

## **Modelo de Espacio Vectorial con ponderación basada en frecuencia relativa de eventos de navegación en una instancia de Open edX para caracterización del estudiantado**

Fabián Andrés Anacona<sup>1</sup>, Mario Solarte<sup>1</sup> and Gustavo Ramírez González<sup>1</sup>

Universidad del Cauca, Popayán, Colombia  
(fanacona, msolarte, gramirez)@unicauca.edu.co

**Resumen.** El uso de los Cursos en Línea Abiertos y Masivos, MOOC, está creciendo en los últimos años, debido a que diversas Universidades en el mundo los están ofreciendo. Esto ha permitido generar una gran cantidad de datos por las actividades desarrolladas de los estudiantes en los cursos. Estas actividades generan patrones que son registrados en el archivo tracking.log para xMOOC de la plataforma Open edX. Hasta el momento no hay un referente claro en la forma de procesar los patrones de navegación de este archivo. Por tal razón en este documento se presenta una propuesta, para caracterizar los patrones de navegación de los estudiantes en cursos xMOOC de la plataforma Open edX, a través del modelo de espacio vectorial y la ponderación basada en la frecuencia relativa de un evento.

**Abstract.** The use of Massive Open Online Courses, MOOCs, is growing in recent years, because several universities around the world are offering them. This has allowed to generate a large amount of data related to the activities developed by students in these courses. These activities generate patterns that are recorded in the tracking.log file for xMOOC of the Open edX platform. So far there is no clear reference in the way to process the navigation patterns of this file. For this reason, this document presents a proposal to characterize the navigation patterns of students in xMOOC courses of the Open edX platform, through the vector space model and weighting based on the relative frequency of an event.

**Palabras claves:** xMOOC, tracking.log, ponderación, patrón de navegación, evento, caracterización.

### **1 Introducción**

Los Cursos en Línea Abiertos y Masivos (MOOC, Massive Open Online Courses) son una propuesta basada en el Conectivismo, para universalizar la educación y ofrecer educación gratuita y de calidad a las personas desde cualquier lugar en el mundo [1]. Su principal objetivo es brindar alternativas educativas soportadas en tecnologías de Internet de acceso abierto [2-4], además, los estudiantes inscritos a estos cursos virtuales tienen la oportunidad de aprender con los docentes de diferentes Universidades [5]. Los MOOC presentan ciertas características similares como: empleo de material de aprendizaje (mapas de conocimiento, videoconferencias, textos, entre otros) [6], por lo general son de corta duración [7], atraen a miles de

estudiantes de diferentes lugares del mundo [8] y algunos emiten un certificado de aprobación, el cual puede tener un valor monetario [8, 9].

En la plataforma Open edX, se ofrecen xMOOCs, en este curso es el profesor quien crea la mayor parte del material de aprendizaje, el cual generalmente es organizado para el curso; la plataforma cuenta con una estructura lineal de arriba hacia abajo, en secciones, subsecciones y unidades, mostrando a través de la plataforma al estudiante, la organización del curso que plantea el docente. Por otra parte, los xMOOCs son criticados por replicar la enseñanza tradicional, basada en conferencias o clases magistrales y no tener presente las diferentes estrategias de aprendizaje, como la preferencia del estudiante por una estructura más o menos lineal en la organización del material [10, 11].

En el contexto del aprendizaje en línea, existen estudiantes que se sienten seguros en entornos de aprendizaje no lineales, es decir en la definición de sus propios caminos de aprendizaje, lo que indica que navegan libremente sin seguir necesariamente lo sugerido por los creadores de contenido o la estructura de la plataforma. Hay otros estudiantes, que prefieren seguir un camino de aprendizaje definido externamente, como lo impone un profesor o el entorno de aprendizaje en línea [10, 11], las navegaciones de los estudiantes generan patrones de navegación [12] los cuales son registrados en un archivo de texto llamado tracking.log, para la plataforma Open edX [13].

Ahora bien, el auge de los MOOC ha implicado indirectamente un crecimiento de los datos en educación y no hay un referente claro en la creación de un conjunto de datos [14], el cual permita la identificación de patrones de navegación acorde a la ponderación basada en la frecuencia relativa de un evento (EF-ISF) [12], por esta razón, en este documento se presenta una propuesta para caracterizar los patrones de navegación de los estudiantes, registrados en el archivo de texto tracking.log de cursos xMOOC de la plataforma Open edX, a través del Modelo de Espacio Vectorial y EF-ISF.

