	discuss.php?d=65132	1
Forum	add discussion	discuss.php?d=65234	1
Forum	add post	discuss.php?d=59107&parent=133626	1
Forum	add post	discuss.php?d=58993&parent=135758	1
Forum	add post	discuss.php?d=58034&parent=135759	1
Forum	add post	discuss.php?d=57733&parent=135761	1

Forum	add post	discuss.php?d=57822&parent=135765	1
Forum	add post	discuss.php?d=57993&parent=135766	1
Forum	add post	discuss.php?d=58948&parent=135767	1
Forum	add post	discuss.php?d=57807&parent=135769	1
Forum	add post	discuss.php?d=58111&parent=135771	1
Forum	add post	discuss.php?d=59107&parent=135772	1
Forum	add post	discuss.php?d=58070&parent=135775	1
Forum	add post	discuss.php?d=58948&parent=135777	1
Forum	add post	discuss.php?d=58884&parent=135778	1
Forum	add post	discuss.php?d=58873&parent=136442	1
Forum	add post	discuss.php?d=58004&parent=136444	1
Forum	add post	discuss.php?d=59107&parent=136446	1
Forum	add post	discuss.php?d=64734&parent=143837	1
Forum	add post	discuss.php?d=64757&parent=143996	1
Forum	add post	discuss.php?d=64788&parent=143997	1
Forum	add post	discuss.php?d=64796&parent=143999	1
Forum	add post	discuss.php?d=64932&parent=144355	1
Forum	add post	discuss.php?d=65132&parent=144710	1
Forum	add post	discuss.php?d=64901&parent=144712	1
Forum	add post	discuss.php?d=64901&parent=144713	1
Forum	add post	discuss.php?d=65234&parent=145346	1
Forum	delete discussion	-----	0
Forum	delete post	discuss.php?d=58948	1
Forum	delete post	discuss.php?d=64901	1
Forum	user report	-----	0
Forum	unsubscribe	view.php?f=13106	1
Forum	unsubscribe	view.php?f=14589	1
Forum	update post	discuss.php?d=57807#p130419&parent=130419	1
Forum	update post	discuss.php?d=58993#p132849&parent=132849	1
Forum	update post	discuss.php?d=59107#p133091&parent=133091	2
Forum	update post	discuss.php?d=64734#p143792&parent=143792	2
Forum	update post	discuss.php?d=64757#p143846&parent=143846	1
Forum	search	search.php?id=28984&search=	4
Blog	view	index.php?filtertype=course&filtersselect=28984&postid=0&tagid=0&tag	1
Resource	view	view.php?id=148011	10
Resource	view	view.php?id=157416	66
Resource	view	view.php?id=158061	15
Resource	view	view.php?id=159268	14
Resource	view	view.php?id=159791	3
Resource	view	view.php?id=159871	5
Resource	view	view.php?id=158251	1
Resource	view all	index.php?id=28984	9
User	view	view.php?id=1753&course=28984	3
User	view	view.php?id=1755&course=28984	3

User	view	view.php?id=7067&course=28984	3
User	view	view.php?id=41301&course=28984	2
User	view	view.php?id=16236&course=28984	3
User	view	view.php?id=5787&course=28984	3
User	view	view.php?id=7070&course=28984	5
User	view	view.php?id=29215&course=28984	4
User	view	view.php?id=4246&course=28984	1
User	view	view.php?id=4249&course=28984	2
User	view	view.php?id=1751&course=28984	2
User	view	view.php?id=5972&course=28984	1
User	view all	index.php?id=28984	26
User	update	view.php?id=1755&course=28984	1
Utpl	role_assign	admin/roles/assign.php?contextid=450088&roleid=5	16
Utpl	role_assign	admin/roles/assign.php?contextid=450088&roleid=3	1
Total			1874

Tabla 39. Datos de interacción en el EVA de los estudiantes de las materias previas a la minería.

Con el procesamiento de los datos extraídos de la tabla de Logs del EVA de la UTPL, con un número total de 126 estudiantes de la carrera de Informática que visitaron e interactuaron con las actividades del sistema; foros, tareas, mensajes, y demás herramientas disponibles en los cursos del sistema, de los cuales, podemos observar que la ruta de navegación que tienen los estudiantes cuando interactúan con el sistema, es que con mayor frecuencia ingresan a revisar las diferentes actividades propuestas por el tutor, para luego visitar la actividad donde el docente haya ingresado algún trabajo, ya sean tareas, foros, cuestionario, recursos, etc. y para la comunicación con el mismo el uso de herramientas como: mensajería, twitter, blogs, chat, etc. Los valores donde existe mayor número de interacción por parte de los estudiantes con las actividades del sistema son las páginas de los foros, esto podría deberse a que los foros son la actividad que les permiten expresar ideologías, pensamientos y conocimientos acerca de algún tema propuesto por el docente con sus compañeros.

ANEXO E

TABLA CON LOS DATOS PREVIOS A LA MINERÍA

La tabla representa el total de accesos a las herramientas y acciones realizadas en los cursos y permitirán determinar las nuevas entidades de la base con los indicadores de navegación.

Fuente: Ruiz, F., (2014).

N°	Id Estudiante	Assignment	Blog	Course	forum	Quiz	Resource	twitter	message	utpl	Chat	User
1	23611	24	0	112	82	108	6	0	0	0	0	11
2	4242	8	0	22	55	10	3	0	0	0	0	4
3	4216	0	0	29	44	0	6	0	0	0	0	3
4	35744	14	0	22	61	15	6	0	0	0	0	1
5	16056	0	0	31	18	1	3	0	0	0	0	2
6	5641	0	1	18	33	0	4	0	0	0	0	1
7	4255	2	0	26	93	0	6	0	0	0	0	2
8	4254	0	0	21	83	0	3	0	0	0	0	0
9	34752	0	0	34	47	0	4	0	0	0	0	1
10	7067	4	0	34	29	3	5	0	0	0	0	3
11	4218	8	0	17	73	12	2	0	0	0	0	2
12	57887	6	0	58	87	9	8	0	0	0	0	5
13	4198	0	0	28	50	0	3	0	0	0	0	1
14	2185	4	0	49	55	2	5	0	0	0	0	9
16	38	32	0	135	20	147	1	0	0	0	0	16
17	15865	4	0	98	32	4	2	0	0	0	0	4
18	1751	0	0	24	17	0	0	0	0	0	0	6
19	1750	0	0	74	69	0	1	0	0	0	0	19
20	4293	3	0	72	39	6	4	0	0	0	0	8
21	67079	0	0	16	30	1	0	0	0	0	0	3
22	4246	0	0	10	1	0	0	0	0	0	0	0
23	6190	0	0	12	5	1	0	0	0	0	0	3
24	4232	3	0	34	10	4	1	0	0	0	0	20
25	6576	0	0	23	39	0	0	0	0	0	0	1
26	5641	0	0	24	28	1	1	0	0	0	0	5
27	41301	3	0	36	14	4	1	0	0	0	0	9
28	4243	0	0	5	0	0	0	0	0	0	0	0
29	4240	5	0	59	37	8	2	0	0	0	0	0
30	4244	3	0	42	63	4	2	0	0	0	0	13
31	30035	16	0	62	43	8	1	0	0	0	0	23
32	4249	6	0	57	71	12	4	0	0	0	0	3
33	34752	0	0	26	12	0	0	0	0	0	0	6
34	16011	0	0	9	7	0	0	0	0	0	0	0
35	4224	1	0	20	27	0	0	0	0	0	0	0
36	7070	7	0	73	125	4	1	0	0	0	0	14

37	5636	1	0	19	30	2	0	0	0	0	0	8
38	50632	1	0	6	19	1	0	0	0	0	0	0
39	4234	7	1	46	43	6	3	0	0	0	0	5
40	6672	7	0	0	0	10	0	0	0	0	0	0
42	3	44	0	215	29	65	32	4	0	0	0	0
43	56105	0	0	13	20	0	6	0	0	0	0	7
44	5972	8	0	24	10	13	28	5	0	0	0	1
45	4246	0	0	1	0	0	0	0	0	0	0	0
46	6190	0	0	8	0	0	4	0	0	0	0	5
47	6408	0	0	18	8	0	19	0	0	0	0	0
48	4232	0	0	30	9	0	11	0	0	0	0	16
49	42898	6	0	38	9	8	9	3	0	0	0	2
50	18	0	0	83	23	0	30	0	0	0	0	38
51	41301	3	0	38	13	4	13	3	0	0	0	0
52	1488	0	0	79	29	0	23	0	0	0	0	10
53	48077	6	0	70	25	8	9	3	0	0	0	10
54	30035	15	0	44	13	8	11	3	0	0	0	12
55	4249	4	0	45	48	9	16	4	0	0	0	0
56	4224	0	0	29	6	0	38	0	0	0	0	0
57	4218	16	0	18	24	16	9	4	0	0	0	0
58	38897	6	0	30	9	14	23	4	0	0	0	8
59	29215	1	0	46	16	0	24	0	0	0	0	8
60	7078	3	0	31	20	1	16	1	0	0	0	2
61	16236	2	0	21	18	0	2	0	0	0	0	4
62	22865	3	0	14	4	4	13	3	0	0	0	8
63	5257	0	0	1	0	0	0	0	0	0	0	0
64	1496	1	0	14	26	0	3	0	0	0	0	20
65	4282	1	0	14	16	0	5	0	0	0	0	4
66	6672	6	0	0	0	9	0	4	0	0	0	0
68	1419	0	0	4	0	42	0	0	0	0	0	0
69	34803	1	0	74	1	2	0	0	0	0	0	24
70	1496	0	0	16	1	0	0	0	0	0	0	27
71	30841	0	0	34	2	0	0	0	0	0	0	1
72	4295	0	0	22	1	0	0	0	0	0	0	0
73	2185	3	0	18	1	1	0	0	0	0	0	6
74	5257	0	0	3	0	0	0	0	0	0	0	0
75	4198	0	0	23	2	0	0	0	0	0	0	0
76	4291	0	0	17	0	0	0	0	0	0	0	3
77	30838	2	0	28	2	1	0	0	0	0	0	0
78	4228	4	0	14	1	6	0	0	0	0	0	0
79	1754	3	0	20	0	0	0	0	0	0	0	7
80	31697	1	0	101	8	0	0	0	0	0	0	2
81	1752	0	0	12	1	0	0	0	0	0	0	0
82	16976	3	0	26	2	1	0	0	0	0	0	4
83	4251	6	0	16	2	10	0	0	0	0	0	1

84	5519	1	0	25	1	2	0	0	0	0	0	0
85	4210	1	0	20	10	1	0	0	0	0	0	0
86	2852	0	0	20	1	0	0	0	0	0	0	0
87	47013	6	0	10	0	4	0	0	0	0	0	0
88	7065	0	0	21	0	3	0	0	0	0	0	0
89	5743	0	0	30	3	0	0	0	0	0	0	0
90	48077	6	0	64	2	8	0	0	0	0	0	0
91	76997	2	0	65	6	2	0	0	0	0	0	0
92	48084	8	0	19	1	9	0	0	0	0	0	0
93	47965	6	0	11	0	9	0	0	0	0	0	2
94	56358	0	0	5	0	2	0	0	0	0	0	0
95	77056	14	0	91	10	20	0	1	0	0	0	19
96	38906	4	0	9	0	2	0	0	0	0	0	0
97	15788	9	0	6	0	9	0	0	0	0	0	0
98	4222	0	0	20	0	1	0	0	0	0	0	1
99	67385	12	0	15	0	14	0	0	0	0	0	0
100	58401	3	0	4	0	3	0	0	0	0	0	0
101	51278	0	0	9	4	0	0	0	0	0	0	0
102	40738	3	0	10	1	4	0	0	0	0	0	0
103	5669	10	0	13	0	16	0	0	0	0	0	0
104	37	42	0	56	5	48	0	0	0	0	0	1
106	1753	0	0	25	37	0	1	0	0	0	0	2
107	5787	0	0	138	16	12	11	0	0	0	0	2
108	2462	30	0	258	224	280	34	0	0	0	0	2
109	1755	7	0	11	6	9	2	0	0	0	0	2
110	15865	4	1	94	23	30	14	0	0	0	0	1
111	1751	0	0	25	13	1	2	0	0	0	0	1
112	5972	23	0	24	14	14	0	0	0	0	0	0
113	4246	0	0	6	2	0	0	0	0	0	0	0
114	41301	4	0	33	10	5	6	0	0	0	0	0
115	4249	4	0	53	70	15	4	0	0	0	0	1
116	7067	3	0	25	17	18	2	0	0	0	0	2
117	7070	5	0	72	64	45	18	0	0	0	0	2
118	29215	2	0	49	17	9	34	0	0	0	0	2
119	7078	5	0	50	36	14	4	0	0	0	0	1
120	8254	0	0	7	4	0	0	0	0	0	0	0
121	16236	3	0	16	13	0	1	0	0	0	0	2
122	5257	0	0	2	0	0	0	0	0	0	0	0

Tabla 40. Datos de interacción de los estudiantes en el EVA de las materias previas a la minería.

