

# Análisis de Datos en Mundos Virtuales Educativos

Juan Cruz, Roberto Therón, Emilianita Pizarro, Francisco J. García-Peñalvo

**Title**—Data Analytics in Educational Virtual Worlds.

**Abstract**—Este artículo presenta el proceso de investigación asociado al análisis de datos basado en Mundos Virtuales Educativos. Este análisis busca facilitar la comprensión a docentes y responsables de estos Mundos Virtuales sobre el uso y aprovechamiento de estas plataformas educativas por parte de los estudiantes, así como poder evaluar la utilidad o adecuación de estos sistemas en el proceso educativo que se haya diseñado. En este artículo se incluyen diversos aspectos del análisis como los referentes a los procesos de recolección de información, objetivos de análisis que pueden plantearse, y un conjunto de propuestas de posibles soluciones para comprender el conocimiento extraído en estos procesos.

**Index Terms**—Educational Virtual Worlds, Mundos Virtuales Educativos, Data Analytics, Análisis de Datos, Learning Analytics, Análisis de Aprendizaje, Data Mining, Minería de Datos, SecondLife, OpenSim

## I. INTRODUCCIÓN

LOS Mundos Virtuales ofrecen nuevos métodos de interacción y uso de recursos a través de Internet, los usuarios tienen su representación virtual mediante *avatares* o *alter ego virtuales* y se les proporcionan una serie de posibilidades de comunicación, colaboración, e interacción casi de cualquier tipo, tanto con el entorno virtual 3D como con el resto de usuarios que tienen presencia en dichas plataformas. Además, ofrecen un entorno inmersivo, donde los usuarios experimentan una sensación de estar *dentro* de las situaciones y los eventos que ocurren, de un modo totalmente distinto a cómo se perciben en otras plataformas como la Web, o servicios informáticos en la Red.

Juan Cruz es miembro del GRupo de Investigación en InterAcción y eLearning (GRIAL) de la Universidad de Salamanca. Instituto Universitario de Ciencias de la Educación (IUCE), Paseo de Canalejas, 169, 37008 Salamanca (España) (e-mail: [juancb@usal.es](mailto:juancb@usal.es))

Roberto Therón es miembro del GRupo de Investigación en InterAcción y eLearning (GRIAL) y Profesor Titular de la Universidad de Salamanca. Facultad de Ciencias, Plaza de los Caídos s/n 37008 Salamanca (España) (e-mail: [theron@usal.es](mailto:theron@usal.es))

Emilianita Pizarro es Directora del Servicio de Inserción Profesional, Prácticas y Empleo de la Universidad de Salamanca. Patio de Escuelas 1, 37008, Salamanca (España) (e-mail: [mili@usal.es](mailto:mili@usal.es))

Francisco J. García es Director del GRupo de Investigación en InterAcción y eLearning (GRIAL) y Profesor Titular de la Universidad de Salamanca. Facultad de Ciencias, Plaza de los Caídos s/n 37008 Salamanca (España) (e-mail: [fgarcia@usal.es](mailto:fgarcia@usal.es))

DOI (Digital Object Identifier) Pendiente

Estas características hacen que pueda presuponerse como un entorno óptimo para llevar a cabo labores de aprendizaje, a través de la representación de la realidad en estos Mundos Virtuales, de hecho, en los últimos años se ha escrito bastante sobre las ventajas de entornos virtuales multi-usuario para la enseñanza debido a su relación con el concepto de *aprendizaje experimental* [1] [2] [3] [4], a pesar de ello, también se observa que no termina de arraigar este uso educativo de forma generalizada en este tipo de plataformas [5] [6].

Según esta teoría del Aprendizaje Experimental, el aprendiz consigue una experiencia y conocimiento más profundo a través de un ciclo de aprendizaje que consiste en cuatro etapas: *Experiencia Concreta*, *Observación Reflexiva*, *Conceptualización Abstracta* y *Experimentación Activa*.

Dentro del aprendizaje que se da en los entornos virtuales multi-usuario (como los Mundos Virtuales), se incluyen más actores además del alumno, entre otros roles se pueden encontrar docentes, profesionales de distintas materias que colaboran en el planteamiento de contenidos o eventos en el Mundo Virtual, administradores de sistemas, investigadores relacionados con estos entornos, etc. Para todos estos actores que intervienen en estos procesos es importante conocer qué ocurre dentro de estos entornos virtuales 3D y poder extraer conclusiones sobre ellos, de cara a la mejora o evaluación de los eventos, prácticas o sistemas planteados para el aprendizaje. Por ello es interesante poder definir y realizar mediciones asociadas a los eventos y situaciones que se producen dentro del Mundo Virtual y conocer el uso que se realiza de él.

Partiendo de esa motivación, en este artículo se pretenden expresar una serie de consideraciones previas al análisis de datos, unos objetivos que se puedan definir de cara a la recogida y análisis de datos, unas nociones sobre los procesos de extracción de datos de los Mundos Virtuales y diversas propuestas y ejemplos de cara al análisis de los datos recopilados y procesados. Además se incluirá un caso real con la aplicación de alguna de las propuestas que se realicen, y un apartado que exprese las conclusiones obtenidas de todo el proceso.

## II. ANÁLISIS DE DATOS

### A. ¿Por qué analizar?

En los Mundos Virtuales, especialmente en los comúnmente usados para docencia y educación como pueden ser los

actualmente basados en SecondLife [3] [7] u OpenSim [8] [9], existen una serie de problemas asociados a la información que se puede extraer de ellos.

