

## **ScientoPy for MOOCs: A Scientometric Review**

Luis Cruz-Ordonez, Mario Solarte, Gustavo Ramirez-Gonzalez

Departamento de Telemática, Universidad del Cauca,  
luisalejo@unicauca.edu.co

**Resumen.** Los MOOCs están siendo foco de investigación en diferentes áreas de la ingeniería y la educación, aumentando el nivel de calidad de la educación en entornos virtuales. Este documento presenta una revisión sistemática de la bibliografía recolectada de los últimos años relacionada directamente con la temática MOOC, usando la herramienta ScientoPy. La revisión aporta a los estados del arte para los investigadores que tienen como temática los MOOCs. Esta consiste en un procesamiento de las publicaciones a partir de las bases de datos de Scopus y WoS con una herramienta de procesamiento de datos. Los resultados obtenidos permiten identificar los principales focos de investigación, palabras claves, investigaciones relacionadas y los lugares e investigadores más destacados en la actualidad.

**Abstract.** MOOCs are being a focus of research in different areas of engineering and education, increasing the level of quality of education in virtual environments. This document presents a systematic review of the literature collected from the last years directly related to the MOOC topic, using the ScientoPy tool. The review contributes to the state of the art for researchers whose subject matter is MOOCs. It consists of processing the publications from the Scopus and WoS databases with a data processing tool. The results obtained allow to identify the main research focuses, keywords, related researches and the most outstanding places and researchers at present.

**Palabras clave:** Education, MOOC, ScientoPy, Scientometry.

### **1 Introducción**

Al pasar de los años, los profesores e investigadores han estado cada vez más interesados en el aporte de la tecnología para la transformación de la educación [1] [2]. Muchos de estos aportes han sido consolidados bajo el término E-Learning, el cual hace referencia al apoyo de la informática y las redes a los procesos formativos no necesariamente presenciales. Una vertiente significativa de E-Learning han sido los MOOC (Massive Online Open Courses), los cuales han recibido una importante acogida en los últimos años [3]. Desde el 2012, la oferta de MOOC ha venido creciendo a través de diferentes plataformas que soportan educación en línea en ambientes de masividad. La estrategia MOOC abre nuevas posibilidades de investigación en diferentes áreas, como, por ejemplo, la educación, la pedagogía y ciencias de la computación entre otras [4]. Dicho lo anterior, es imperativo poder identificar a través de análisis de publicaciones cuáles han sido las trayectorias de la investigación en MOOC y cuáles pueden ser las tendencias en los próximos años. Por lo tanto, como objetivo central de este artículo se plantea la siguiente pregunta de investigación: ¿Cuáles han sido las áreas de publicación más frecuentes en la literatura MOOC desde su aparición?

## **2 Marco Teórico.**

### **2.1 MOOC (Massive Online Open Courses)**

El término MOOC es acuñado por primera vez por Dave Cormier en 2008 [5] para referirse a los cursos ofrecidos en línea, abiertos y masivos. Los MOOC fueron utilizados por primera vez por George Siemens y Stephen Downes en el 2008 y desde ese entonces se han dado a conocer como una oportunidad de aprendizaje extra para los alumnos. ¿Quién puede acceder a un MOOC? Básicamente, un estudiante que tenga conexión a internet y se una a uno de los cursos ofrecidos en las plataformas MOOC. Es por eso que su nombre define sus características. Masivo (Massive), referido a la cantidad de estudiantes registrados por MOOC, que puede oscilar desde los cientos de estudiantes hasta cerca de los 150.000 estudiantes [6] Abierto (Open), definido así por la libertad con la que los estudiantes pueden registrarse a ellos independientemente de su localización, edad, nivel de aprendizaje, cultura o cualquier otro factor que los distinga entre ellos. El término de abierto también puede hacer referencia los recursos abiertos utilizados en cada MOOC por los profesores [7], como diapositivas, videos, apuntes, imágenes, etc. El otro término es Online (En línea), relacionado directamente con la accesibilidad de los cursos, que se resume al hecho de tener como mínimo una conexión a internet para poder acceder al mismo y de este modo establecer una relación síncrona de interacción y asíncrona entre los participantes del curso [8]. Por último, el término Curso (Course) permite organizar los anteriores términos en un plan de estudios estructurado, que no sólo consiste en escuchar una clase pregrabada sino en el seguimiento detallado de la brecha Estudiante-Profesor para alcanzar el objetivo principal que es el aprendizaje.