La **Tabla 40**, muestra las herramientas y actividades disponibles en el EVA, para el aprendizaje del estudiante y comunicación con su tutor y demás compañeros, en las cuales se detallan las páginas más frecuentadas por los estudiantes cuando interactúan con el EVA y su docente, para

su aprendizaje según los datos obtenidos con la herramienta de Excel, son los *foros, recursos, tareas, exámenes y el perfil de usuario*.

ANEXO F

CONEXIÓN DE LA BASE DE DATOS DEL EVA CON LA HERRAMIENTA WEKA

Los pasos para conectar WEKA con base de datos MySQL se los presenta a continuación:

La versión de Weka con la cual se está trabajando en esta tesis es Weka versión 3.6.3

REQUISITOS:

- Weka (weka.jar)
- MySQL Server instalado y corriendo
- Conector de MySQL para Java, versión 5.1.13

CLASSPATH & SCRIPT de INICIO de WEKA

Pasos de configuración para levantar Weka

El CLASSPATH se lo configura con la ruta a WEKA y al conector de JAVA, el cual se lo debe configurar en un script de arranque de WEKA.

Es importante el path donde van a ubicarlo, se lo puse en home de usuario: /home/Tita Ruiz

Por otro lado, el directorio que contiene a WEKA se encuentra en: /home/Tita Ruiz/PROGRAMAS/weka_2

El conector de JAVA y MySQL (tar.gz), hay que descomprimir el archivo y abrir el archivo .jar, dentro de la carpeta mysql-connector-java-5.1.13-bin.jar y copiarlo adentro del mismo directorio de WEKA, donde se encuentra el weka.jar.

El script quedara de la siguiente manera:

```
#!/bin/bash
WEKA_PATH=/home/Tita Ruiz/PROGRAMAS/weka_2
CP="$CLASSPATH:/etc/alternatives/java_sdk:$WEKA_PATH/mysql-connector-java-5.1.13-
bin.jar:$WEKA_PATH/weka.jar"
java -cp $CP -Xmx500m weka.gui.explorer.Explorer
```

Se escribe el comando: "java -jar weka", para que no se sobre monte las variable CLASSPATH que se creó. Además se levanta directamente el Explorer de WEKA.

DATABASEUTILS.PROPS

Hay que descomprimir el archivo weka.jar de la siguiente dirección:

```
# jar -xf weka.jar
```

Se creará un nuevo directorio llamado "Weka", y dentro de él un directorio "experiment", allí se encuentra un archivo denominado *DataBaseUtils.props*, al cual se lo debe copiar el mismo path, donde se encuentra el script de inicio, en el home de usuario, y al directorio de Weka:

```
# cp /home/Tita Ruiz/PROGRAMAS/weka_2/weka/experiment/DatabaseUtils.props /home/Tita Ruiz
# cp /home/Tita Ruiz/PROGRAMAS/weka_2/weka/experiment/DatabaseUtils.props /home/Tita Ruiz/PROGRAMAS/weka_2
```

En las tres ubicaciones debe ser modificadas y tener el mismo contenido dentro del archivo en /weka/experiment/

El archivo quedara de la siguiente manera:

```
# JDBC driver (comma-separated list)
jdbcDriver=org.gjt.mm.mysql.Driver
# database URL

jdbcURL=jdbc:mysql://localhost:3306/mysql

#Text
CHAR=0
TINYTEXT=0
TEXT=0
VARCHAR=0
LONGVARCHAR=0
BINARY=0
VARBINARY=0
LONGVARBINARY=0
BLOB=0
MEDIUMTEXT=0
MEDIUMBLOB=0
LONGTEXT=0
LONGBLOB=0
#
###Number types
#BIT=1
NUMERIC=2
DECIMAL=2
FLOAT=2
DOUBLE=2
TINYINT=3
SMALLINT=4
#SHORT=4
SHORT=5
INTEGER=5
INT=5
INT_UNSIGNED=5
MEDIUMINT=5
BIGINT=6
```

LONG=6

##Data Types

REAL=7

DATE=8

TIME=8

TIMESTAMP=0

DATETIME=0

##mappings for table creation

CREATE_STRING=VARCHAR(8000)

CREATE_INT=INT

CREATE_DOUBLE=DOUBLE

##database flags

checkUpperCaseNames=false

setAutoCommit=true

createIndex=false

#

##flags for loading and saving instances using DatabaseLoader/Saber

nominalToStringLimit=50

idColumn=auto_generated_id

#

#

#

#

other options

CREATE_DOUBLE=DOUBLE

CREATE_STRING=TEXT

CREATE_INT=INT

checkUpperCaseNames=false

checkLowerCaseNames=false

checkForTable=true

de la cual, se debe editar sólo la línea

jdbcURL=jdbc:mysql://localhost:3306/mysql

que es la base de datos a la cual se quieren conectar y el servidor.

LEVANTANDO WEKA y CONECTANDO A LA BASE DE DATOS

Finalmente se puede correr WEKA, para ello corremos el script de inicio que se creó en el segundo paso.

./weka (estando parados en /home/Tita Ruiz)

Debería haberse abierto el EXPLORADOR de WEKA:

- vamos a la opción "Open DB"
- En la URL digitamos la dirección de la base de datos a explorar.

jdbc:mysql://localhost:3306/MySQL

- Se digita el nombre de usuario y contraseña para la conexión con la base de datos
- y luego CONNECT
- Finalmente, se escribe una consulta a la base de datos, por ejemplo:

SELECT

mdl_log.userid, mdl_log.course, mdl_log.module, mdl_log.action, mdl_log.url,

FROM mdl_log

WHERE (module = 'course' AND (action = 'view'))

AND course = 51749

Y estará lista la conexión para trabajar con MySQL en WEKA.

ANEXO G

CONFIGURACIÓN DE GEPHI PARA EL ANÁLISIS GRAFICO

En este Anexo se presenta la configuración y adquisición de los datos de interacción de los estudiantes y profesor con las herramientas del EVA durante su navegación por el sistema.

Primeramente importamos el archivo .csv, el cual se lo exportó desde la base del EVA mediante consultas SQL de la tabla de los Log. una vez cargado el archivo la **Fig. 39** Muestra el grafo inicial con los datos del archivo; luego se procedió a configurar la herramienta de Gephi para obtener las interacciones de los estudiantes con las herramientas del EVA.

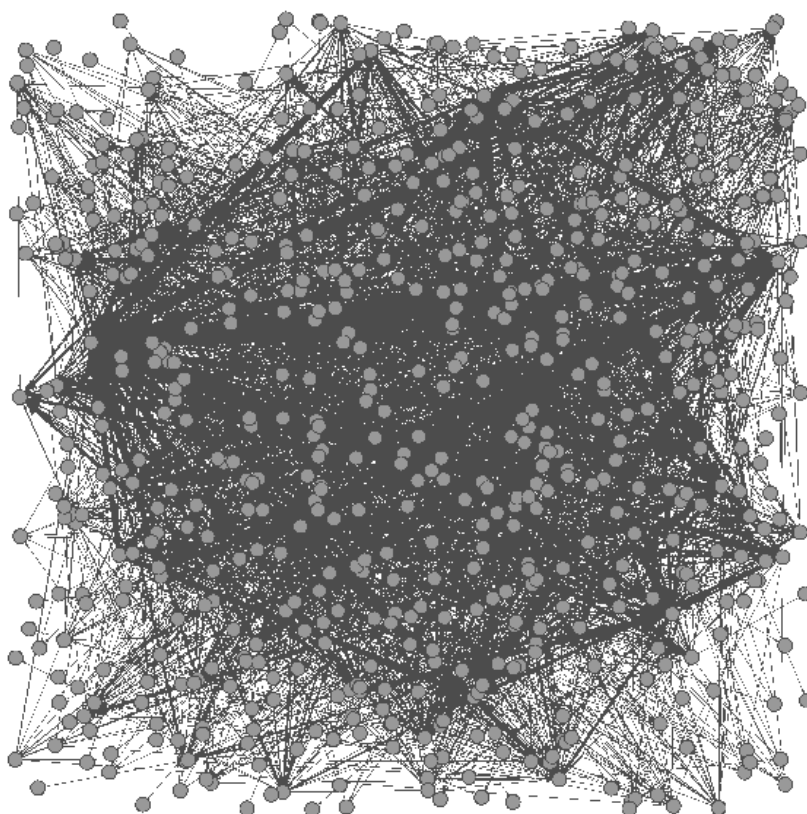


Fig. 39. Ponderada red de estudiantes que interactúan con la plataforma del EVA

Como se puede ver en la **Fig. 39**, el grafo de la interacción de los estudiantes en el EVA, presenta la posición de los nodos del grafo aleatoriamente y sin ningún detalle en particular del proceso de interacción ni los datos tanto de las herramientas a las que acceden los estudiantes dentro de los cursos, es por ello que se lo debe configurar según lo que se desee obtener del grafo.

Dentro de la gráfica se manipuló varias de las opciones del diseño gráfico, ya que los resultados que se obtengan del grafo dependerán bastante de los datos que se manipulen dentro de la herramienta; se escogió el algoritmo 'Force Atlas' para la distribución del grafo con los valores por defecto y con una fuerza de repulsión de 10000, se lo puede apreciar en la **Fig. 39.**, ya que este algoritmo presenta los grafos más compactos, legibles y además puede mostrar las 'authorities' más en el centro que los 'hubs' (opción Distribución de Atracción); la auto-estabilización mejora la convergencia al final del algoritmo. Esto con la finalidad de diseñar el grafo con el nivel de interacción de los estudiantes con las herramientas del EVA

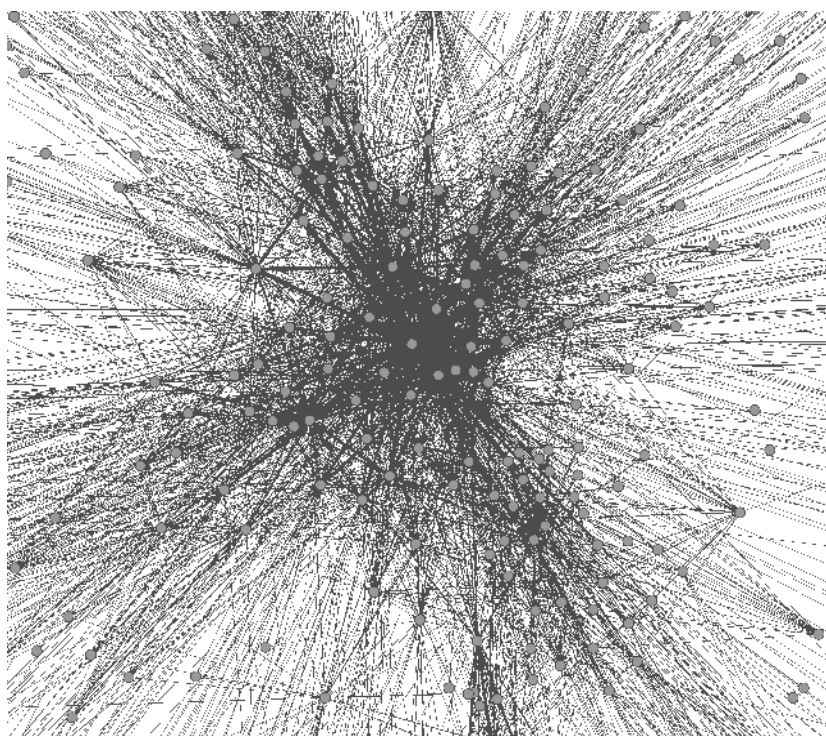


Fig. 40. Grafo con la aplicación del algoritmo 'Force Atlas'

Para poder diferenciar cada uno de los nodos y las conexiones que se tienen de aristas con nodos, se le personalizo con colores altos y el tamaño más grande en relación a los demás nodos, en lo que tiene que ver con los nodos más visitados y las aristas que mayor conexiones se las puso curvas y con diferentes colores según el nodo del cual se conecta, para lo cual dentro de Gephi se copió 'Degree' como parámetro de rango.