En este tipo de entornos masivos y multi-usuario, mientras un docente, profesional o responsable de una actividad está conectado al entorno virtual, observando el desarrollo de tareas por parte de los alumnos o público al que se dirigen las actividades, es posible que se perciba el aprovechamiento exacto de la plataforma, los recursos, y el uso que se hace en general del entorno virtual, evaluando, si es el caso, lo observado. En cambio, cuando una persona de ese perfil docente no está presente, la evaluación de las actividades realizadas se vuelven mucho más complejas, obligando al docente a obtener la información por otros medios, entre los que destacan los test realizados dentro del entorno virtual, las evaluaciones mediante controles antes y después del uso del mundo virtual, de modo que se conozca el aprovechamiento que se ha hecho de las actividades [10], etc.

En cambio, existen más opciones de evaluación [11], como pueden ser las basadas en el análisis del uso que hacen los usuarios de una plataforma educativa de este estilo. En un mundo virtual educativo, los alumnos tienen ciertas tareas asignadas, deben moverse, interactuar con el entorno virtual, realizar acciones sobre determinados objetos, comunicarse con otras personas allí presentes, etc. Este tipo de acciones, pueden ayudar al docente o analista a evaluar las actividades, sin orientar únicamente el proceso hacia la adquisición de conocimiento, sino conocer otras variables que pueden ayudar a comprender el uso que hacen los alumnos del sistema y el aprovechamiento de las herramientas [12], a modo de *feedback* hacia las personas que plantean las experiencias inmersivas [13].

En estos casos, es necesario conseguir muchos más datos del entorno virtual, de modo que deben establecerse de forma inequívoca los objetivos del análisis, para trazar después la estrategia de recogida de datos.

### B. Objetivos del Análisis de Datos

Antes de plantear cualquier solución, técnica o metodología aplicable a la información que se obtiene o se puede obtener del entorno virtual, es capital realizar una especificación de requisitos en cuanto a *qué* información se quiere obtener de los datos, es decir, qué objetivos se persiguen a través de la comprensión y análisis de los datos que brindan estas plataformas.

En estos objetivos que se establecen en cuanto a la recuperación, análisis y comprensión de la plataforma que sustente el entorno virtual, pueden darse distintos enfoques:

1) *Comprensión del uso global de la plataforma*: en este grupo de objetivos que ahondan en las cuestiones más puramente de análisis de uso, se pueden englobar métricas como tiempo de uso por usuario, tiempo medio de uso global, fechas de uso de la plataforma, características de exploración del entorno [14], horarios de conexión, zonas geográficas virtuales más visitadas, flujos de movimiento por parte de los usuarios entre territorios del mundo virtual, etc. En general irán destinados a trazar una serie de perfiles de usuarios en

función del uso del Mundo Virtual, que ayuden a conocer o prever posibles patrones de uso del sistema.

2) *Análisis de la interacción con el entorno*: en este conjunto de objetivos se plantea la necesidad de conocer cómo interactúa el usuario con el entorno virtual, sobre qué objetos intenta actuar, qué resultados obtiene, que secuencias de interacción aparecen. Este tipo de objetivos pueden ayudar a trazar grafos de interacción (Fig.1) que especifiquen la sucesión de acciones que realizan los usuarios en el uso de objetos y recursos virtuales en las actividades concretas, grado de uso de elementos informativos, visibilidad de elementos de interacción [15], etc. y de este modo conocer qué hacen los usuarios, qué *no hacen* y podrían hacer, y qué decisiones toman los usuarios de la plataforma cuando se dan en cualquiera de los dos casos.

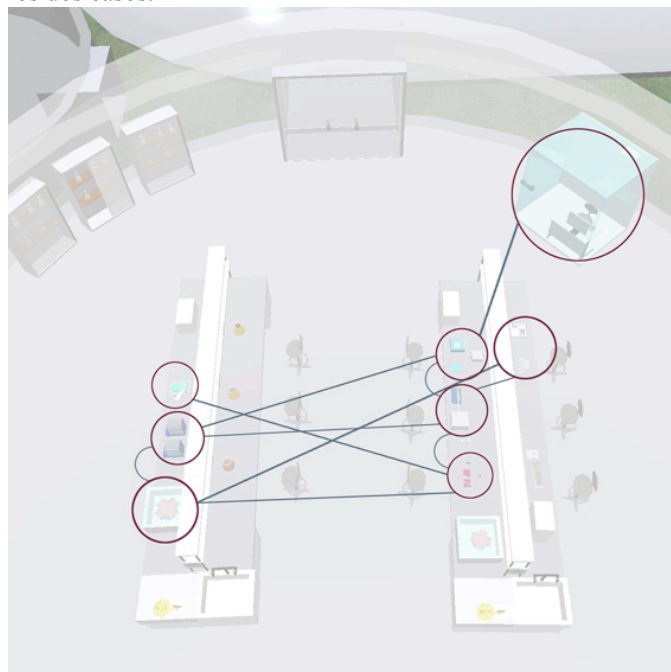


Fig. 1. Posible grafo de interacción entre elementos 3D que se daría al utilizar un usuario los materiales y objetos disponibles en una práctica de aprendizaje.

3) *Análisis de la comunicación e interacción entre usuarios*: conocer cómo usan los usuarios los canales de comunicación en el entorno virtual puede dar una aproximación bastante acertada en cuanto al nivel de colaboración entre usuarios, tan importante en estos entornos de aprendizaje [16].