Los MOOCs enfrentan hoy, problemas como la deserción, la rigidez pedagógica, el seguimiento estudiantil que entre otras problemáticas son la motivación para los investigadores que buscan aportar en la evolución de la educación virtual

### **2.2 Bibliometría**

Es el estudio de medición y análisis de la literatura científica mediante la medición del impacto de la innovación y la comprensión de la relevancia de estas citas científicas para esta innovación [3]. Este estudio hace uso de toda la producción científica de una temática específica para brindarles a los investigadores un apoyo para la realización de su estado del arte o trabajos relacionados en sus investigaciones. Gracias a estas técnicas han sido creadas herramientas software que proveen un procesamiento de datos que facilitan las revisiones bibliográficas.

### **2.3 ScientoPy**

ScientoPy es un script de Python que genera y reporta automáticamente los tópicos principales (basados en las palabras clave de los autores), los autores y los países e instituciones de las publicaciones, junto con los documentos relacionados. Esta síntesis automática de datos evita el sesgo potencial como en los estudios individuales.

Se ha comprobado que los resultados obtenidos en diferentes procesamientos de datos han servido de mucha ayuda a los investigadores [9]. Sin embargo, el análisis del nombre del autor (como la lista principal del autor) tiene un riesgo de sesgo en los estudios debido a las posibles similitudes en los nombres. Los autores de esta revisión conocen y advierten sobre esta posibilidad de similitud entre los nombres de autor de los documentos, que es parte de la limitación de cualquier estudio de Bibliometría; por lo tanto, en este momento no todos los autores y bases de datos tienen un identificador de autor único, como el ORCID, asociado a todas las entradas.

Este script tiene la capacidad de:

- Leer archivos CSV de bases de datos como Web of Science y Scopus.
- Filtrar publicaciones por tipo de documento.
- Encontrar y eliminar documentos duplicados.
- Ilustrar por medio de gráficas el histórico de los principales tópicos (Palabras claves, autores y países).
- Ilustrar por medio de gráficas el historial de elementos seleccionados dentro de un tema.
- Encontrar temas de tendencia usando la tasa de crecimiento promedio superior (AGR).
- Calcular el h-index para autores y países

### **3 Pre-Procesamiento**

#### **3.1 Data Set**

En esta sección se muestra el DataSet con el que la herramienta realiza el procesamiento de datos. Es importante una buena definición de estos datos como primer paso de este proceso, pues es la fuente de procesamiento de la información. Por esta razón, se decidió realizar la búsqueda de publicaciones científicas tanto en la base de datos de Scopus [10] como de la de WoS (Web of Science) [11]. A continuación, se muestra el criterio de búsqueda utilizado para las bases de datos:

**TITLE-ABS-KEY (“MOOC\*” OR “Massive Open Online Course\*” OR “Massive Online Open Course\*”).**

Debido a que el término MOOC es relativamente nuevo, es muy probable que no se encuentre un gran número de publicaciones como en otras temáticas más conocidas, sin embargo, se quiere agotar cualquier término con el que se puede también encontrar en las publicaciones. El asterisco mostrado en la búsqueda hace referencia a buscar no solamente el término MOOC sino también aquellos que van seguidos de otros caracteres, por ejemplo: MOOCS, MOOC-Education, MOOC-Learning, etc. Se ha utilizado el conector “OR” para también buscar aquellas publicaciones en donde sus autores no han utilizado la sigla MOOC sino las palabras que conforman el término y además se ha considerado algún cambio en el orden de las palabras que puede variar según algunos países. Estas consideraciones ayudarán a tener un campo más amplio de datos. Estos datos fueron extraídos el 15 de Julio de 2018 y la Tabla 1 muestra el número total de documentos encontrados en ambas bases de datos según el tipo de documento.

**Tabla 1.** Documentos encontrados.

| Source | Conference Paper | Article | Review Article | in Press | Total |
|--------|------------------|---------|----------------|----------|-------|
| WoS    | 0                | 1029    | 29             | 0        | 1058  |
| Scopus | 1977             | 1370    | 83             | 54       | 3484  |

### 3.2 Duplicados y Simplificación

Cabe la posibilidad de que, en el total de documentos mostrado en la subsección anterior, haya duplicados en las bases de datos, o que se encuentre el mismo documento en ambas bases de datos. Para esto, la herramienta Scientopy aplica una técnica de Pre-procesamiento de datos que elimina todos los artículos duplicados. La tabla 2 muestra que, en total, fueron encontradas 704 muestras duplicadas. Otra de las funciones de esta primera técnica es la simplificación de nombres de autores. Un problema común que se ha percibido en las publicaciones es la inconsistencia de los nombres y apellidos de los autores. A través del preProces, ScientoPy es capaz de simplificar caracteres especiales en los nombres, acentos y abreviaturas utilizadas.