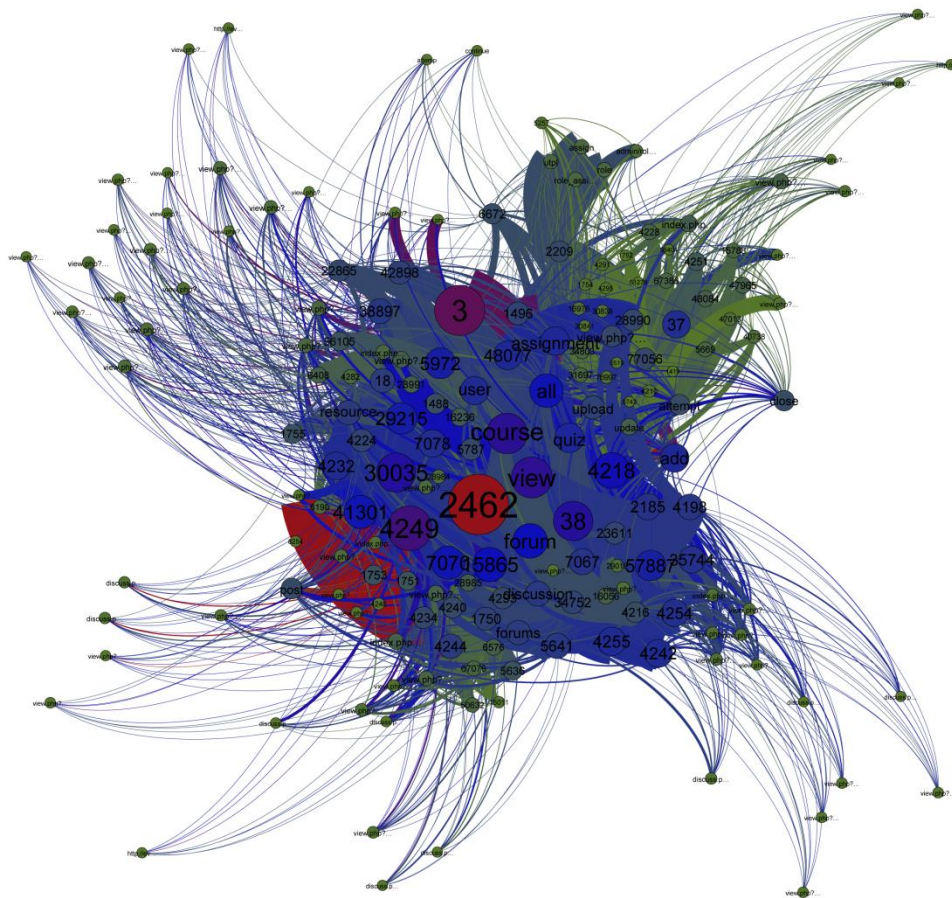


Fig. 41. Ponderada red de estudiantes que interactúan con la plataforma del EVA

El grafo mostrado en la **Fig. 41**, Muestra los colores de cada uno de los ciclos y los accesos por parte de los estudiantes a cada una de las herramientas dentro de la plataforma del EVA. Los colores más fuertes representan los nodos con mayor frecuencia de ingreso a las diferentes páginas dentro de los cursos de noveno ciclo; los nodos con mayor tamaño representan los nodos (estudiantes, materias, herramientas, páginas, acciones, etc.) con mayor número de accesos dentro de la plataforma y los nodos más pequeños son los nodos con menor número de frecuencias por parte de los estudiantes dentro del EVA.

PAPER

Análisis de patrones de navegación para determinar el comportamiento en los estudiantes de la UTPL que hacen uso del EVA, mediante la aplicación de minería de datos.

Ing. Priscila Valdiviezo Díaz
*Instituto de Ciencias de la
Computación*
Universidad Técnica
Particular de Loja
Loja, Ecuador
pmvaldiviezo@utpl.edu.ec

Ing. Héctor F. Gómez
*Instituto de Ciencias de la
Computación*
Universidad Técnica
Particular de Loja
Loja, Ecuador
hfgomez@utpl.edu.ec

Fanny Ruiz Chávez
*Escuela de Ciencias de la
Computación*
Universidad Técnica
Particular de Loja
Loja, Ecuador
feruiz@utpl.edu.ec

Resumen—En el presente trabajo se pretende analizar y aplicar varias técnicas y métodos de minería de datos con el fin de determinar patrones de comportamiento en la navegación de los estudiantes durante su interacción con el EVA. Actualmente existen varias investigaciones sobre la aplicación específica de técnicas de minería de datos basados en sistemas e-Learning.

Para el presente proyecto se pretende hacer un estudio de diferentes algoritmos que demuestran el uso de técnicas como: Reglas de Asociación (A-priori, GSP), Clasificación (Árboles de decisión, J-48) Clustering (SimpleKMeans, EM).

Palabras clave: Patrones de navegación, minería de datos, e- Learning, EVA, Moodle, reglas de asociación, secuencias de patrones, clustering, algoritmos, técnicas de minería.

Introducción

El aprendizaje en entornos virtuales hoy en día, demanda gran cantidad de tiempo y recursos de parte de los estudiante, ya que no existe una comunicación directa entre el profesor y estudiantes, y es por eso que se han incrementado herramientas que dispongan de varias alternativas para el aprendizaje y comunicación del estudiante con el tutor, algunas de las herramientas que disponen de varios medios de interacción entre tutor-estudiante es la herramienta Moodle (EVA), la cual tiene implementada la Universidad Técnica Particular de Loja, la misma que dispone de actividades como: foros, chat, mensajería, twitter, cuestionarios, exámenes entre otras; y permiten la comunicación constante entre los participantes del EVA.

Con esta investigación se pretende buscar y consolidar patrones de navegación interesantes en la interacción de los estudiantes con el EVA; para lo cual se pretende realizar un análisis en los modelos de acceso de los estudiantes a la información que se obtiene de los ficheros logs. Estos ficheros, almacenan las actividades que se

producen dentro de tablas en la base de datos del EVA como consecuencia directa de la navegación de estudiantes, y a estos datos aplicar técnicas de minería de datos, para encontrar el conocimiento sobre la forma en que los estudiantes navegan a través del sistema.

Existen varias técnicas de minería de datos, pero este trabajo se centrara en la minería de utilización Web, con herramientas de e-learning; ya que su objetivo es guiar a los estudiantes durante su aprendizaje para maximizarlo; y concretamente se desarrollará con el descubrimiento de reglas de asociación en los registros Log, y análisis de secuencias ya que son las técnicas más utilizadas en sistemas de e-learning, para lograr determinar ciertos comportamientos en los estudiantes, y de esta manera determinar la herramienta o técnica con la que se va a trabajar para conseguir estos patrones en la iteración con el EVA; aunque también se describen algunas otras técnicas diferentes como minería de contenidos y minería de texto, detección de anomalías, etc.

Entre los algoritmos mejor conocidos dentro de e-Learning están las reglas de asociación, las cuales permiten encontrar dependencias entre elementos, muchas veces estas reglas se usan como punto de comienzo cuando no se sabe exactamente qué tipos de patrones son los que se desea buscar, también se utilizan estas técnicas de minería de reglas de asociación y filtrado colaborativo para descubrir patrones de navegación útiles, también existen otros algoritmos como los de agrupamiento (clustering) como los vecinos más cercanos o k-means que encuentran relaciones entre grupos de objetos con ciertas afinidades, los algoritmos de árboles de clasificación permiten encontrar estructuras en espacios de cierta dimensionalidad.

El presente trabajo consta de cuatro partes las cuales se detallan a continuación: los trabajos relacionados al presente estudio, Estado del arte, Análisis de la plataforma de la UTPL, y el Estudio de las técnicas y herramientas, y el proceso de minería para el análisis de interacción en el EVA.

Trabajos Relacionados

Según [Romero, Ventura, de Castro & García, 2005] la minería de datos aplicada a SHA y STI para educación basados en Web, dispone de mayor información en los logs de la interacción entre el estudiante y el sistema sobre actividades realizadas, aciertos y fallos, notas y niveles de conocimiento, etc. Además la capacidad de estos sistemas para la enseñanza adaptada, hacen que la aplicación de minería de datos se diferencie más de la aplicada en comercio electrónico, al estar mucho más orientados al problema del proceso de aprendizaje.

Reglas de asociación

Uno de los pioneros de la utilización de técnicas de minería Web en sistemas de e-learning es **Osmar Zaïne** que propone utilizar agentes recomendados para recomendar actividades de aprendizaje en línea o atajos en un curso Web basándose en los historiales de acceso y mejorar el proceso de aprendizaje en línea. Concretamente utiliza minería de reglas de asociación para entrenar al agente recomendador y construir un modelo que representa el comportamiento de acceso o asociaciones entre actividades de aprendizaje en línea.

La utilización de métodos heurístico como regresión lineal en combinación con reglas de asociación es propuesta por **Freyberger** y otros para buscar buenos modelos de transferencia de aprendizaje de estudiantes a partir de los ficheros Logs de interacción de sistemas tutores inteligentes. El modelo de transferencia relaciona preguntas con los componentes del conocimiento necesarios para responderlas correctamente.

Clustering

Otra propuesta es la de análisis de caminos es el realizado por **Sung Ho Ha** y otros que utilizan el análisis de caminos transversales de páginas Web para la personalización de la enseñanza y la asociación de páginas Web para la estructuración de conocimiento. El descubrimiento de caminos se realiza desde dos puntos de vista: caminos individuales y caminos agregados que incluye el proceso de clustering utilizando mapas auto organizativo de **Kohonen**.

Análisis de secuencias

El análisis de patrones de navegación en entornos de aprendizaje basado en Web es utilizado por **Karin Becker** y otros dentro de una herramienta de minería de utilización Web para el análisis de patrones y pre procesamiento de datos de

utilización de entornos de aprendizaje basados en Web. Las técnicas de descubrimiento de patrones utilizadas son asociación y secuencia, y la secuencia de patrones describen accesos a páginas relaciones en un orden específico. Otra propuesta de análisis de caminos es el realizado por **Sung Ho Ha** y otros que utilizan el análisis de caminos transversales de páginas Web para la personalización de la enseñanza y la asociación de páginas Web para la estructuración de conocimiento.

TÉCNICAS DE MINERÍA WEB

Antes de aplicar cualquier técnica de minería de datos es conveniente realizar una transformación de los datos para que estos puedan ser manejados eficientemente; a continuación se detallan las técnicas de minería de datos permitan detectar patrones.

Las reglas de asociación

Las reglas de asociación permiten determinar patrones en los conjuntos de datos en los que ocurren transacciones de datos. Con esta técnica se puede encontrar relaciones sin que exista intervención alguna por parte de algún operador. El aprendizaje de reglas de asociación se divide en dos fases: [Fuentes, 2007].

1. Extracción de los conjuntos de ítems que cumplen con la cobertura requerida a partir de los datos.
2. Generación de las reglas a partir de estos documentos.

Las reglas de asociación descubren relaciones entre atributos dentro de un conjunto de datos que superan unos determinados umbrales. Su aplicación a sistemas de e-Learning permite descubrir relaciones o asociaciones entre distintas páginas Web visitadas. [Romero, Ventura, de Castro & García, 2005].

El análisis de secuencias

El análisis de secuencias de patrones es una técnica de minería de datos que descubre sucesiones dentro de un conjunto de datos. Su aplicación a sistemas de e-Learning permite analizar secuencias de páginas visitadas durante distintas sesiones de un mismo usuario. [Romero, Ventura, de Castro & García, 2005].

Árboles de decisión

Los árboles de decisión se asemejan a la teoría de divide y vencerás para partir el problema en

subconjuntos. Los árboles de clasificación, son un conjunto de condiciones organizadas jerárquicamente en forma de árbol.

Los árboles son uno de los métodos de aprendizaje inductivo supervisado no paramétrico bastante utilizados, son sencillos y útiles para encontrar estructuras en espacios de alta dimensionalidad y además pueden ser utilizados en diferentes áreas.

ANÁLISIS DE LA PLATAFORMA DEL EVA DE LA UTPL

El EVA está diseñado bajo la plataforma implementada en la UTPL, en la modalidad Presencial y a Distancia, El EVA cuenta con varias herramientas de apoyo, colaboración para disposición del profesor y estudiantes.