4) *Análisis de la usabilidad de la plataforma*: en relación con los “objetivos orientados al análisis de la interacción con el entorno”, se puede establecer como objetivo de análisis los errores, fallos y usos incorrectos de la plataforma virtual. De este modo, si es posible determinar patrones de uso o comportamiento que llevan a situaciones inesperadas, poco claras para el usuario o de fallo, los responsables de la plataforma pueden tener información suficiente para corregir esos errores de usabilidad [17] [18].

De forma completa, si hubiera que dar una lista detallada de las posibles tareas que se persiguen dentro del proceso de

análisis de datos de un Mundo Virtual Educativo, se podría aportar el siguiente listado básico:

- 1) *Conocimiento de los patrones horarios de los usuarios.*
- 2) *Conocimiento de los patrones de fechas en los que los usuarios usan las plataformas.*
- 3) *Tiempo medio de uso por usuario y de cada usuario concreto.*
- 4) *Patrones de exploración del entorno virtual.*
- 5) *Flujos de movimiento de los usuarios entre territorios.*
- 6) *Territorios virtuales más comúnmente visitados.*
- 7) *Objetos 3D más usados en espacios acotados (territorio, habitación o estancia, edificio, etc.).*
- 8) *Patrones de uso de objetos embebidos en prácticas o actividades educativas dentro del Mundo Virtual, en función de los tipos de usuarios, grupos de usuarios, en relación a otros objetos cercanos o pertenecientes a la misma actividad, etc.*
- 9) *Datos referentes a objetos no usados, o usados en menor medida.*
- 10) *Uso de elementos informativos.*
- 11) *Grado del uso de herramientas de comunicación.*
- 12) *Análisis léxico de las comunicaciones públicas que se realizan en determinados entornos.*
- 13) *Obtención de patrones en errores de teleport (un teleport es aquel movimiento que hace un usuario al cambiar de territorio virtual de forma instantánea, también se le puede llamar acción de teletransporte).*
- 14) *Fallos en la interacción con los escenarios virtuales.*
- 15) *Fallos en el seguimiento de las prácticas, tanto por error del sistema, como por error del usuario, discerniendo si el usuario se ha equivocado por falta de preparación de cara a la práctica, o por falta de información proporcionada por los objetos.*
- 16) *Relaciones entre software usado como visor del Mundo Virtual y fallos en la realización de los procesos planteados.*
- 17) *Creación de grupos de usuarios o de perfiles en función de sus semejanzas que nos puedan ayudar a tomar decisiones en el futuro.*

### III. EXTRACCIÓN DE DATOS DE UN MUNDO VIRTUAL

#### A. Métodos de Obtención de Datos en los Mundos Virtuales

Como se ha comentado, en la actualidad hay dos grandes sistemas genéricos que lideran el desarrollo de Mundos Virtuales Educativos, como son OpenSimulator (*OpenSim*) y SecondLife. Existen otras opciones, como los desarrollados a medida (mediante *frameworks* de desarrollo de videojuegos o motores 3D); en este caso, el artículo se centra en las características comunes a estos dos sistemas (extensibles muchas veces también a otros entornos).

En general, estas dos plataformas que permiten crear Mundos Virtuales, tienen opciones similares, de hecho, OpenSim surge como alternativa libre a SecondLife, e imita y adopta parte de sus características. Sin embargo, para el tema del que trata este artículo hay una diferencia capital entre los

dos sistemas: en OpenSim los administradores tienen acceso completo a los datos que se generan dentro del Mundo Virtual (a través de archivos *log*, bases de datos, etc.), mientras que en SecondLife, esto no es posible, ya que la empresa dueña, *LindenLab*, tiene una política bastante cerrada en cuanto a estos aspectos [19].

Como resumen de la obtención de datos, se puede reseñar que en SecondLife, es necesario crear objetos que *escuchen* y detecten lo que ocurre a su alrededor, y almacenen en archivos o envíen a servidores externos los datos recogidos; no es posible tener acceso a ningún tipo de analítica o informe acerca del uso de regiones o espacios 3D alojados en esta plataforma. A ello se suma, que estos objetos virtuales están limitados por los recursos (*API* [20], etc.) que proporciona el entorno virtual para la programación y configuración de los mismos. Por ello, SecondLife dificulta enormemente la obtención de datos en cantidad y calidad adecuadas para realizar la recolección, procesado y posterior análisis de los datos.

Por el contrario, OpenSimulator, como se reseñaba anteriormente, se trata de un sistema *Open Source*, pensado para usarse desde un servidor propio y no dependiente de ningún agente o empresa externo, lo cual hace que sea perfectamente posible la consulta, manipulación o recuperación de cualquier dato que el sistema recoja en cualquiera de sus archivos de *log*, bases de datos, etc. También, en contraposición a SecondLife, OpenSimulator permite la programación en varios lenguajes, tanto en *LSL*, como en *C#* u *OSL* [21], proporcionando más posibilidades en la programación de objetos virtuales.

#### B. Extracción de Datos en OpenSimulator

En cualquier caso, y de manera general, se pueden definir dos tipos de extracción de datos del entorno virtual:

1) *Datos extraídos de ficheros de servidor y bases de datos:* en este tipo de extracción de datos, se examinan archivos de *log*, bases de datos, y otros posibles ficheros o sistemas de persistencia de datos donde se pueda encontrar información relevante sobre la ejecución del Mundo Virtual. Este tipo de extracción se aplica típicamente, como se ha comentado, en el entorno de OpenSim, ya que SecondLife, no contempla proporcionar acceso a este tipo de recursos.