Las abreviaturas o simplificación de nombres también se perciben en los demás tópicos como países, keywords, instituciones, por lo que se recomienda definir a través de un solo término cada país para hacer un buen procesamiento posteriormente.

**Tabla 2.** Documentos finales.

| Source | Conference Paper | Article | Review Article | in Press | Total |
|--------|------------------|---------|----------------|----------|-------|
| WoS    | 0                | 1027    | 29             | 0        | 1056  |
| Scopus | 1977             | 729     | 53             | 53       | 2782  |

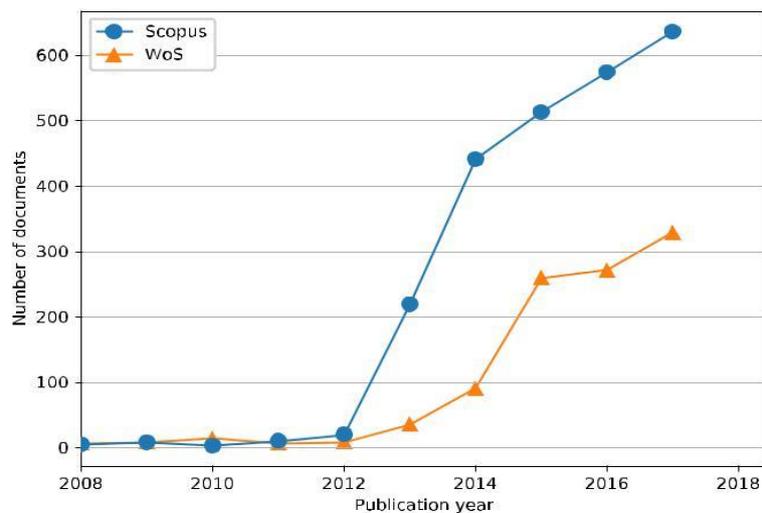
## 4 Procesamiento

En esta sección se presentan los análisis de procesamientos realizados que permiten percibir la situación actual de la investigación en MOOCs y las tendencias previstas

### 4.1 Crecimiento de publicaciones

Las investigaciones de los MOOCs comenzaron a darse a conocer desde mediados de 2012 justo después del lanzamiento del primer MOOC oficial (Connectivism and Connective Knowledge) [12]. De esta manera, es razonable comenzar el análisis de las publicaciones de MOOCs a partir del 2012 hasta el año 2017, pues se considera un año en el que ya se han incluido todas las publicaciones presentes en las bases de datos. Esto se aplicará en todos los criterios de análisis a excepción del análisis de crecimiento de publicaciones que se procesará desde el año 2008 (año donde fue acuñado el término MOOC).

La Figura 1 muestra dicho crecimiento para ambas bases de datos y es notable la fuerza que está tomando la investigación en el campo de los MOOCs. Esto impulsa a los investigadores a seguir apostando a aportes a nivel de tecnología y pedagogía en la enseñanza online. Se puede notar claramente que los aportes de publicaciones comenzaron a darse a partir del 2012 como se esperaba que fuera por lo anteriormente expuesto.



**Figura 1.** Crecimiento de publicaciones en MOOC

#### **4.2 Análisis de palabras claves definidas por autores**

Como se ha descrito anteriormente, ScientoPy tiene la capacidad de brindar información no solamente de la totalidad en número o porcentaje de documentos, sino también de algunos tópicos o argumentos de búsqueda de las publicaciones como las keywords que aparecen en cada publicación. Hay que reconocer que como investigadores las keywords y las citas generan enlaces con otras publicaciones relacionadas.

En esta sección se analizarán las principales palabras claves definidas en la temática MOOC. Para ello se ha recomendado unir por grupos los términos similares como por ejemplo educación y enseñanza, o parejas como “Mobile Learning” “M-Learning”. Este tipo de ejercicios manuales ayudan a organizar mejor la información y a evitar duplicidad en los términos.

Según la figura 2 la palabra clave más usada es “Education”, era de esperarse que el procesamiento de datos muestre este término global y directamente relacionado con la temática como el principal. Después se presentan palabras claves importantes que sirven como ayuda para saber cuáles elegir para documentos en nuestras investigaciones. Aunque en la figura sólo se muestran las primeras 10, ScientoPy permite visualizar cualquier cantidad de keywords deseadas.