• **Tablas de información relacionadas al estudiante**

En esta sección se detalla las tablas que se relacionan entre sí para poder crear un escenario de la navegación del estudiante cuando este interactúa con entorno virtual, con páginas que se encuentran enlazadas entre objetos o recursos relacionados, y a partir de ello poder descubrir la secuencia de navegación que tiene un determinado modelos de estudiante en las acciones que en el EVA.

• **Diagrama Entidad – Relación**

En el proceso de análisis se representan las actividades que el estudiante efectúa cuando interactúa con el EVA, y las paginas a las que ingresa cuando navega por las herramientas y actividades del sistema.

Para determinar las tablas necesarias de Moodle y acoplarlas en el escenario, se empezará con la realización del modelo Entidad – Relación, el cual detalla un subconjunto de tablas y atributos de la base de datos, que se correlacionan entre sí, cuando el usuario realiza peticiones en el EVA, y toda esta información se almacena en la entidad de los Logs, dentro de la base de datos de la plataforma del EVA, y es de allí donde se filtrara la información necesaria para el proceso de minería.

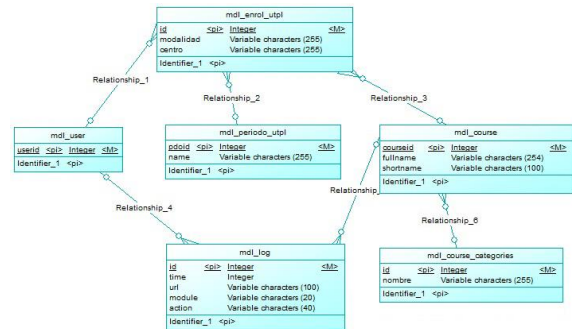


Fig. 2 Diagrama E–R de las tablas del EVA.
Fuente: Ruiz, F., (2014).

La Fig. 2 detalla las tablas para el proceso de minería de datos, estas tablas contienen los registros de las actividades que realizan los estudiantes en su navegación por el sistema, para obtener un seguimiento detallado de las acciones que los estudiantes han realizado, accesos a los cursos, tareas, foros, etc., que ayuden a la identificación de patrones de navegación.

• **Tablas del EVA para la minería de datos**

Fuente: Ruiz, F., (2014).

Tabla	Descripción
mdl_user	Información de los estudiantes.
mdl_role	Entidades que tienen acceso al EVA.
mdl_log	Registros de las acciones de los usuarios.
mdl_course	La información de los cursos en el sistema.
mdl_centro	Centro universitario al cual pertenece.
mdl_periodo_utpl	Periodo académico que cursa el estudiante.
mdl_modules	Información de las herramientas en el EVA.
mdl_log_display	Información de las acciones de los usuarios.

Tabla 1. Tablas del EVA previas al proceso de minería.

• **Descripción de los atributos de interacción en el EVA**

Para la obtención de los datos de interacción en el EVA, hay que extraer los datos de interés de los estudiantes durante su navegación en el sistema, y las páginas a las que accede con mayor frecuencia para realizar actividades dentro de las herramientas en los cursos.

ANÁLISIS Y SELECCIÓN DE HERRAMIENTAS

Selección de datos para el análisis

En esta actividad, dentro del análisis de los archivos Logs, se tomarán un conjunto determinado de datos de la base de datos, las mismas que son algunos atributos dentro de cada una de las tablas que forman parte de la interacción del estudiante de la UTPL cuando interactúa con el EVA. Estos datos se tomarán como datos de entrada que nos ayudara a

determinar la forma de navegación de los estudiantes dentro del Entorno Virtual, para de esta manera poder identificar las características, propiedades, preferencias e intereses de cada uno de los estudiantes; y a partir de ello crear un nuevo modelo de usuario.

- **Datos de usuario**

Se determinan los datos que necesitaremos de la base de datos para realizar el análisis respectivo de la navegación del estudiante por el EVA. Estos atributos nos permiten definir los parámetros de acceso del usuario al entorno de aprendizaje y de esta manera poder descubrir los patrones de navegación durante su interacción con el EVA de la UTPL, detallados a continuación:

- Id de usuario.
- Centro universitario.
- Ciclo en el que esta matriculado
- Modalidad de estudio a la cual pertenece.
- Periodo académico.
- Materias en las que esta enrolado en el sistema.

- **Datos de interacción**

Los datos que se detallan en esta sección son las interacciones que el usuario realiza cuando solicita algún recurso de la plataforma del EVA, el mismo que definen los niveles de utilidad de los estudiantes de la UTPL en los diferentes contenidos de las herramientas en cada una de las asignaturas que el usuario esté cursando en el periodo de estudio, y además la frecuencia de interacción del estudiante con el sistema, este último es uno de los atributos que más nos ayudan a definir el nivel de importancia para el aprendizaje del usuario con respecto al EVA, ya que este factor nos permite identificar la frecuencia de peticiones y acciones que el usuario (sujeto) realiza sobre los recursos (objetos) de la plataforma de soporte a la educación (EVA), los datos de interacción que realizan los estudiantes con el sistema se detallan a continuación:

- Fecha de acceso al sistema.
- Herramientas a las que ingresa en el EVA.
- Actividades a las que ingresa.
- Acciones que realiza el estudiante en el EVA.

- Información de las páginas visitadas por el estudiante cuando interactúa en el EVA.
- Páginas a las que acceden con mayor frecuencia.

Estos datos se almacenan automáticamente dentro de los archivos Logs, en la entidad llamada *mdl_log*, dentro de la base de datos del EVA, y son algunos de los datos que nos servirán para la inferencia de las rutas de navegación de los estudiantes que navegan en el EVA.

De acuerdo a los objetivos planteados para la presente tesis, en los cuales principalmente esta encontrar las rutas de navegación que los usuarios usualmente tienen cuando interactúan con las herramientas disponibles en la plataforma de la Universidad (EVA), mediante técnicas básicas de aprendizaje automático, es decir la minería de datos, tratar de extraer el aprendizaje del estudiante con un mejor detalle con los algoritmos de aprendizaje propuestos anteriormente para el presente análisis.

Para descubrir patrones de navegación una de las técnicas más eficaces en este tipo de análisis es la aplicación de la Identificación de Patrones Secuenciales en el proceso de extracción de los datos de navegación. Este proceso de patrones secuenciales permitirá aplicar técnicas de reglas de asociación secuenciales, reglas que expresen patrones de comportamiento secuencial en la navegación de los estudiantes que interactúan con el EVA, en distintos instantes del tiempo y de esta manera clasificar a los estudiantes que visitan las diferentes páginas del sistema en grupos con características similares según las rutas de navegación que tengan, es decir que permita agrupar estudiantes con una conducta similar y que tengan intereses similares al momento de acceder al sistema

Algoritmos seleccionados para la experimentación

Los datos tomados para el desarrollo de este trabajo, corresponden a la base de datos de la UTPL, los cuales son datos reales tomados del periodo académico *abril 2013 – agosto 2013*, de los estudiantes de noveno ciclo, Modalidad a Distancia de la carrera de Informática.

Fuente: Ruiz, F., (2014).

Herramienta de Minería	Técnicas	Algoritmos
Weka	Asociación	GSP A priori
	Clustering	SimpleKmeans EM
	Clasificación	J48 (Árboles)

Tabla 2. Técnicas para la minería de datos del EVA.

La razón por la cual se realizará el análisis con diferentes técnicas de minería, es porque estos algoritmos generan modelos descriptivos que agrupan a los estudiantes en función de sus acciones u alguna otra variable de navegación, en este caso se lo realiza por herramientas y páginas visitadas dentro del sistema; y las Reglas de asociación que permiten relacionar estudiantes según sus características y similitudes. A partir de estas dos técnicas de minería se realizará el estudio de búsqueda de patrones mediante un análisis del comportamiento de navegación de los estudiantes a través de las acciones registradas en los archivos Log, con la ayuda potencial descriptiva de los modelos de Weka: Clustering, Asociación y Clasificación.

MINERÍA DE DATOS: EXPERIMENTACIÓN Y RESULTADOS

Para descubrir patrones de navegación en los datos de interacción de los estudiantes, una técnica muy eficaz en este tipo de análisis es la aplicación de la Identificación de Patrones Secuenciales en el proceso de extracción de los datos de navegación. Este proceso de patrones secuenciales permitirá aplicar técnicas de reglas de asociación, reglas que expresen patrones de comportamiento secuencial en la navegación de los estudiantes que interactúan con el EVA, en distintos instantes de tiempo y de esta manera clasificar a los estudiantes que visitan las diferentes páginas del sistema en grupos con características similares según las rutas de navegación que tengan, es decir que permita agrupar estudiantes con una conducta e intereses similares en el sistema.

Clasificación de páginas del EVA

Para determinar las rutas de navegación de los estudiantes en el EVA, se consideró conveniente clasificar tres tipos de dominios, en las cuales se encuentran las herramientas disponibles en el sistema para la interacción de los estudiantes con su profesor, detalladas a continuación:

Fuente: Ruiz, F., (2014).

CATEGORÍA	SERVICIOS	ACCIONES
Información General	Perfil de usuario Calendario Library Glossary	view, view all, update add, delete, edit. mailer. view, view all, add comment
Páginas de Actividades	Foros Cuestionarios Tareas	view forum, add discussion. view, attempt, close attempt view, view all, upload.
Páginas de Apoyo	Recursos Subir archivos Blogs utpl	view, view all. upload view, add, comment, delete. role_assign
Comunicación y colaboración	Twitter Chat Mensajería	view all view, talk, update, add contact. contact, unblock contact, write

Tabla 3. Clasificación de páginas en el EVA.

En la **Tabla 3**, se detalla cada una de las cuatro categorías que se clasificaron para determinar las páginas por las cuales navega el estudiante cuando interactúan con las herramientas del EVA, y de esta manera determinar cuales son las preferencias de los estudiantes según las acciones que realice dentro de las páginas por las que navega y de esta manera encontrar las rutas de navegación de los estudiantes por el sistema para su aprendizaje e interacción y comunicación con los demás estudiantes y profesor.

Luego de obtener los datos necesarios para el pre-proceso de minería, para cumplir con los objetivos propuestos en la presente tesis y las herramientas que nos permita realizar dicho proceso, el cual se detalla más adelante con cada una de las técnicas y algoritmos propuestos para obtener la información necesaria que nos ayuden a determinar los indicadores y presentar un pequeño modelo con los indicadores obtenidos en el minado de los datos.

Descripción del Escenario

Las materias de noveno ciclo seleccionadas para el experimento cuenta con un total de 132 usuarios, de los cuales están estudiantes y docentes de Noveno ciclo de la carrera de Informática de la modalidad a Distancia de la Universidad Técnica Particular de Loja (UTPL); de los cuales 84 son estudiantes que corresponden al noveno ciclo de Informática, 6 son docentes de los seis cursos, y los 28 restantes son estudiantes que se encuentran enrolados en más de un curso dentro de las materias de noveno ciclo. De los 116 estudiantes, 112 estudiantes interactuaron con las herramientas dentro de la plataforma (*foros, tareas, recursos, cuestionarios, twitter, chat, perfil de usuario, fuentes de información, etc.*) en el sistema. Estos datos fueron tomados del periodo académico *Abril 2013 – Agosto 2013*.

En los datos extraídos en lenguaje SQL, se determino que en su mayoría los estudiantes tuvieron acceso a los cursos de Noveno ciclo, y realizaron las siguientes actividades:

- Interactuaron en las herramientas dentro de los cursos, como son: foros, cuestionarios, tareas y los recursos subidos por los tutores en los cursos.
- En los cinco cursos la mayor parte de los estudiantes interactúan con las mismas herramientas: cursos, foros, cuestionarios, recursos, tareas, perfil de usuario, utpl, subida de archivos, y sólo uno de los cinco cursos, interactuó con el blog y el twitter.
- De un total de 116 estudiantes enrolados en los diferentes cursos, 112 estudiantes interactúan con las herramientas del EVA.
- En el proceso de interacción de los estudiantes con las herramientas del sistema, y durante la navegación por el EVA, los estudiantes acceden a revisar con mayor frecuencia los recursos y actividades subidos al entorno por el tutor responsable de cada uno de los cursos, en el EVA.
- Durante el ingreso a los cursos, los estudiantes revisan los recursos, tareas y foros, antes de realizar alguna acción sobre el mismo.
- Cuando tienen que responder el cuestionario los estudiantes realizan más de un intento para el desarrollo de los mismos.
- Casi todos los estudiantes participaron de los foros y cuestionarios propuestos en el curso por cada uno de los tutores responsables.
- La mayor parte de los estudiantes participan en las discusiones dentro de los foros propuestos por el profesor de los cursos.