2) *Datos extraídos de objetos 3D del propio Mundo Virtual:* en este tipo de extracción de datos son los propios objetos 3D del Mundo Virtual los que informan de alguna forma de la actividad que registran. Los métodos de información y extracción de datos en este tipo se basan en el paso de mensajes a través de Internet hasta llegar a servidores que los recogen.

En el caso de la extracción de datos sobre sistemas persistentes que se proporcionan por defecto, hay que tener muy en cuenta la estructura interna del sistema, para comprender cómo se organiza la recogida de datos que realiza la plataforma. Se propone como ejemplo una estructura de tipo *grid* en OpenSim [22] [23], ya que en casi todos los Mundos Virtuales creados a través de esta plataforma con previsiones

de carga de servidor significativas y planificación de extensibilidad y crecimiento, usan esta arquitectura.

Tomando esta arquitectura, existen dos archivos de *log* donde se recogen los datos que generan cada programa servidor (*Robust.exe* y *OpenSim.exe*). Los datos generados por el programa servidor *Robust.exe* quedan recogidos en el fichero *Robust.log*, de la misma forma los generados por el programa *OpenSim.exe* se almacenan en el fichero *OpenSim.log*.

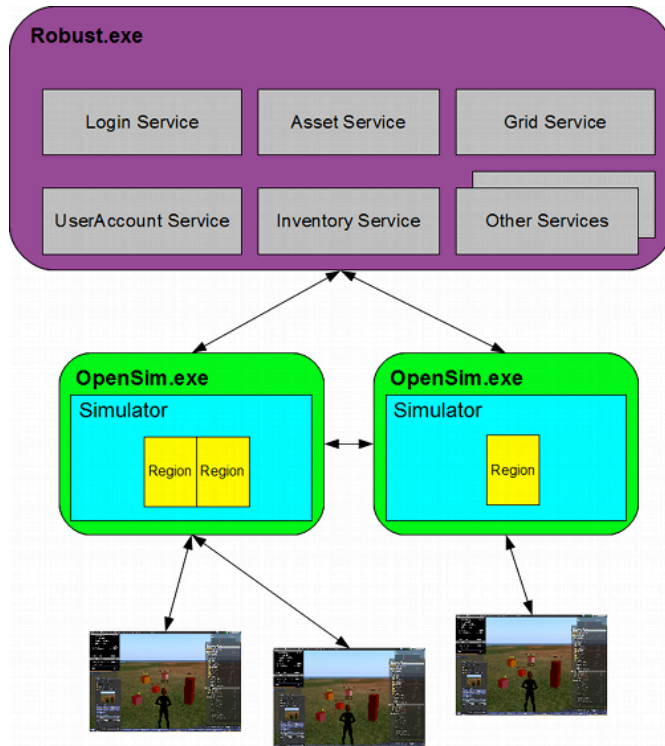


Fig. 2. Esquema de la arquitectura de servidores con topología Grid en Mundos Virtuales desarrollados con OpenSim

Como se observa en el esquema de la arquitectura mostrado en la Fig.2, el programa servidor *Robust.exe* se encarga de gestionar los eventos de *login*, recuperación de información sobre un territorio del Mundo Virtual, el registro de actividad en una cuenta de usuario etc. Toda esta información, se guarda en el fichero *Robust.log*, de modo que será a este archivo al que se recurrirá para obtener datos relacionados con esos eventos registrados

En el caso del programa de servidor *OpenSim.exe*, su respectivo archivo de datos *OpenSim.log* recoge eventos de búsqueda de territorios, salida/llegada entre territorios o regiones virtuales, llamadas a los servidores que gestionan los servicios de VoIP, registro de actividades referentes a los grupos de usuarios, etc.

Por otra parte, en cuanto a la extracción de datos mediante la comunicación entre objetos virtuales y servidores de Internet que recojan los datos que se envíen, se puede obtener información acerca de eventos, como *clicks* sobre los objetos de las prácticas, datos acerca de la posición de los distintos usuarios, tiempo que está un usuario en una determinada zona, recogida de información sobre conversaciones, uso de los

sistemas informativos, etc.

#### IV. PROPUESTAS PARA EL ANÁLISIS DE DATOS

Una vez planteados los objetivos del análisis de datos y las fuentes de datos de donde se puede extraer conocimiento e información de relevancia, hay que plantear qué métodos de análisis se deben seguir para poder resolver las tareas y objetivos de análisis trazados en la sección II. Además, estos métodos de análisis deben tener muy en cuenta la naturaleza de los datos que genera un Mundo Virtual, ya que en caso de que exista un uso regular y continuado, se enfrentarán a grandes cantidades de datos dinámicos que varían a gran velocidad, por lo que deben ser capaces de soportar esas características.

Algunos de los objetivos que se han planteado, no requieren de grandes procesos de ingeniería, ni métodos complejos de tratamiento de datos, sin embargo, otros sí que plantean ciertos retos en cuanto al procesado que debe hacerse para extraer conocimiento de ellos.

Por ejemplo, para el objetivo llamado "*Tiempo medio de uso por usuario y de cada usuario concreto*", bastará con calcular el tiempo que dura cada sesión de conexión por parte de un usuario y hacer la media en el total. En el caso de tiempo medio por usuario concreto, habrá que realizar la misma operación, pero aplicando el método solo a las conexiones realizadas por el usuario seleccionado. Otros objetivos como "*Objetos 3D más usados en espacios acotados*" también incluyen tratamientos sencillos, por ejemplo en este que se plantea será necesario tan solo realizar el cálculo de *clicks* segmentando los datos obtenidos por el ID de objeto.