Este artículo está organizado de la siguiente manera: la sección 2, describe la estructura general del archivo tracking.log; la sección 3, caracterización del conjunto de datos y un ejemplo; la sección 4, describe los pasos generales de algoritmos para el descubrimiento de patrones de navegación y la sección 5, conclusiones y trabajo futuro.

## **2 Tracking.log de la plataforma Open edX**

Log es un registro en un archivo o medio de almacenamiento; cada vez que se realiza alguna actividad en el Sistema de Gestión de Aprendizaje (Learning Management System - LMS), se crea un historial o registro en un archivo de texto llamado tracking.log. Las actividades pueden ser las siguientes [13]:

1. Hacer clic en un enlace.
2. Ver o descargar un video.
3. Leer o descargar documentos, pdf y ppt, entre otros.

Existen varios beneficios de almacenar registros para el Sistema de Gestión de Aprendizaje, entre los que se encuentran: monitoreo del progreso del estudiante, verificar el uso del material de aprendizaje en el LMS por parte del estudiante; comprobar el uso de LMS por los estudiantes, como han utilizado los estudiantes el material de aprendizaje y el foro en LMS y consultas personalizadas, responder a preguntas a partir de la información almacenada en el archivo de texto [13].

La entrada de registro, es una estructura de datos que almacena pocos campos, según el tipo de entrada o tipo de actividad, en el que la actividad también se conoce como evento. La entrada de registro es clave - valor (key - value) para los campos y sus valores, el cual se registra en formato JSON [13].

Los eventos que se guardan son generados por el servidor, navegador o dispositivo móvil, el archivo tracking.log en la plataforma Open edX se encuentra en la ruta /edx/var/log/tracking/; a pesar de que dichos eventos pueden representar acciones distintas tienen un conjunto de campos comunes, cuya estructura general se presenta a continuación [12]:

```
{ "username": "", "event_source": "", "name": "", "accept_language": "", "time": "",
"agent": "", "page": "", "host": "", "session": "", "referer": "", "context": { "user_id":, "org_id":
"", "course_id": "", "path": "" }, "ip": "", "event": "", "event_type": "" }
```

**Fig. 1.** Estructura general del archivo tracking.log de Open edX. Tomado de [12].

En la siguiente tabla se describe cada uno de los eventos.

**Tabla 1.** Descripción de campos. Tomado de [15].

<b>Campo</b>	<b>Tipo</b>	<b>Descripción</b>
accept_language	string	Determina el lenguaje por defecto del navegador.
agent	string	Reconoce al principal agente de usuario que inicia el evento.
context	object	Tiene un conjunto de sub campos que describen información del evento, los sub campos pueden ser comunes para todos los eventos como se muestra en la Tabla 2; sin embargo, pueden contener campos adicionales, ver Tabla 3.
event	object	Reconoce los campos específicos de cada evento iniciado, los campos pueden variar dependiendo del tipo de evento, los cuales pueden ser: de navegación y de inscripción, entre otros.
event_source	string	Determina el origen que genero el evento, los cuales pueden ser: "browser", "mobile", "server" y "task".
event_type	string	Identifica el tipo de evento registrado acorde al campo de "event_source".
host	string	Registra el sitio visitado por el usuario.
ip	string	La dirección "ip" del usuario cuando genera el evento.
name	string	Registra el tipo de evento lanzado.

page	string	Registra la URL de la página visitada al momento de generar el evento.
referer	string	Registra la cabecera HTTP del cual llego el evento.
session	string	Registra una cadena de 32 caracteres de la sesión del usuario.

**Tabla 2.** Campos del evento context. Tomado de [15].

Campo	Tipo	Descripción
course_id	string	Registra el curso en el que se generó el evento.
org_id	string	Registra la organización que oferta el curso.
path	string	identifica la URL que genera el evento.
user_id	number	Registra el usuario que género el evento.

**Tabla 3.** Sub campos adicionales del evento context. Tomado de [15].

Campo	Tipo	Descripción
course_user_tags	object	Registra la clave y el valor de la tabla user_api_usercoursetag para el usuario en el curso.
module	object	Registra información de los componentes que se involucran en el evento generado por el servidor.