Casi en su totalidad, los estudiantes interactúan sólo con aquellas herramientas que los tutores/profesores consideran para el aprendizaje de los estudiantes dentro de los cursos.

Obtención de datos de navegación a partir de los datos del usuario

Los datos de navegación de los estudiantes en la plataforma se detallan a continuación:

Fuente: Ruiz, F., (2014).

DATOS DE INTERACCIÓN DE LOS ESTUDIANTES
Páginas en las que el estudiante inicia sesión de usuario.
Páginas visitadas por el estudiante en los cursos.
Páginas visitadas por el estudiante en el perfil del usuario.
Páginas en las que el estudiante realiza vistas de los recursos.
Páginas que el estudiante ingresa a revisar los recursos.
Páginas donde los estudiantes han visitado los foros.
Páginas en las que el estudiante se ha suscripto en los foros.
Páginas en las que el estudiante ha eliminado la suscripción.
Páginas que el estudiante inicia una conversación en los foros.
Páginas que el estudiante responde en debates de los foros
Páginas donde el estudiante agrega un debate en los foros.
Páginas que el estudiante ha borrado un debate en el foro.
Páginas que el estudiante ha adicionado una nota en los foros
Páginas en las que el estudiante actualiza notas en los foros
Páginas donde el estudiante postea un mensaje en el twitter.
Páginas donde el estudiante da respuestas a los foros.
Páginas en las que el estudiante revisa las tareas.
Páginas en las que el estudiante sube las tareas en el EVA.
Páginas que el estudiante actualiza las tareas.
Páginas en las que el estudiante ingresa al chat en el curso.
Páginas en las que el estudiante revisa los cuestionarios.
Páginas que el estudiante realiza un intento en el cuestionario
Páginas que el estudiante cierra el cuestionario que realizó.
Páginas que el estudiante ingresa a la mensajería del EVA.
Páginas que el estudiante ingresa al blog dentro del sistema.
Páginas en las que el estudiante ingresa a revisar los blogs.
Páginas que el estudiante ingresa a la información del EVA.

Tabla 4. Datos de navegación de los estudiantes.

Los datos de interacción de los estudiantes, se obtuvieron mediante consultas SQL para determinar las acciones realizadas por el estudiante cuando interactúa en el EVA.

Selección de las técnicas de minería

Para la obtención de resultados de la minería aplicada a los datos de entrada y previa al diseño del modelo de datos con los indicadores encontrados después de la minería, se consideró conveniente desarrollar según las características mencionadas anteriormente de los algoritmos escogidos para el proceso de minado (*Asociación, Clustering y Clasificación*), se los creara con los datos extraídos de las interacciones de los estudiantes dentro de las herramientas y servicios a los cuales accede con mayor frecuencia cuando navega en la plataforma del EVA, y a estos datos se les aplicara varias de las técnicas de aprendizaje de minería, descritas en la **Tabla 2.**, para luego realizar una tabla comparativa con el resultado de los algoritmos propuestos.

Experimentación con la técnica de Asociación

Con la finalidad de encontrar patrones de comportamiento de los estudiantes frente al sistema cuando interactúan con el mismo, se creyó necesario antes de realizar la minería buscar algoritmos que permitan implementar los indicadores seleccionados, como ya se lo considero en el estado del arte, que se trabajaría con la

herramienta de WEKA (cual cuenta con un amplio conjunto de algoritmos que sirven para la generación de indicadores), la técnica empleada será la de Asociación y los algoritmos seleccionados fueron el GSP y A priori, los mismos que permite encontrar patrones en los datos ingresados.

- **Aplicación del algoritmo GSP**

Para el análisis de los datos dentro del algoritmo GSP en Weka, se lo realizará como primer experimento con los valores que tiene por defecto el modelo, (valor de semilla $S=0.9$, longitud de las secuencias $l=0$ y filtros dentro de los atributos $F=-1$).

Se realizará tres experimentos con diferentes datos, en este caso se cambiara el valor del número de las secuencias, con la finalidad de encontrar las secuencias más frecuentes en la navegación de los estudiantes en las herramientas y servicios de la plataforma del EVA, el segundo experimento será con los valores de ($S=0.9$, $l=0$, $F=1$), y con el tercer experimento será con los valores de ($S=0.9$, $l=0$, $F=2$) como lo indica la tabla comparativa de resultados. La implementación de este algoritmo permite encontrar dentro de un conjunto de entrenamiento encontrar las secuencias más frecuentes de los estudiantes durante su navegación en la plataforma, y a partir de estos resultados obtener indicadores navegación en el EVA.

```

Associator output
-----
GeneralizedSequentialPatterns
=====
Number of cycles performed: 4
Total number of frequent sequences: 15
Frequent Sequences Details (filtered):
- 1-sequences
[1] <{'(-inf-28987.5)'}> (10)
[2] <{assignment}> (10)
[3] <{view all}> (9)
[4] <{index.php?id=28984}> (9)
- 2-sequences
[1] <{'(-inf-28987.5)',assignment}> (10)
[2] <{'(-inf-28987.5)',view all}> (9)
[3] <{assignment,view all}> (9)
[4] <{'(-inf-28987.5)',index.php?id=28984}> (9)
[5] <{assignment,index.php?id=28984}> (9)
[6] <{view all,index.php?id=28984}> (9)
- 3-sequences
[1] <{'(-inf-28987.5)',assignment,view all}> (9)
[2] <{'(-inf-28987.5)',assignment,index.php?id=28984}> (9)
[3] <{'(-inf-28987.5)',view all,index.php?id=28984}> (9)
[4] <{assignment,view all,index.php?id=28984}> (9)
- 4-sequences
[1] <{'(-inf-28987.5)',assignment,view all,index.php?id=28984}> (9)

```

Fig. 3. Resultados con el Algoritmo GSP.
Fuente: Ruiz, F., (2014).

La Fig. 3 muestra una lista ordenada de secuencias (*item-sets*) realizada por los estudiantes cuando navegan por el sistema e interactúan con las herramientas.

Resultados obtenidos de la experiencia

En la Fig. 3, se puede apreciar los resultados que se obtuvieron del tercer experimento con el algoritmo GSP, el cual presentan una lista de secuencias de navegación de los estudiantes en la plataforma. ($S=4$), la longitud de las secuencias se la determina con el número de elementos o ítem-sets que pertenecen a determinada secuencia, que en este caso es ($l=2$), y uno de los cuatro estados representan las acciones realizadas por los estudiantes en las páginas del EVA, es decir, un ítem set o elemento de la secuencia es una página Web a la cual accedió el estudiante, y una secuencia es el total de páginas Web visitadas por el estudiante cuando navega por el sistema en determinado tiempo.

Según los resultados de la aplicación del algoritmo **GSP**, los datos resultantes de la secuencia frecuente de acceso dentro de las rutas de navegación de los estudiantes en el EVA se detallan de la siguiente manera: En el segundo experimento realizado los resultados presentados en la Fig. 3 muestran que el curso con más interacciones por parte de los estudiantes es el curso de **Redes y Sistemas Distribuidos [A]**, con mayor frecuencia de accesos en la página de los **Foros y Tareas** que ingresaron a revisar las tareas y revisar los debates en los foros en el curso dentro del EVA.

- **Aplicación con el algoritmo A-priori**

Antes de empezar con el minado, es conveniente realizar un pre-proceso en los datos preliminar a la ejecución del algoritmo en WEKA, esta herramienta permite realizar varias manipulaciones sobre los datos; y lo primero que se realizó en ellos es la aplicación de filtros, un filtro no supervisado sobre los atributos de los datos, ya que las operaciones son independientes del análisis del algoritmo.

La función del filtro 'Discretize' sobre los datos es transformar los atributos numéricos seleccionados en atributos simbólicos, con una gran variedad de etiquetas que resultan de dividir la amplitud total del atributo en intervalos.

Se realizaron tres experimentos el primero con los valores por defecto del modelo ($N=10$, $C=0.9$, $M=0.1$ y $S=-1.0$).

En el segundo experimento se modifico el valor de confianza, el del límite superior de apoyo mínimo y el valor del número de las reglas que se desea obtener ($N=10$, $C=0.9$, $M= 0.0$ y $S=-$

1.0), los resultados fueron nulos ya que no se encontraron reglas en el conjuntos de elementos dentro de los datos, por lo cual, se dejó el mismo valor de confianza y los valores del límite inferior y superior de apoyo mínimo del modelo, así ($N=10$, $C=0.5$, $M=0.1$ y $S=-1.0$).

Y el último experimento, igual se dejó los valores por defecto del modelo y sólo se modificó el número de reglas a obtener ($N=50$, $C=0.9$, $M=0.1$ y $S=1.0$), el cual presenta resultados similares a los del primer experimento, pero con mayor número de reglas y algunos valores diferentes.

En el tercer experimento nos produce relaciones bastante evidentes, y con reglas igualmente significativas entre curso, actividad y la acción que realizan los estudiantes. Al ejecutar el algoritmo con los parámetros por defecto sólo nos presenta una regla; Esto se debe a que el modelo pone como umbral mínimo aceptable para una regla el 90%, si se pone un parámetro menor al del modelo de $C = 50\%$ y los demás valores con los mismos del modelo por defecto, y nos arroja el siguiente resultado.

```

15:36:19 - Apriori
==== Run Information ====
Scheme: www.associations.Apriori -0 10 -0 0 0.9 -0 0.05 -0 0.0 -0 0.0 -0 0.0 -0 1.0 -0 1
Instances: 13974
Attributes: 5
  USUARIO
  CURSO
  ACTIVIDAD
  ACCION
  PAGINA_INGRESADA
==== Associator Model (full training set) ====

Apriori
=====
Minimum support: 0.05 (695 instances)
Minimum metric 'confidence': 0.9
Number of cycles performed: 39
Generated sets of large itemsets:
Size of set of large itemsets L(1): 23
Size of set of large itemsets L(2): 46
Size of set of large itemsets L(3): 25
Size of set of large itemsets L(4): 4
Best rules found:
1. ACCION=view_forum 2406 ==> ACTIVIDAD=forum 2406 conf:(1)
2. ACCION=view_discussion 2231 ==> ACTIVIDAD=forum 2231 conf:(1)
3. USUARIO=[inf-7175-9] ACCION=view_forum 132 ==> ACTIVIDAD=forum 1321 conf:(1)
4. USUARIO=[inf-7175-9] ACCION=view_discussion 1293 ==> ACTIVIDAD=forum 1293 conf:(1)
5. CURSO=redes y sistemas distribuidos ACCION=view_forum 1219 ==> ACTIVIDAD=forum 1219 conf:(1)
6. PAGINA_INGRESADA=view.php?id=51755 1141 ==> CURSO=redes y sistemas distribuidos 1141 conf:(1)
7. PAGINA_INGRESADA=view.php?id=51755 1141 ==> ACTIVIDAD=course 1141 conf:(1)
8. PAGINA_INGRESADA=view.php?id=51755 1141 ==> ACCION=view 1141 conf:(1)
9. PAGINA_INGRESADA=course PAGINA_INGRESADA=view.php?id=51755 1141 ==> CURSO=redes y sistemas distribuidos 1141 conf:(1)
10. ACTIVIDAD=course PAGINA_INGRESADA=view.php?id=51755 1141 ==> CURSO=redes y sistemas distribuidos 1141 conf:(1)

```

Fig. 4 Resultados con el Algoritmo A priori.
Fuente: Ruiz, F., (2014).

Como se puede apreciar en la Fig. 9, el resultado del algoritmo presenta 10 reglas descubiertas. Según los resultados obtenidos de la implementación de A priori se puede concluir que el modelo tiene la ventaja de ser de fácil utilización para el usuario, ya que sólo requiere como parámetro inicial el número de reglas que se desee obtener. A su vez este algoritmo posee varias desventajas para este tipo de análisis, y es que en los resultados el modelo dispara gran cantidad de datos redundantes, es decir, reglas con la misma información en reglas con bajo soporte, además, el tiempo de ejecución es muy alto.

Con esta información el profesor puede tomar decisiones con respecto al manejo de las

herramientas y prestar más atención a las herramientas con las que interactúan más los estudiantes.