Sin embargo, muchos de los otros objetivos que se han propuesto plantean retos mayores que los de calcular una media o *contar* números de *clicks*. Por ejemplo, el objetivo con nombre "*Patrones de exploración del entorno virtual*" supone un paso más allá en el análisis de datos, ya no es necesario solo sumar el número de veces que aparece un territorio o que un usuario visita determinadas islas, sino que se pretende extraer conocimiento más avanzado, como es el conocer qué patrones o características migratorias tienen los usuarios dentro del Mundo Virtual. Para este caso se propone la aplicación de técnicas de Minería de Datos [24], concretamente de *Reglas de Asociación*, como se comenta en [25]. Los dominios de aplicación de las reglas de asociación varían desde el soporte de decisiones, a diagnóstico y predicción de eventos o comportamientos de que se representan mediante un grupo de datos de estudio. Mediante la aplicación en este caso se pueden obtener predicciones, con mayor o menor porcentaje de acierto y confianza, sobre el destino del movimiento de un usuario entre territorios dado su lugar de origen, de manera global o especializando el resultado solo para dicho usuario.

En otros casos, como son, por ejemplo, el descubrimiento de patrones de error como en el objetivo "*Obtención de patrones en errores de teleport*", pueden ser necesarios otros algoritmos, como pueden ser los de clasificación. En este caso,

en lugar de intentar encontrar unas reglas de fallo, se puede intentar encontrar un patrón más completo de funcionamiento, que ayude a reproducir el comportamiento completo de ciertos fallos. Como se expresa en [26], los árboles de decisión (un algoritmo concreto de clasificación) se pueden aplicar a cualquier contexto de clasificación; en este caso, cabe destacar que lo relevante de los árboles de decisión es que se presentan con una serie de casos relevantes en la tarea de clasificación e implementan a través de ellos un árbol de decisiones *top-down* (de la raíz a las hojas), guiado por la frecuencia de aparición de esta información relevante. Este tipo de técnicas de clasificación, así como posiblemente otras (otros árboles por ejemplo) podrían tomarse como válidas en algún otro objetivo que se ha planteado, por ejemplo en el denominado “*Patrones de uso de objetos embebidos en prácticas o actividades educativas dentro del Mundo Virtual en función de los tipos de usuarios, grupos de usuarios, en relación a otros objetos cercanos o pertenecientes a la misma actividad, etc.*”, en el cual se podría evaluar de la misma forma las decisiones o caminos que ha ido eligiendo el usuario para llegar a un resultado final.

Por último, en objetivos similares al planteado como “*Creación de grupos de usuarios o de perfiles en función de sus semejanzas*”, se puede hacer uso de otra técnica típica de la Minería de Datos: el *clustering*. El *clustering* aplicado en grupos de usuarios puede ayudar a construir una aproximación de distintos conjuntos y modelos que se corresponden con los patrones de comportamiento para distintos tipos de usuarios [27]. Realizar agrupaciones (*clustering*) de usuarios en comunidades con características de comportamiento similares es el primer paso para establecer perfiles y tipologías de usuarios, ya que se pueden obtener mediante las características comunes de estos clústeres. A través de esta creación de comunidades, conjuntos o clústeres de usuarios, se puede tener, como se comentó anteriormente una mejor información sensible sobre los usuarios, pueda ayudar a tomar decisiones sobre el uso, diseño o usabilidad del entorno virtual, por ejemplo.

## V. UN EJEMPLO REAL DE APLICACIÓN

### A. Consideraciones previas

A continuación se expone un ejemplo con datos reales sobre parte de lo expresado anteriormente; se tomará un conjunto de datos existente y se resolverán dos de los objetivos planteados anteriormente, uno será el objetivo “*Patrones de exploración del entorno virtual*” aplicado en usuarios concretos del entorno virtual; el otro será el denominado “*Obtención de patrones en errores de teleport*”.

Como consideraciones previas, el conjunto de datos usado corresponde al que se ha generado en el Campus Virtual de Prácticas de la Universidad de Salamanca, el entorno USALSIM [28] [29], el cual se desarrolló mediante OpenSim. Como características relevantes para la extracción de datos, este sistema está implementado con una arquitectura de tipo *grid*. En el momento de este experimento, este sistema había

sido usado por unos 140 usuarios divididos en cinco ramas de conocimiento distintas, que habían participado en doce experiencias de aprendizaje y prácticas profesionales durante un periodo de pruebas de dos meses y medio aproximadamente.

En cuanto a la extracción de datos, se han usado solo los archivos de *log* generados por los servidores *Robust.exe* y *OpenSim.exe*. Para este ejemplo, se han extraído datos referentes a la localización de usuarios, por ejemplo, los datos de los *teleport* (salida de un territorio y llegada a otro), llegada a un territorio o isla (*login* o llegada de una *teleport*), salida de un territorio (*logout* o territorio de salida en un *teleport*). Estos datos extraídos se han guardado en un archivo *arff* [30] para su tratamiento con una herramienta de Minería de Datos; en cualquier caso, dependiendo del proceso de minería que se plantee, habrá que evaluar otras opciones de almacenamiento persistente.

En este caso se ha hecho uso de la herramienta de Minería de Datos *RapidMiner* [31], aunque podría haberse hecho uso de otras como puede ser *Weka* [32] u otras aplicaciones típicas para estas tareas.