### 3 Caracterización de los patrones de navegación con el Modelo de espacio vectorial y EF-ISF

Teniendo en un archivo de texto plano los eventos de navegación de los estudiantes, se procede a construir la Matriz de ponderación basada en la frecuencia relativa de un evento (Matriz EF-ISF), a través del Modelo de Espacio Vectorial, el cual permite encontrar la relación de los eventos de navegación de los estudiantes en un curso MOOC. Para dar un valor a los eventos de navegación se utiliza, la ponderación basada en la frecuencia relativa de un evento (EF-ISF), con lo que se otorga mayor relevancia a los eventos menos frecuentes en el conjunto de estudiantes; pero a su vez más frecuentes en cada estudiante, como se observa en la siguiente ecuación [12]:

$$w_{ij} = (f_{ij}) \left( \log \frac{N}{n_j} \right) \quad (1)$$

donde  $w_{ij}$  representa el peso del evento  $i$ ésimo del estudiante ( $\vec{s}_j$ ) [12].

La Matriz EF-ISF, es una matriz dispersa de pesos dispuestos en  $m$  eventos (filas) y  $n$  estudiantes (columnas), en el que elemento  $ij$  de la Matriz es denotado por  $w_{ij}$ , que corresponde al peso del evento  $i$  en el estudiante  $j$ , este peso es calculado de acuerdo a la ecuación (1) [12], la cual se muestra a continuación [16]:

$$\text{Matriz EF - ISF} = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1j} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2j} & \dots & w_{2n} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ w_{i1} & w_{i2} & \dots & w_{ij} & \dots & w_{in} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ w_{m1} & w_{m2} & \dots & w_{mj} & \dots & w_{mn} \end{pmatrix} \quad (2)$$

Al procesar los eventos de navegación con la ecuación (1), en la Matriz EF-ISF, se guarda la caracterización de los patrones de navegación de los estudiantes, acorde a la ponderación basada en la frecuencia relativa de un evento.

Ahora se muestra un ejemplo del proceso de construcción de la Matriz EF-ISF, a partir de dos eventos generados por un estudiante del curso de Astronomía Cotidiana del Grupo B, del año 2017 primer periodo académico de la Universidad del Cauca.

Los eventos son obtenidos del archivo tracking.log de la plataforma Selene, de ellos se tomarán algunos campos y al estudiante se lo llamara anónimo, esto se muestra a continuación:

#### Primer evento

**username:** anónimo

**name:** pause\_video

**time:** 2017-05-19T03:23:26.966429+00:00

**referer:** <http://selene.unicauca.edu.co/courses/course-v1:Unicauca+AstronomiaCotidianaGrupoB+2017-I/courseware/9ee2d4e6ba4f4c8cb5a1aea3b66220a8/83d11edf15c446a5be18be0014144fcb/>

**event:** {"code": "0xIv1RoSXNk", "id": "721d0f414ed44cdea917313b29848d48", "currentTime": 0.793626168800354}

#### Segundo evento

**username:** anónimo

**name:** load\_video

**time:** 2017-05-19T03:21:00.319791+00:00

**referer:** <http://selene.unicauca.edu.co/courses/course-v1:Unicauca+AstronomiaCotidianaGrupoB+2017-I/courseware/9ee2d4e6ba4f4c8cb5a1aea3b66220a8/83d11edf15c446a5be18be0014144fcb/>

**event:** {"code": "P2uUPX2y8Ks", "id": "067290f9c4374d5da91cf317e5ccb863"}

De los fragmentos del registro, se toman siete (7) campos los cuales se denominarán eventos, como se muestra a continuación:

evento1 = pause\_video,

evento2 = AstronomiaCotidianaGrupoB

evento3 = 9ee2d4e6ba4f4c8cb5a1aea3b66220a8

evento4 = 83d11edf15c446a5be18be0014144fcb

evento5 = 0xIv1RoSXNk

evento6 = load\_video  
evento7 = P2uUPX2y8Ks

Aplicando la ecuación (1), se obtiene la siguiente Matriz EF-ISF

$$\text{Matriz EF - ISF} = \begin{pmatrix} 0.3 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0.3 & 0 \\ 0 & 0.3 \\ 0 & 0.3 \end{pmatrix}$$

Con base en los resultados de la Matriz EF-ISF, se evidencia que el primer evento del estudiante anónimo difiere del segundo, en los eventos 1 y 5; el segundo evento del estudiante anónimo se diferencia del primero, en los eventos 6 y 7; pero ambos tienen valores similares de 0.3 en la Matriz, de esta forma el patrón de navegación del estudiante, acorde a la ponderación basada en la frecuencia relativa de un evento, puede ser cualquiera de las dos navegaciones siguientes:

**Navegacion1:**

AstronomiaCotidianaGrupoB -> 9ee2d4e6ba4f4c8cb5a1aea3b66220a8 ->  
83d11edf15c446a5be18be0014144fcb -> pause\_video -> 0xIv1RoSXNk

**Navegacion2:**

AstronomiaCotidianaGrupoB -> 9ee2d4e6ba4f4c8cb5a1aea3b66220a8 ->  
83d11edf15c446a5be18be0014144fcb -> load\_video -> P2uUPX2y8Ks

#### 4 Algoritmos para el descubrimiento de patrones de navegación

Para el descubrimiento de patrones de navegación en un conjunto de estudiantes, acorde a la ponderación basada en la frecuencia relativa de un evento (EF-ISF), se pueden utilizar algoritmos como: Análisis Semántico Latente (LSA), LexRank con Umbral, Máxima Cobertura y Mínima Redundancia (MCMR) y Moderador, entre otros [16].

A continuación, se muestran los pasos generales del algoritmo LexRank con Umbral [16]:

1. Representar el conjunto de estudiantes en la Matriz EF-ISF y construir la Matriz de similitudes.
2. Aplicar el umbral a la Matriz de Similitudes.
3. Distribuir la centralidad del estudiante con los que tiene alguna relación, creando la Matriz Estocástica.
4. Transformar la Matriz Estocástica a irreducible y aperiódica.
5. Encontrar la Matriz Estacionaria.

Ahora se muestran los pasos generales del algoritmo Análisis Semántico Latente (LSA) [16]:

1. Representar el conjunto de estudiantes en la Matriz EF-ISF.
2. Aplicar Descomposición de Valores Singulares (SVD) en la Matriz EF-ISF.
3. Escoger los estudiantes que más aporten al tópico de la matriz  $U$  o  $V^T$  sin repetir estudiantes.
4. Elegir los estudiantes que tengan mayor similitud con el vector centroide.

Luego se muestran los pasos generales del algoritmo Máxima Cobertura y Mínima Redundancia (MCMR) [16]:

1. Representar el conjunto de estudiantes en la Matriz EF-ISF y construir la Matriz de similitudes.
2. Utilizar el algoritmo de Optimización por Enjambre de Partículas (PSO) para elegir los mejores estudiantes.
3. Evaluar la función objetivo con los estudiantes elegidos.
4. Repetir los pasos 2 y 3 hasta el número de iteraciones.
5. Escoger los estudiantes que tienen el mayor valor de la función objetivo.

Posteriormente, se muestran los pasos generales del algoritmo Moderador, basado en la Toma de Decisiones en Grupo (GDM) con consenso suave con doble ronda retroalimentada [16]:

1. Representar el conjunto de estudiantes en la Matriz EF-ISF y construir la Matriz de similitudes.
2. El moderador aplica proceso de consenso en dos rondas, en cada ronda se aplica fase de agregación.
3. El moderador sigue al proceso de selección y aplica fase de explotación.
4. El moderador genera los patrones de navegación de la GDM con consenso suave.

## **5 Conclusiones**

Se describió una alternativa para caracterizar los patrones de navegación de los estudiantes, registrados en el archivo de texto tracking.log de xMOOC de la plataforma Open edX, a través del modelo de espacio vectorial y la ponderación basada en la frecuencia relativa de un evento.

Con la caracterización de los patrones de navegación de los estudiantes, registrados en el archivo de texto tracking.log de xMOOC de la plataforma Open edX, a través del modelo de espacio vectorial y la ponderación basada en la frecuencia relativa de un evento, se pueden determinar patrones de navegación acorde a EF-ISF.

Se describieron los pasos generales de los algoritmos: Análisis Semántico Latente (LSA), LexRank con Umbral, Máxima Cobertura y Mínima Redundancia (MCMR) y Moderador; para identificar patrones de navegación de un conjunto de estudiantes acorde a EF-ISF.

Como trabajo a futuro se plantea la implementación de la cartelización de los patrones de navegación, del archivo de texto tracking.log de xMOOC de la plataforma

Open edX, con el modelo de espacio vectorial y la ponderación basada en la frecuencia relativa de un evento. De igual forma, la implementación de los algoritmos: Análisis Semántico Latente (LSA), LexRank con Umbral, Máxima Cobertura y Mínima Redundancia (MCMR) y Moderador.