Resultados obtenidos de la experiencia

La tarea de asociación permite descubrir patrones en forma de reglas, los cuales muestran las interacciones que ocurren con mayor frecuencia en el conjunto de datos de entrada. Como se puede ver el algoritmo busca relaciones entre los atributos de los datos para descubrir reglas de asociación expresadas en el soporte y confianza del modelo. La Fig. 10 muestra que hay reglas poco interesantes (regla 5), reglas redundantes (la regla 3, que es igual a la regla 4 y 5), también hay reglas similares, (reglas con el mismo elemento en el antecedente y el consecuente, pero intercambiados, tales como las reglas 4, 9, y las reglas 1 y 2); también hay reglas que pueden ser muy importantes para los propósitos educativos, es decir, para saber las herramientas con la que a los estudiantes más les gusta trabajar, o son más aptas para su aprendizaje, como son las reglas que se ajustan o tienen mayor nivel de interacción y visitas (si un estudiante interactúa bastante con los foros, puede ser porque expresando sus conocimientos o leyendo el aporte de otros estudiantes aprende de mejor manera, como la regla 1, 2 y algunas reglas similares).

- **Aplicación con el algoritmo SimpleKMeans**

La razón por la cual se pusieron valores de semilla más altos es debido a que los datos de semilla con valores de (10, 50 y 100) presentan un valor muy alto tanto en el error de la suma de los cuadrados como en el valor de clústeres mal clasificados, por lo cual se realizaron experimentos con valores de semilla más altos para determinar si existe un decremento en el valor de los datos considerados en el análisis de experimentación.

```

Clusterer output
=====
Number of iterations: 6
Initial cluster sum of squared errors: 914.37492476012
Missing values globally replaced with mean/mode

Cluster members:
Attribute      Full Data      Cluster#
=====
0 3485 ( 45%)
1 2306 ( 29%)
2 2282 ( 30%)

Class attributes: REDACCIONES
Classes to clusters:
0 1 2 <-- assigned to cluster
11 145 197 assigned
215 0 0 blog
215 154 1 runtime
349 202 1710 forum
182 122 170 quiz
300 155 24 assignment
0 4 15 code
14 26 40 upload
150 443 180 user
0 4 15 utqi

Cluster 0 <-- course
Cluster 1 <-- quiz
Cluster 2 <-- forum
Incorrectly clustered instances: 3426.0 44.4762 9

```

Fig. 5. Resultados con SimpleKMeans con S=2000.
Fuente: Ruiz, F., (2014).

La **Fig. 5**, muestra los resultados obtenidos de la agrupación de los tres clústeres evaluados según la interacción de los estudiantes con las herramientas disponibles en los cursos minados en EVA.

Para el proceso de experimentación se realizó varios experimentos, del cual el primero se lo realizó con los parámetros por defecto del algoritmo ($I=100$, $N=-1$, $S=100$), y el modo de evaluación 'Use training set' que utiliza el propio conjunto de entrenamiento para que indique el porcentaje de instancias que se van a cada grupo.

Los resultados del experimento presentaron datos poco significativos para la evaluación por lo cual, se modificó los valores de semilla, el número de clúster y el modo de evaluación, el cual se lo realizó por Clases to clusters evaluation para que realice la comparación de los clústeres con varios atributos de clasificación de los datos.

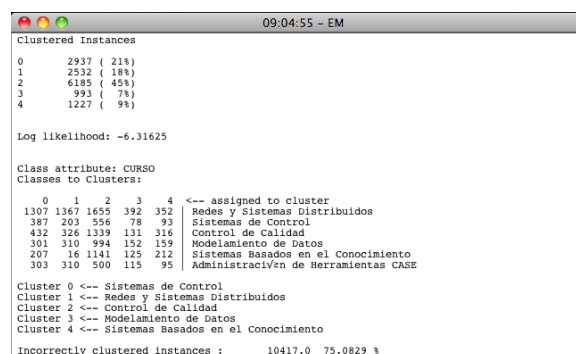


Fig. 6. Resultados con el algoritmo EM con $S=10$
Fuente: Ruiz, F., (2014).

Debido a la complejidad que tiene este algoritmo, y a que se requiere más operaciones, se realizó varios experimentos con el valor del clústeres, la **Fig. 10**, muestra que el algoritmo realizó la evaluación por cursos, en el cual se puede apreciar que el curso con mayor nivel de interacción es '**Redes y Sistemas Distribuidos**' con un total de 1843 datos bien clasificados y seguido con 1163 la materia de '**Control de Calidad**'.

De los tres grupos obtenidos el grupo con mayor número de accesos realizados dentro del sistema por parte de los estudiantes de la UTPL de Informática, de la Modalidad Abierta, de un total de 112 estudiantes, navegan con mayor frecuencia en la herramienta de los foros en las materias de *Redes y Sistemas Distribuidos* y *Control de Calidad* con un total bien diferenciado en comparación a las demás herramientas, en las cuales ingresan a *revisar los foros* propuestos por el docente a revisar las discusiones.

Con los resultados obtenidos del modelo **SimpleKmeans** determinó mayor eficiencia en los

resultados obtenidos y en un intervalo de tiempo más corto al momento de ejecutarlos, además el modelo permitió obtener el indicador del nivel de interés y colaboración que tienen los estudiantes con las diferentes herramientas de la plataforma para su aprendizaje; en los resultados obtenidos de la implementación del algoritmo EM se agrupó a los estudiantes de los cursos en varios grupos según las características de comportamiento que tienen con las actividades realizadas en las herramientas del EVA, y a partir de este comportamiento análogo que tienen los estudiantes determinar patrones de navegación durante la interacción con el sistema. Con los datos obtenidos del proceso del modelo se determinaron indicadores estadísticos del nivel de interacción de los estudiantes con las páginas de Actividades (*foros, cuestionarios, tareas*) y las páginas de Apoyo (*recursos, subida de archivos y utp*), además se determinó el nivel de colaboración que tiene los estudiantes ya que interactúan con las herramientas de colaboración (*foros, twitter, blog*), con los foros tienen un nivel de interacción alto, pero con los blog y el twitter existe un porcentaje bajo de contribución, por lo tanto, se puede decir que mediante la aplicación de la técnica de clustering se obtuvieron indicadores de interacción y colaboración que tienen los estudiantes con las herramientas.

Experimentación de la técnica Clasificación con Árboles de decisión

La aplicación de esta técnica permite encontrar a partir de reglas, las rutas de navegación de los estudiantes cuando interactúan con las herramientas de la plataforma; y para ello se utilizará la técnica de **árboles de decisión** ya que son de gran ayuda para el minado, este es un modelo de predicción muy utilizado en el ámbito de la inteligencia artificial, a partir de una base de datos se construyen estos diagramas de construcciones lógicas, muy similares a los sistemas de predicción basados en reglas, que sirven para representar y categorizar una serie de condiciones que suceden de forma sucesiva, para la resolución de un problema. El algoritmo que se usara para el análisis con esta técnica es el Algoritmo *J-48*, con el cual se realizará 3 experimentaciones.

- **Aplicación con el Algoritmo J-48**

El análisis con el algoritmo *J-48* se lo realizó con diferentes valores en el factor de confianza C , utilizado para la poda del árbol; la semilla es usada para la aleatorización de los datos; el valor de M , número mínimo de instancias por hoja; y la función de evaluación, que usa un determinado atributo para la evaluación de los resultados.

Los tres experimentos se realizaron con valores de factor de confianza pequeños, ya que son más eficientes en los resultados. Los datos que variaran en los tres experimentos serán el factor de confianza y la función de evaluación.

El primer experimento se lo realizó con el factor de confianza de $C=0.50$, que presenta el nivel de error más bajo, y los demás valores son los valores por defecto del modelo, y la función de evaluación se la configuró para clasificar los datos por *cursos*. El segundo experimento se lo realizó con un valor de confianza más bajo $C=0.25$, y la función de evaluación en los datos, por *herramientas*. El tercer experimento se incremento el factor de confianza en $C=1.0$, ya que tiene el mínimo valor de error en los resultados, y la función de evaluación en las páginas más visitadas.

Fuente: Ruiz, F., (2014).

Algoritmo J48	(C = 0.25)
Función a evaluar	cursos
Número total de instancias	6
Modo de evaluación	Cross-validation 10
Número de hojas	1002
Tamaño del Árbol	1003
Tiempo	0.36 seg.
Instancias bien clasificadas	13618
Porcentaje	98.15%
Instancias mal clasificadas	256
Porcentaje	2.74%
Error estadístico	0.98%
Error relativo absoluto	2.74%

Tabla 8. Análisis de resultados con el algoritmo J-48.

Se puede apreciar en la **Tabla 8**, que los resultados más eficientes son los del primer experimento, $C=0.50$, $M=2$ y la función de evaluación que se lo configuro el atributo *cursos* para la evaluación de los resultados.

```

Number of Leaves : 85
Size of the tree : 149
Time taken to build model: 0.31 seconds
=== Evaluation on training set ===
=== Summary ===
Correctly Classified Instances 10737      77.884 %
Incorrectly Classified Instances 3137      22.6106 %
Kappa statistic 0.7047
Mean absolute error 0.0967
Root mean squared error 0.2198
Relative absolute error 37.1549 %
Root relative squared error 65.9557 %
Total Number of Instances 13874

=== Detailed Accuracy By Class ===
   TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
   -----  -
   0.988    0.14    0.798    0.898    0.871    0.96    Redes y Sistemas Distribuidos
   0.736    0     1     0.736    0.848    0.989    Sistemas de Control
   0.981    0.02    0.87    0.896    0.886    0.989    Control de Calidad
   0.908    0.044    0.787    0.908    0.832    0.986    Modelamiento de Datos
   0.984    0.025    0.787    0.884    0.847    0.989    Sistemas Basados en el Conocimiento
   0.952    0.067    0.483    0.952    0.503    0.929    Administrativn de herramientas CASE
Weighted Avg. 0.774    0.07    0.789    0.774    0.769    0.962

=== Confusion Matrix ===
  a  b  c  d  e  f  <-- classified as
4862  0  0  0  9  202  a = Redes y sistemas Distribuidos
285  965  0  0  0  83  b = Sistemas de Control
267  0  1477  159  183  456  c = Control de Calidad
0  0  120  1739  114  43  d = Modelamiento de Datos
146  0  201  311  960  81  e = Sistemas Basados en el Conocimiento
336  0  0  57  0  730  f = Administrativn de herramientas CASE
  
```

Fig. 7. Resultados con el algoritmo J-48 con $C=0.50$

Fuente: Ruiz, F., (2014).

En el primer caso se pudo observar durante la experimentación cualquiera que sea el valor del factor de confianza y la semilla, igual presentaba el mismo resultado en todos los casos.

En el segundo caso se dejo el factor de confianza el valor por defecto del modelo ya que este presento los resultados más óptimos, ya que al incrementar el valor de $C=0.25$, también incrementaba el valor de error y las instancias mal clasificadas. En el tercer caso se pudo observar que el resultado más eficiente presento el valor de confianza más alto, ya que a medida de que se iba incrementando el factor de confianza más grande era el árbol y el número de hojas, pero sin embargo disminuía el total de instancias mal clasificadas y por ende el porcentaje de error.

Otra forma de variar el tamaño del árbol es el parámetro que especifica el mínimo número de instancias por nodo, si bien es menos distinguido puesto que depende del número de instancias en el conjunto de partida.

Análisis e interpretación de resultados

Los resultados obtenidos de la implementación del Algoritmo J48, los cuales se pueden apreciar en la **Fig. 13**, del primer experimento que fue el que arrojó los mejores resultados, es que el algoritmo los clasificó por cursos, y a partir de ellos se desprenden las rutas de navegación del estudiante según las peticiones realizadas a las páginas del EVA en cada una de las herramientas a las que frecuentan los estudiantes.

En los resultados se puede ver que la mayor parte de los estudiantes, interactúan solo con algunas de las herramientas dentro del sistema, y a continuación se interpretan los resultados obtenidos de la matriz de confusión.

a = Redes y Sistemas Distribuidos: Según el algoritmo éste es el curso con mayor índice de ingresos e interacción por parte de los estudiantes. Los valores en la diagonal son los aciertos, y el resto de valores son los errores, por lo tanto de los 13874 interacciones realizadas por los estudiantes en el EVA, 5073 accesos son correctamente clasificados y 0 errores.

b = Sistemas de Control: Éste es el segundo curso con mayor índice de ingresos por parte de los estudiantes, con un total de 1317 accesos, 1285 están correctamente clasificados y 32 son errores.

c = Control de Calidad: Con un total de 2544 ingresos, 2443 son accesos a las actividades de

dicho curso correctamente clasificadas, y 101 son errores.

d = Modelamiento de Datos: El total de accesos por parte de los estudiantes es de 1916, de los cuales 1890 son correctamente clasificados, y 26 son errores. Este clase es la que tiene el más bajo índice de errores en la clasificación.

e = Modelamiento de Datos: En éste curso hay un total de 1701 datos clasificados, de los cuales 1643 son instancias bien clasificadas por el algoritmo, mientras que 58 son errores.

f = Modelamiento de Datos: Por último está éste curso, con el menor índice de interacciones por parte de los estudiantes. Con un total de 1323 datos clasificados, 1284 fueron correctamente clasificados y los 39 restantes son errores de clasificación.