### B. Minería de datos aplicada al objetivo “*Patrones de exploración del entorno virtual*”

Para este objetivo se ha hecho uso de un proceso de Reglas de Asociación. Para ello, se ha establecido una cadena de operadores (Fig. 3) que realizan las operaciones de lectura del archivo de datos (*arff*), conversión de los datos de tipo Nominal a Binomial, de modo que se obtenga un valor de soporte del conjunto de elementos igual al valor esperado de una variable aleatoria [33] y así puedan ser interpretados por el algoritmo *FP-Growth*, el cual necesita valores no nominales, sino que sigan una distribución de probabilidad discreta. Una vez convertidos, se aplica un operador de tipo *FP-Growth*, que interpreta los datos binomiales en términos de frecuencia de aparición de estos atributos en el conjunto de datos, de modo que después el algoritmo de generación de Reglas de Asociación pueda evaluar estas frecuencias de aparición de datos y establezca relaciones entre ellos, dando lugar así al descubrimiento de patrones y relaciones ocultas en los datos. La salida (*output*) de este operador se conecta finalmente a un operador “*Create Association Rules*”, que será el encargado de crear finalmente estas reglas que asocian distintos datos y hallar los valores de soporte y confianza (comprendido como porcentaje de veracidad) en cada una de ellas.

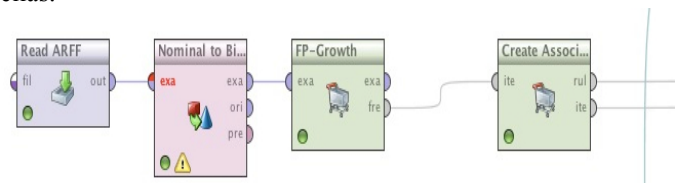


Fig. 3. Operadores usados para la generación de Reglas de Asociación en la herramienta RapidMiner.

Como configuración básica indicar que para encontrar todas las reglas posibles, se establece como valor de soporte 0.1 en el operador *FP-Growth*; de la misma forma se establece en el

operador que crea las reglas de asociación un valor de confianza en las reglas de mínimo 50%. Estos valores se eligen para obtener el mayor número de reglas posible (soporte igual a 0.1) y que ofrezcan una porcentaje de confianza de al menos el 50%.

Una vez ejecutado el proceso, se muestran los resultados mediante una tabla completa con todas las reglas encontradas. Entre esas reglas de asociación se pueden ver algunas muy interesantes, como las referentes a los lugares (territorios) habituales de *login* de un usuario, la frecuencia con la que realiza determinados desplazamientos, etc. Como muestra se ilustra la regla reflejada en la Fig. 4.

18	OrigenTeleport = Animal_Recovery_Center		
CLASE = LlegadaRegion, usuario = Admin_SIPPE		0.049	0.676

Fig. 4. Regla referente a la llegada a un territorio (*Animal\_Recovery\_Center*) por parte de un usuario (*Admin\_SIPPE*), con un soporte de 0.049 y una fiabilidad de 0.676.

En dicha regla se hace referencia a que hay un 67.6% de posibilidades que quien realice una llegada a la región *Animal\_Recovery\_Center* sea el usuario *Admin\_SIPPE*, a pesar de que cuenta con un valor de soporte muy bajo. Evaluando la regla desde el conocimiento de los datos y los sucesos ocurridos en el entorno virtual, un usuario analista podría concluir que es cierta, ya que esa región ha sido un territorio en desarrollo durante mucho tiempo, y no ha tenido una gran afluencia de usuarios corrientes, ya que el que la visitaba principalmente era este administrador del sistema.

### C. Minería de datos aplicada al objetivo “Obtención de patrones en errores de teleport”

En el caso del objetivo del que se trata este apartado se ha aplicado una técnica de Árbol de Decisión. A través de esta técnica se ha podido determinar qué camino, o qué elecciones ha tomado un usuario para llegar a una situación de error en un *teleport* o como habitualmente se le denomina, un *teleport incompleto*.

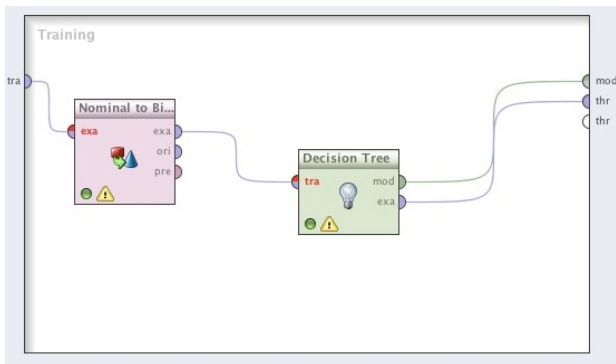


Fig. 5. Etapa de entrenamiento correspondiente al proceso de validación de datos, donde se procesan los datos de valores nominales a binomiales y se genera el Árbol de Decisión.

Análogamente al proceso anterior, se ha definido una secuencia de operadores, que realizan una lectura de los datos

contenidos en un fichero *arff*, se han procesado y se les ha aplicado un proceso de validación de resultados, a través del cual se ha generado un entrenamiento como el observado en la Fig.5, y que ha tenido como resultado la generación de un árbol de decisión.

Cuando se ejecuta el proceso descrito se obtiene un árbol similar al que se muestra en la Fig. 6.