## **Agradecimientos**

Los autores agradecen el apoyo recibido por el proyecto MOOC-Maker Construction of Management Capacities of MOOCs in Higher Education (561533-EPP-1-2015-1-ESEPPKA2-CBHE-JP) financiado por la Comisión Europea a través del Programa Erasmus+ para la realización y divulgación de los resultados expuestos en el presente artículo.

De igual manera agradecer al proyecto VRI 49694 MOOCMenTES "Construcción de capacidades para la gestión de MOOC para la formación profesional, el desarrollo rural y nuevas generaciones de estudiantes rurales en el Mejoramiento de su Tránsito a la Educación Superior", co-financiado en el marco de alianzas rurales por el Ministerio de Educación Nacional de Colombia.

## **Referencias**

- [1] Siemens, G.: Connectivism: A learning theory for the digital age. *Elearnspace Everything Learning* (14), 1-8 (2004).
- [2] Guàrdia, L., Maina, M., and Sangrà, A.: MOOC design principles: A pedagogical approach from the learner's perspective. *eLearning Papers* (33), 1-6 (2013).
- [3] Hernández, S.: MOOC: Estado de la situación actual, posibilidades, retos y futuro. *SCOPEO INFORME* (2), Universidad de Salamanca (2013).
- [4] Liyanagunawardena, T., Williams, S., and Adams, A.: The impact and reach of MOOCs: a developing countries' perspective. *eLearning Papers Special Edition*, 38-46 (2014).
- [5] Sonwalkar, N.: The first adaptive MOOC: A case study on pedagogy framework and scalable cloud Architecture—Part I. In *MOOCs Forum* (1), 22-29 (2013).
- [6] Fasihuddin, H., Skinner, G., and Athauda, R.: Towards adaptive open learning environments: Evaluating the precision of identifying learning styles by tracking learners' behaviours. *Education and Information Technologies* 22(3), 807-825 (2017).
- [7] Liyanagunawardena, T., Adams, A., and Williams, S.: MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open and Distributed Learning* 14(3), 202-227 (2013).
- [8] Kizilcec, R.: Collaborative Learning in Geographically Distributed and In-person Groups. In *AIED 2013 Workshops Proceedings Volume* (67), 67-74 (2013).
- [9] McAuley, A., Stewart, B., Siemens, G., and Cormier, D.: The MOOC model for digital practice. *Digital ways of knowing and learning*, University of Prince Edward Island (2010).
- [10] Guo, P., and Reinecke, K.: Demographic differences in how students navigate through MOOCs. *Proceedings of the first ACM conference on Learning@ scale conference*, 21-30 (2014).



- [11] Jaramillo, D., Solarte, M., and Ramírez, G.: Follow-Up of Learning Activities in Open edX: A Case Study at the University of Cauca. In: Delgado Kloos, C., Jermann, P., Pérez-Sanagustín, M., Seaton, D. (eds.) *Digital Education: Out to the World and Back to the Campus. EMOOCs 2017, LNCS*, vol. 10254, pp. 217-222. Springer, Cham (2017).
- [12] Anacona, F., Solarte, M., and Ramírez, G.: Descubrimiento de patrones de navegación en Open edX—una aproximación arquitectónica. *Ingeniería e Innovación* 5(1), 43-50 (2017).
- [13] Yadav, R.: *Understanding Logs in edX for Monitoring Student Progress*. Doctoral Thesis, Indian Institute of Technology Bombay (2014).
- [14] Burbano, M., Anacona, F., Solarte, M., and Ramírez, G.: Informe sobre tecnologías Web Semántica y Social en cursos MOOC. *MOOC-Maker Construction of Management Capacities of MOOCs in Higher Education WPD1.14*, 1-17 (2016).
- [15] Palta, R., and Vásquez, J.: Descubrimiento de patrones de interacción en cursos MOOC en entornos ONLINE: Un enfoque utilizando Minería de Procesos. Caso de Estudio: “Curso de la metodología DICREVOA en Open edX”. Monografía Facultad de Ingeniería, Universidad de Cuenca (2016).
- [16] Anacona, F., Cobos, C., and Mendoza, M.: Algoritmo para generación automática de resúmenes extractivos genéricos de múltiples documentos basado en consensos. Monografía Departamento de Sistemas, Universidad del Cauca (2015).