En resumen, de acuerdo a los resultados obtenidos con la técnica de clasificación, permitió determinar las rutas de navegación (características más comunes de navegación) que tienen los estudiantes con el sistema y las herramientas con las que más interactúan y usan para su aprendizaje y comunicación con el profesor/tutor responsable de los cursos dentro del sistema.

Resumen de los resultados obtenidos en la minería de los datos

Dentro del pre-procesamiento que se realizó en los datos, se puede sacar en resumen todas las interacciones realizadas por los estudiantes durante la navegación dentro de la plataforma del EVA, se detalla las herramientas a las cuales acceden con mayor frecuencia y las acciones que realizan dentro de cada una de ellas para interactuar con el docente y sus compañeros.

Según los resultados obtenidos se puede decir que los estudiantes más interactúan con las páginas de las actividades (*foros, tareas, cuestionarios*), las páginas de apoyo (*recursos, utpl, subir de archivos*), y con un menor porcentaje de interacción están las páginas de colaboración y comunicación (*chat, blog, mensajería, twitter*), y con un índice más bajo de interacción están las páginas sobre información general (*perfil de usuario, calendario, noticias, eventos, etc.*) dentro de la plataforma del EVA, y realizan las siguientes acciones dentro de los cursos en los que se encuentran enrolados.

Para la interpretación de los resultados obtenidos de la implementación de las técnicas, a continuación se presenta los datos cualitativos y cuantitativos que se determinaron:

El curso con mayor índice de interacción por parte de los estudiantes es la de “Sistemas y Redes Distribuidos”.

Las herramientas a las que acceden con mayor frecuencia los estudiantes y las acciones que realizan dentro de cada una de las mismas, en los diferentes cursos son los foros (*revisar foros, revisar discusiones*), tareas (*revisar, subir tareas*) y recursos (*revisar*) y al perfil de usuario y quiz (*revisar*).

- En su mayoría los estudiantes en los cursos, prefieren iniciar según el tema planteado por el docente, realizar sus propias conversaciones a postear en los foros de otros compañeros del curso.
- Pocos son los estudiantes que se remitan a dar respuesta a temas dentro de los foros iniciados.
- Los estudiantes sólo acceden a las herramientas donde el docente ingresa para dejar alguna actividad como: foros, cuestionarios, recursos y muy pocos al blog.
- La mayor parte de los estudiantes usan las mismas herramientas.
- Todos los estudiantes revisan los mensajes ingresados en el curso por el tutor.
- Casi en su totalidad los estudiantes no prestan interés en los mensajes comentados por los compañeros.
- Todos los estudiantes una vez que ingresan en los cursos revisan sólo las herramientas en las que el tutor sube o deja alguna tarea, como foros, recursos, tareas y exámenes.
- La interacción más activa de los estudiantes dentro de las herramientas del curso es con la herramienta de los foros.

Los resultados obtenidos de los datos procesados mediante las técnicas con las cuales se realizó el proceso de minería se presenta en la **Tabla 9**, en la que se puede apreciar que los estudiantes navegan con mayor frecuencia por las páginas de actividades (*foros, tareas, cuestionarios*) y luego con las páginas de Apoyo (*recursos, subida de archivos y utpl*); y con un nivel muy bajo de interacción están las páginas de colaborativas (*twitter, blog y foros*).

Fuente: Ruiz, F., (2014).

Accesos	Resultados técnica de Asociación	Resultados técnica de Clustering	Resultados técnica de Clasificación
Páginas de información general <i>Perfil de usuario.</i>	914 914	443 443	483 483
Páginas de Actividades <i>Foros</i> <i>Cuestionarios</i> <i>Tareas</i>	13178 6282 901 780	3127 2240 524 363	14388 9824 1324 800
Páginas de Colaboración <i>Twitter</i> <i>Blog</i>	4 0 4	3 0 3	4 0 4
Páginas de Apoyo <i>Recursos</i> <i>Subida de archivos</i> <i>Utpl</i>	959 629 198 132	600 491 90 19	956 600 192 128

Tabla 9. Resultados obtenidos con las técnicas de minería.

Los resultados obtenidos de las tres técnicas de minería no varían mucho los valores obtenidos, en las técnicas de **Asociación** la herramienta con mayor número de interacción es las **tareas**, mientras que en la técnica de **Clustering** es los **foros**; además varían las materias en las cuales navegan con mayor frecuencia los estudiantes, como por ejemplo en la técnica de Clasificación interactúan más con la materia de **Redes y Sistemas Distribuidos**, y en la técnica de Clustering es el curso de Modelamiento de datos, y **Sistemas basados en el conocimiento**, y en la técnica de Asociación el curso con mayor interacción por parte de los estudiantes es **Control de Calidad**.

Gephi para la visualización de las interacciones de los usuarios en el EVA

GEPHI es una herramienta para la exploración, navegación y análisis de grafos; de acuerdo a [Medrano J.; Alonso J.; Figuerola C., s.a] que permite a los usuarios interactuar con las distintas representaciones, manipular las estructuras, formas, y colores que revelan propiedades ocultas. Utiliza un motor de renderizado 3D para mostrar las grandes redes en tiempo real y para acelerar la exploración. el objetivo es ayudar a los analistas de datos a hacer hipótesis, descubrir patrones, aislar singularidades en las estructuras o encontrar fallas en los datos.

La gráfica muestra de forma visual mediante grafos la interacción de los estudiantes y los profesores dentro de la plataforma del EVA en las herramientas disponibles en las materias de noveno ciclo de la UTPL, Modalidad Abierta que es la muestra que se ha tomado de la base del EVA, con la finalidad de cumplir los objetivos planteados en la presente tesis.

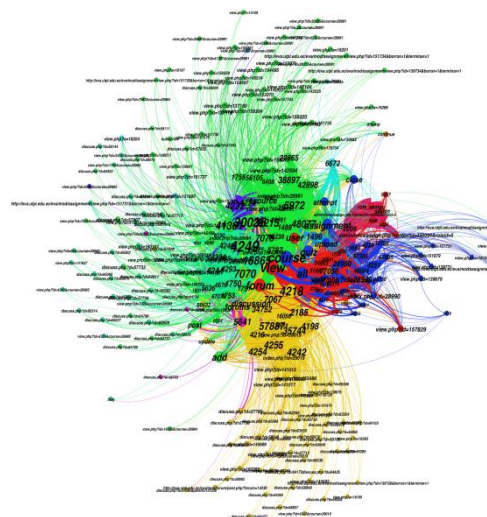


Fig. 8. Interacción de los estudiantes en el EVA. Fuente: Ruiz, F., (2014).

Los datos mostrados en la Fig. 8, son el resultado de la interacción que tienen los estudiantes en los diferentes cursos, en el cual tanto los estudiantes que ingresan a realizar alguna acción en el sistema, la interacción se muestran en diversos colores, más adelante en la sección de la minería de datos se detallará mediante grafos la interacción que los estudiantes tienen con cada una de las herramientas del EVA, el grafo describe todas las páginas visitadas por parte de los estudiantes cuando ingresan a realizar alguna acción en las herramientas y su interactúan con las mismas.

Interpretación de los datos del grafo

En la interacción de los estudiantes con las herramientas del EVA, se las clasificó por secciones por páginas: *páginas de información general*, *páginas de Actividades*, *páginas de comunicación y colaboración* y *las páginas de Apoyo*; para de esta manera según las rutas de navegación determinar las herramientas más frecuentadas por los estudiantes, se puede ver en la Fig. 8, que tanto los profesores como los estudiantes no acceden a todas las herramientas cuando navegan en el sistema, las páginas de actividades tienen el 43.84% de ingresos; para revisar las actividades planteadas por el profesor. El 26.85% ingresan en los *foros*; con menor interacción ingresan a los *cuestionarios*, 12.53%; otra de las actividades con mayor interacción son los *recursos* con 5.83%, y luego están las *tareas* con el 5.51% de interacción, seguido están los ingresos al *perfil del usuario* que tiene un 5.40%, y por último está los blogs con un mínimo del 0.03% y al servicio del *twitter* con un porcentaje casi nulo.

Los datos obtenidos de la Fig. 8, muestran las materias que fueron tomadas para desarrollar el

estudio de minería de datos para determinar la interacción de los estudiantes con las herramientas del EVA, las cuales corresponden a noveno ciclo de Informática de la modalidad abierta, presentan un mayor índice de interacción con las herramientas del EVA del periodo académico Agosto 2010– febrero 2011.

Fuente: Ruiz, F., (2014).

MATERIAS	N° ESTUDIANTES	N° INTERAC.
Administración de Herramientas CASE	21	1331
Control de Calidad	24	1946
Modelamiento de Datos	31	1719
Sistemas Basados en el Conocimiento	11	1339
Sistemas de Control	24	1364
Redes y Sistemas Distribuidos	13	2234

Tabla 10. Datos de interacción en las materias de Informática.

Según los resultados obtenidos de la gráfica realizada con Gephi, las materias con mayor número de interacción son las materias de **Modelamiento de datos [A]** con 1946 interacciones y **Sistemas basados en el conocimiento [A]** con 1719 interacciones en el sistema; luego esta el curso **Planeación estratégica [A]** con un total de 1339 interacciones. Con menor nivel de interacción está el curso de **Teoría de colas [A]**, y por último el curso **Administración de herramientas [A]**. Además están las páginas más accedidas por los estudiantes dentro del EVA, las páginas de actividades como son los foros, cuestionarios y las tareas; y las páginas de apoyo como son los recursos y links de descargas; éstas son las páginas de las herramientas con las cuales los estudiantes navegan para la comunicación con el tutor y aprendizaje dentro de los cursos.

Obtención de resultados de interacción en los cursos de experimentación

Los datos obtenidos de la interacción de los estudiantes que navegan en las páginas dentro de las herramientas del EVA se detallan más adelante, y se presentan los resultados de manera estadísticos y probabilísticos (minería) y visual (grafos).

Cursos.- Esta herramienta permite al estudiante ingresar y revisar todas las actividades propuestas por el profesor/tutor ya que en aquí se almacena toda la información relacionada a los cursos en los que se ha enrolado el estudiante y las cuales están disponibles dentro de la plataforma. Esta información es útil para tener un registro de

información de todas las actividades y eventos realizados por los usuarios en base a las materias, categorías en esta herramienta dentro de la plataforma.

Fuente: Ruiz, F., (2014).

RESULTADOS DE INTERACCIÓN EN LOS CURSOS		
Acción realizada por el estudiante	Estudiantes que visitan la página	Total de visitan las páginas
Revisa páginas de Administración de herramientas.	13	1717
Revisan páginas de Control de Calidad.	31	2544
Revisa páginas de Modelamiento de datos	24	2362
Revisan páginas de Redes y sistemas distribuidos.	22	5073
Revisa páginas de Sistemas basados en el conocimiento	24	2193
Revisa páginas de Sistemas de control.	12	1317

Tabla 11. Datos de Interacción de los estudiantes en los cursos.

En los resultados obtenidos de la interacción de los estudiantes con las páginas de las actividades dentro de la herramienta de los cursos del EVA, la cual se puede apreciar en la **Tabla 11**, muestra las acciones realizadas por los estudiantes dentro de las páginas del curso, y como se puede apreciar la mayor parte de los estudiantes ingresan al sistema a revisar las diferentes actividades propuestas por el tutor dentro de los diferentes cursos minados.

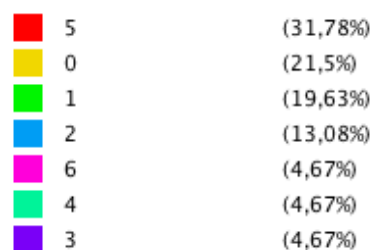


Fig. 9. Valores de rango de interacción de los estudiantes en los cursos.

Fuente: Ruiz, F., (2014).