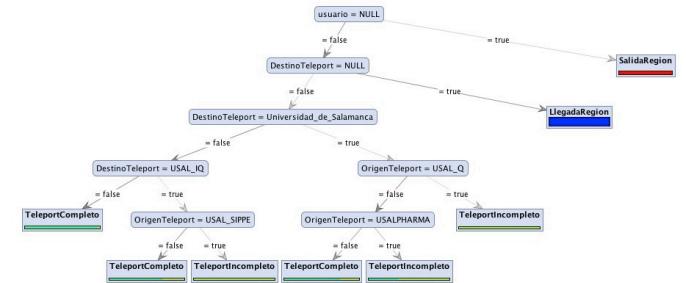


Fig. 6. Árbol de decisión que muestra los caminos o decisiones que llevan a teleports completos o incompletos

Este árbol de decisión expresa claramente cómo se ha llegado a situaciones de *teleport* completo o incompleto, y qué porcentaje se da de acierto en cada uno de estas situaciones (barra de colores verdes en los eventos *TeleportCompleto* / *TeleportIncompleto*). Por ejemplo, si el origen de *teleport* ha sido la región *USAL\_SIPPE*, el territorio de destino ha sido la región *USAL\_Q* y ha habido un usuario concreto implicado en el proceso, se puede asumir que, con los datos de los que se disponen, este proceso casi siempre (con un gran porcentaje) falla.

De la misma forma que en el caso anterior, sobre los flujos de usuarios, un usuario con rol de analista en el sistema, que cuente con conocimiento experto e información completa sobre lo que ocurre en el Mundo Virtual, podría explicar este tipo de resultados que arrojan las técnicas y algoritmos propuestos, ya que por ejemplo, conocería que esta región *USAL\_Q* tuvo problemas relacionados con *firewalls* y puertos de acceso, ya que estuvo restringida al uso de manera general durante un buen periodo de tiempo.

Por ello, se podría decir que de esta unión entre conocimiento experto (del usuario analista) y los resultados de los algoritmos y técnicas planteadas, se da la posibilidad de tomar decisiones respecto del entorno virtual y diseñar acciones correctivas sobre estos eventos no deseados descubiertos, consiguiendo así mejoras constantes en el Mundo Virtual de Aprendizaje.

## VI. CONCLUSIONES

La recogida de datos y su análisis mediante cualquiera de las técnicas propuestas, o con otras, permite a los Mundos Virtuales de aprendizaje dar un salto de calidad en cuanto a la evaluación del aprendizaje y conocimiento de los eventos que ocurren dentro del mundo virtual, permitiendo que se iguale en ese aspectos a muchas otras herramientas usadas habitualmente para *eLearning*. El uso de técnicas avanzadas

para el análisis, como pueden ser las presentadas en este artículo como ejemplo, u otras relacionadas con la minería de datos, permite realizar acciones más complejas, como el descubrimiento de reglas, patrones o conocimiento oculto que de otra forma sería muy complejo de descubrir o tener en cuenta; la agrupación de un conjunto de técnicas de este tipo da una funcionalidad sin limitar el análisis solo a estadísticas de frecuencia, o revisión de archivos de *log* o bases de datos de una forma básica.

Por otra parte, el proceso de análisis sobre un conjunto de datos reales, demasiado grande, complejo, dinámico y de rápido crecimiento, como puede ser el que se genera en un Mundo Virtual de Aprendizaje con un volumen de uso significativo, no puede ser solo abordable desde el punto de vista de la Minería de Datos, ya que es necesario integrar en este proceso de análisis, como ya se ha comentado, el conocimiento del analista humano que interprete los resultados y pueda sacar conclusiones sobre todos estos datos que se le presentan. Por ello, además de las técnicas propuestas en este artículo, no debe perderse la perspectiva de cómo se le presenta a este usuario analista (o de otro perfil que necesite analizar lo que ocurre) la información; por muy potentes que sean los algoritmos aplicados a los datos, si no hay presentación visual de los mismos [34] [35], el usuario nunca será capaz de entenderlos de forma clara y transparente, y tendrá que realizar un gran esfuerzo en revisarlos y comprenderlos. Desde este punto de vista, se propone como trabajo futuro que complementa este trabajo de análisis de necesidades, planteamiento de posibles objetivos y propuestas de soluciones de análisis, el estudio propuesta y construcción de herramientas, principalmente visuales, que hagan de este proceso de descubrimiento de conocimiento y análisis de datos una experiencia rica y clara de cara al usuario.

#### AGRADECIMIENTOS

Los autores agradecen el apoyo del Ministerio de Ciencia e Innovación (proyectos FFI2010-16234 y TIN2010-21695-C02-01) y de la Consejería de Educación de la Junta de Castilla y León (proyecto SA294A12-2).