La **Fig. 9**, muestra 7 comunidades, las cuales representan los grupos con el número de ingresos por parte de los estudiantes en los diferentes cursos, herramientas, servicios y acciones realizadas en cada uno de ellos, además estas comunidades presentan los valores de rango según el número de interacciones y accesos que tienen los estudiantes a las páginas de cada uno de los cursos, el valor porcentual de accesos y el número de enlaces realizados a cada una de las herramientas y acciones dentro del EVA por parte del estudiante. Esta gráfica se la obtuvo con la herramienta de Gephi, permite observar mediante valores porcentuales las herramientas con mayor número de visitas de manera global en todos los

cursos tomados para el proceso de minería. Para obtener estos datos se exporto en formato .CSV desde la base del EVA.

Resultados:

Los resultados obtenidos con Gephi concuerdan con la mayoría de los resultados obtenidos con las técnicas de minería, entre ellas se destacan las siguientes:

Todos los estudiantes ingresan a revisar las actividades propuestas por los profesores de los cursos en los que se encuentran enrolados. Los estudiantes con mayor número de interacciones en las herramientas de la plataforma son los enrolados en las materias de: Modelamiento de datos, Sistemas basados en el conocimiento y Teoría de colas.

Los cursos con menor interacción con las herramientas del EVA son los estudiantes pertenecientes a las materia de Administración de herramientas CASE y Planeación estratégica.

Las herramientas más utilizadas por los estudiantes cuando interactúan en la plataforma del EVA son: Foros, Cuestionarios, Recursos, Tareas, Perfil de usuario y Subida de archivos.

Según los resultados obtenidos de la minería de datos el curso con mayor nivel de interacción en la plataforma es la asignatura Modelamiento de datos, a diferencia de los resultados obtenidos con Gephi que presenta que la asignatura con mayor nivel de interacción es Teoría de colas y la asignatura con el mínimo nivel de interacción es Administración de herramientas CASE.

Ambos resultados concuerdan que las páginas más frecuentadas en los seis cursos por los estudiantes son en las páginas de las actividades (*foros, cuestionarios, tareas*) y los recursos; y las páginas con menor nivel de interacción son las páginas de Colaboración y comunicaron (*blog, chat*).

La herramienta más frecuentada por los estudiantes enrolados en los cinco cursos, en ambos caos es la herramienta de los foros.

Los estudiantes que interactúan con los foros primeramente revisan las discusiones propuestas por los demás compañeros, antes de agregar una nueva discusión.

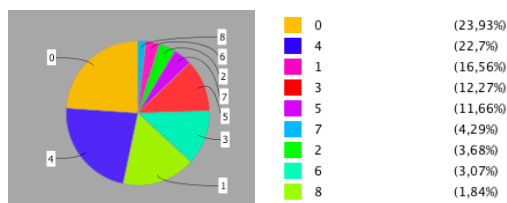


Fig. 10. Resultado de las interacciones realizadas en el EVA.

Fuente: Ruiz, F., (2014).

La Fig. 10, muestra una gráfica en detalle porcentual y mediante un diagrama circular seccionada con las diversas herramientas a las que accede el estudiantes en una división porcentual de acceso según su frecuencia de navegación en las páginas del EVA.

Representación de indicadores y atributos en un modelo de datos

De la extracción de los resultados obtenidos del análisis de minería se pretende extraer indicadores y atributos de navegación de los estudiantes del comportamiento cuando interactúan con las herramientas del EVA, y representarlos en un modelo de datos.

Resultados de los comportamientos de navegación de los estudiantes

Los estudiantes interactuaron, y realizaron varias actividades:

- ✓ Ingreso al curso, para revisar las actividades en el curso.
- ✓ Interacción con las herramientas del curso: foros, cuestionario, tareas, recursos, perfil de usuario, utpl y blog.
- ✓ La maoria de los estudiantes de los seis cursos evaluados, interactúan con la herramienta de los foros.
- ✓ Los estudiantes ingresan a revisar en el curso las actividades que el profesor ha puesto a disposición del estudiante para su desarrollo.
- ✓ Participación de los estudiantes en la herramienta de los foros propuestos por el tutor en el curso.
- ✓ Conforme revisan los foros los estudiantes adicionan una discusión dentro de los foros dentro el curso.
- ✓ Los estudiantes antes de adicionar una discusión en el foro, primero revisan las discusiones ingresadas por otros estudiantes.

- ✓ Durante el proceso de interacción en los cuestionarios los estudiantes realizan un sólo intento antes de terminarlos.
 - ✓ En base a los recursos subidos por el profesor, los estudiantes desarrollan las actividades en el sistema.
 - ✓ En base a las tareas subidas por el profesor, los estudiantes desarrollan las tareas para luego subidas al EVA.
- **Indicadores obtenidos en las actividades de aprendizaje**

Se busca determinar las herramientas con las que el estudiante interactúa con mayor frecuencia dentro del sistema, para luego seleccionar los indicadores de navegación encontrados y representarlo en un modelo de datos del usuario.

Fuente: Ruiz, F., (2014).

INDICADOR	DESCRIPCIÓN	ATRIBUTOS
Nivel de interacción	Este indicador muestra el nivel de interacción que tienen los estudiantes con las herramientas de la plataforma del EVA.	Número de ingresos a los foro Número de ingresos a los cuestionarios Número de ingresos a los recursos Número de ingresos a los tareas Número de ingresos al perfil Número de ingresos a subir archivos Número de ingresos al twitter Número de ingresos a los blogs.
Nivel de interés	Este indicador presenta el nivel de interés que tiene el estudiante con las materias que esta cursando según el número de accesos en las herramientas del EVA.	Número de veces que revisan los foros Veces que revisan los cuestionarios Veces que realizan los cuestionario Número de veces que revisan las tareas Número de veces que suben las tareas Número de veces que revisan los recursos.
Nivel de colaboración	El indicador mide el nivel de colaboración que tienen los estudiantes con las cuales participa dentro del EVA.	Veces que visitan los foros. Número de blogs revisados. Número de conversaciones que el estudiante inicia. Número de conversaciones que el estudiante participa Número de veces que ingresan al twitter. Veces que el estudiante revisa los mensajes. Número de respuestas del estudiante en mensajes.

Tabla 13. Indicadores obtenidos de la interacción de los estudiantes en el EVA.

Los indicadores establecidos para el modelo de datos del usuario se los obtuvo a partir de las interacciones realizadas por los estudiantes en las herramientas del EVA, y el interés que presenta en realizar las actividades propuestas por el docente en cada uno de los cursos.

CONCLUSIONES

El análisis de los resultados se lo obtuvo de los experimentos realizados de la navegación e

interacción con las herramientas con las que cuenta el estudiante en el EVA y de las cuales hace uso.

- La implementación de técnicas de minería de datos aplicados a la educación, los cuales se relacionen con la navegación ayudan a descubrir los intereses que tienen los estudiantes con las herramientas del EVA, para encontrar cuales son las preferencias de los estudiantes para su aprendizaje y colaboración con el profesor y determinar el nivel de aporte del estudiante en las herramientas del sistema
- El algoritmo con resultados más óptimos y eficientes para realizar este tipo de análisis es el Algoritmo GSP y J-48, ya que son algoritmos para determinar las rutas de navegación que tienen los estudiantes durante la interacción con las herramientas, además son algoritmos que determinan el número total de accesos que el estudiante realiza en las páginas del EVA.
- Con el algoritmo **J-48**, se obtuvieron reglas de clasificación según las acciones que el estudiante realiza en las herramientas dentro del conjunto de entrenamiento, existen elementos en común, pero con variación en los tiempos de ejecución y la clasificación de los conjuntos. Este algoritmo es muy sensible en cuestión de los cambios que se realizan en los datos del conjunto de entrenamiento, debido a lo complicado que puede ser determinar, a partir de la "integración" de modelos de minería de datos correspondientes a fuentes de datos homogéneas.
- En los resultados obtenidos de la implementación de la técnica de Clustering, se obtuvieron datos similares entre grupos, con la variación de un clúster y el tiempo de ejecución; en el resultado de los algoritmos de las Reglas de asociación los resultados variaron un poco más, a diferencia del algoritmo **GSP**, el cual devuelve todas las secuencias de ingreso por parte de los estudiantes en determinadas páginas, lo cual no pasa con el algoritmo de A-priori, ya que este algoritmo devuelve sólo la acción realizada pero no el total de accesos que el estudiante ha realizado en las páginas.
- Con el algoritmo **A priori** las reglas que se obtuvieron a partir de los conjuntos de entrenamiento disjuntos se los realizó con valores de soporte y confianza diferentes, para buscar los resultados más óptimos, en los experimentos se obtuvieron similares características en los resultados obtenidos en el conjunto de datos centralizados. En estos

experimentos es posible determinar un modelo de minería de datos válido para el conjunto las reglas obtenidas.

- La técnica Clustering ayuda a agrupar a los estudiantes para encontrar características similares en el comportamiento que tienen cuando interactúan con el sistema, pero no ayuda a determinar rutas de navegación de los estudiantes en el EVA.

Con la implementación de técnicas de minería Web en los datos de interacción de los estudiantes se puede determinar el nivel de aporte que brindan las herramientas en el aprendizaje, la comunicación y relación del estudiante con el profesor en el curso.

Las técnicas visuales mediante grafos, ayudan en la interpretación de los resultados obtenidos para determinar el nivel de aporte que tienen los estudiantes en los cursos.

Los algoritmos con los que se experimentó determinan que los estudiantes tienen un mayor nivel de interacción en los *Foros*, con un 72%, los cuestionarios con el 27%, los recursos el 15%, las tareas con 8%, el blog 0,7% y el twitter 0,3 % de interacción.

TRABAJOS FUTUROS

Se puede considerar en trabajos futuros en el área de aplicación de minería en entornos e-learning, en el aprendizaje-enseñanza para:

- ✓ En base al modelo de datos con los indicadores de comportamiento en la navegación del estudiante propuesto en este trabajo, el profesor podría tomar esta información como referencia que le permita predecir, en base al uso de las herramientas el interés del estudiante en las actividades de las asignaturas.
- ✓ Anexión de nuevos indicadores de navegación de los estudiantes con las demás herramientas con las que dispone el EVA, que no se hayan considerado en este trabajo, y con ello, obtener indicadores de rendimiento en los estudiantes.
- ✓ Realizar un estudio con los indicadores obtenidos para saber en qué medida influye la navegación de los estudiantes en el EVA para su aprendizaje, y el rendimiento académico.
- ✓ Se pueden realizar un análisis comparativo propuestos en otros estudios similares, con

algoritmos iguales o más eficientes en cuanto a tiempo de ejecución y resultados, para determinar cuál es el mejor candidato para mostrar reglas de asociación con datos relativos en entornos e-learning

BIBLIOGRAFÍA

[Brusilovsky, 1996] Brusilovsky, P. (1996). Methods and techniques of adaptive hypermedia. *User Modeling and User Adapted Interaction*, 6(2-3), 87-89.

[Brusilovsky, 1998] Pedro Brusilovsky y Pesin Leónidas. Apoyo a la navegación de adaptación de la Educación Hipermedia: Una evaluación de la ISIS-tutor "en Revista de Informática y Tecnologías de la Información, 6 (1) ,27-38, 1998. 1998.

[Gaudioso E. & Boticario J., 1997] Elena Gaudioso and Jesús G. 'Towards Web-based adaptive learning communities' Boticario Artificial Intelligence Department Facultad de Ciencias, UNED c/ Senda del Rey, 9 28040 Madrid, Spain

[Romero, Ventura & de Castro, 2005] Romero C.; Ventura S.; de Castro C. y García E. (2005, Julio - Diciembre). Algoritmos Evolutivos para descubrimiento de Reglas de Predicción en la Mejora de más Educativos Adaptativos Basados en Web. ADIE, Asociación para el Desarrollo de la Informática Educativa ISSN: 1699-4574, pp 47-60.

[Moreno M.; Miguel L.; García F.; Polo M., s.a] María N. Moreno García*, Luis A. Miguel Quintales, Francisco J. García Peñalvo y M. José Polo Martín, 'APLICACIÓN DE TÉCNICAS DE MINERÍA DE DATOS EN LA CONSTRUCCIÓN Y VALIDACIÓN DE MODELOS PREDICTIVOS Y ASOCIATIVOS A PARTIR DE ESPECIFICACIONES DE REQUISITOS DE SOFTWARE'. Universidad de Salamanca. Departamento de Informática y Automática

[Koch, 2000] Koch, N. (2000).] Software Engineering for Adaptive Hypermedia Systems. Reference Model, Modeling Techniques and Development Process. Unpublished PhD Thesis, Ludwig-Maximilians- Universität München, Munich.

[Pallof & Pratt, 2001] Pallof R. M. and Pratt K. Lessons from the cyberspace classroom. San Francisco: Jossey-Bass. 2001.