#### BIBLIOGRAFÍA

- [1] Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development* (Vol. 1). Englewood Cliffs, NJ: Prentice-Hall.
- [2] Boulos, M. N. K., Hetherington, L., & Wheeler, S. (2007). Second Life: an overview of the potential of 3-D virtual worlds in medical and health education. *Health Information & Libraries Journal*, 24(4), 233-245.
- [3] Warburton, S. (2009). Second Life in higher education: Assessing the potential for and the barriers to deploying virtual worlds in learning and teaching. *British Journal of Educational Technology*, 40(3), 414-426.
- [4] Dede, C. (1995). The evolution of constructivist learning environments: Immersion in distributed, virtual worlds. *Educational technology*, 35(5), 46-52.
- [5] Arnab, S., Berta, R., Earp, J., de Freitas, S., Popescu, M., Romero, M., ... & Usart, M. (2012). Framing the Adoption of Serious Games in Formal Education. *Electronic Journal of e-Learning*, 10(2), 159-171.
- [6] Duncan, I., Miller, A., & Jiang, S. (2012). A taxonomy of virtual worlds usage in education. *British Journal of Educational Technology*, 43(6), 949-964.
- [7] Second Life, R. E. (2). de Mayo del 2013 de: <http://secondlife.com/whatis/>
- [8] Opensimulator, R. E. (2). de Mayo del 2013 de: [http://opensimulator.org/wiki/Main\\_Page](http://opensimulator.org/wiki/Main_Page)
- [9] Allison, C., & Miller, A. (2012). Open virtual worlds for open learning.
- [10] Virvou, M., Katsionis, G., & Manos, K. (2005). Combining software games with education: Evaluation of its educational effectiveness. *Educational Technology & Society*, 8(2), 54-65.
- [11] Wojciechowski, R., & Cellary, W. (2013). Evaluation of learners' attitude toward learning in ARIES augmented reality environments. *Computers & Education*.
- [12] Dalgarno, B., & Lee, M. J. (2010). What are the learning affordances of 3D virtual environments?. *British Journal of Educational Technology*, 41(1), 10-32.
- [13] Roussos, M., Johnson, A., Moher, T., Leigh, J., Vasilakis, C., & Barnes, C. (1999). Learning and building together in an immersive virtual world. *Presence: Teleoperators & Virtual Environments*, 8(3), 247-263.
- [14] Freitas, S. D., & Neumann, T. (2009). The use of 'exploratory learning' for supporting immersive learning in virtual environments. *Computers & Education*, 52(2), 343-352.
- [15] Gómez-Aguilar, D.A., Therón Sánchez, R. and García Peñalvo, F.J., 2013. Reveal the relationships among student's participation and its outcomes on eLearning environments: Case study. *Advanced Learning Technologies*, 2013. ICALT 2013. 13th IEEE International Conference on (in press.).
- [16] De Freitas, S. (2006). *Learning in immersive worlds*. London: Joint Information Systems Committee.
- [17] Virvou, M., & Katsionis, G. (2008). On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. *Computers & Education*, 50(1), 154-178.
- [18] Fetscherin, M., & Lattemann, C. (2008). User acceptance of virtual worlds. *Journal of Electronic Commerce Research*, 9(3), 231-242.
- [19] Second Life, Terms of Service. R. E. (2) de Mayo del 2013 de: <https://secondlife.com/corporate/tos.php>
- [20] LSL, Second Life Language Reference. R.E (1) de Julio del 2013 de: [http://wiki.secondlife.com/wiki/Category:LSL\\_Functions](http://wiki.secondlife.com/wiki/Category:LSL_Functions).
- [21] Scripting Documentation, OpenSimulator. R.E. (1) de Julio del 2013 de: [http://opensimulator.org/wiki/Scripting\\_Documentation](http://opensimulator.org/wiki/Scripting_Documentation)
- [22] OpenSimulator, Architecture configuration. R. E. (2) de Mayo del 2013 de: [http://opensimulator.org/wiki/Configuration\\_-\\_Standalone\\_vs\\_Grid](http://opensimulator.org/wiki/Configuration_-_Standalone_vs_Grid)
- [23] Lesko, C. J., & Hollingsworth, Y. A. (2013). Architecting Scalable Academic Virtual World Grids: A Case Utilizing OpenSimulator. *Journal of Virtual Worlds Research*, 6(1).
- [24] Orallo, J. H., Quintana, M. J. R., & Ramírez, C. F. (2004). *Introducción a la Minería de Datos*. Pearson Prentice Hall.
- [25] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., & Verkamo, A. I. (1996). Fast discovery of association rules. *Advances in knowledge discovery and data mining*, 12, 307-328.
- [26] Quinlan, J. R. (1986). Induction of decision trees. *Machine learning*, 1(1), 81-106.
- [27] Paliouras, G., Papatheodorou, C., Karkaletsis, V., & Spyropoulos, C. D. (2000, June). Clustering the users of large web sites into communities. In *Proceedings of the International Conference on Machine Learning (ICML)*, Stanford, California, USA.
- [28] USALSIM, Campus Virtual de Prácticas de la Universidad de Salamanca, R. E. (2). de Mayo del 2013 de: <http://sim.usal.es>
- [29] Lucas, E. P., Benito, J. C., & Gonzalo, O. G. (2013, January). USALSIM: Learning and Professional Practicing in a 3D Virtual World. In *2nd International Workshop on Evidence-based Technology Enhanced Learning* (pp. 75-82). Springer International Publishing.
- [30] Attribute-Relation File Format (ARFF), Weka. R. E. (2) de Mayo del 2013 de: <http://www.cs.waikato.ac.nz/ml/weka/arff.html>
- [31] RapidMiner. R. E. (2) de Mayo del 2013 de: <http://rapid-i.com/>
- [32] Weka, University of Waikato. R. E. (2) de Mayo del 2013 de: <http://www.cs.waikato.ac.nz/ml/weka/>
- [33] Tong, Y., Chen, L., Cheng, Y., & Yu, P. S. (2012). Mining frequent itemsets over uncertain databases. *Proceedings of the VLDB Endowment*, 5(11), 1650-1661.
- [34] Gómez Aguilar, D. A., Therón, R., & García-Peñalvo, F. J. (2009). Semantic spiral timelines used as support for e-learning. *Journal of Universal Computer Science*, 15(7), 1526-1545.
- [35] Borner, K., Hazlewood, W. R., & Lin, S. M. (2002). Visualizing the spatial and temporal distribution of user interaction data collected in three-dimensional virtual worlds. In *Information Visualisation, 2002. Proceedings. Sixth International Conference on* (pp. 25-31). IEEE.