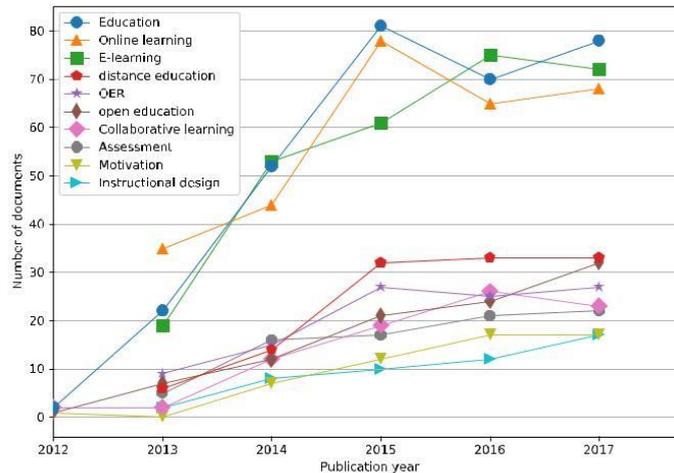


Figura 2. Principales palabras claves de autores

Las palabras claves se pueden categorizar según diferentes criterios, en esta sección se han seleccionado 2 criterios principalmente. El primero, según las tecnologías involucradas en investigaciones MOOC y el segundo, según plataformas encontradas en investigaciones MOOC. La información de los documentos según estas categorías definidas se mostrará a continuación a través de diagramas de barras, dado que los diagramas de barras permiten visualizar un acumulado en los años del uso de los términos.

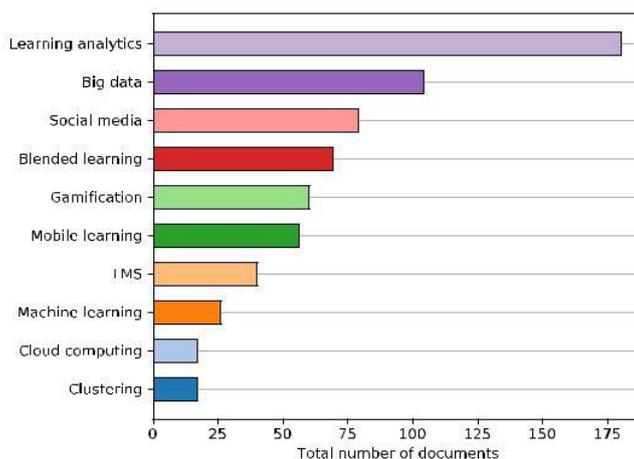
Para investigadores en MOOC que hacen parte de la ingeniería o ciencias de la computación, es de gran interés la figura 3 que muestra la gráfica de las tecnologías aplicadas en los MOOC. También para efectos de analizar cuáles de las plataformas MOOC han venido siendo las más representativas en las publicaciones, se ha hecho el procesamiento, y los resultados se muestran en la figura 4.

### 4.3 Análisis de países y Autores

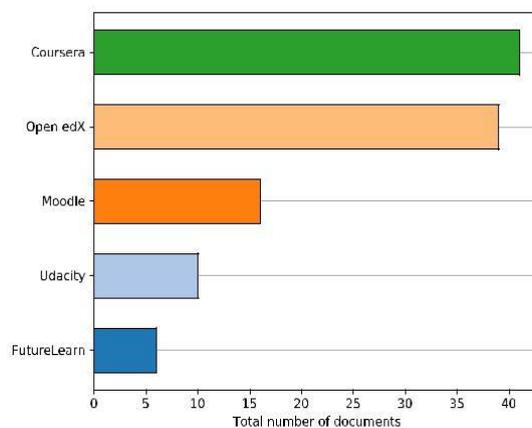
El análisis de países permite determinar cuáles son aquellos países con más influencia en las temáticas, esta información es extraída según la procedencia de los autores o el país al que pertenecen las universidades que bien se especifica en cada publicación. La figura 5 muestra otra forma de representar gráficamente por la herramienta ScientoPy denominada WordCloud, este gráfico es posible obtenerse gracias a que Python posee una librería que permite dibujar aleatoriamente una nube de palabras dotando de mayor tamaño a las más significativas. En la figura se puede percibir que entre los países con más publicaciones en MOOCs está Estados Unidos, España, China y Reino Unido.

Para los investigadores también es muy útil saber los autores más representativos en sus aportes para citarlos en sus aportes o para tener un referente de trabajos relacionados. la tabla 3 muestra el top 30 de los autores más sobresalientes por sus aportes en MOOCs. La tabla también especifica una tasa de crecimiento (AGR) y el

h-index de cada uno. El hecho de citar autores importantes en nuestras investigaciones facilita la divulgación de estas, así como el reconocimiento científico. Entre los autores principales se encuentre Christoph Meinel [13] de la Universidad de Potsdam, Carlos Alario Hoyos [14] de la Universidad Carlos III de Madrid y Mar Pérez [14] de la Universidad Pontificia Católica de Chile.



**Figura 3.** Principales palabras claves de tecnologías



**Figura 4.** Principales palabras claves de plataformas



Figura 5. Países

Tabla 3. Principales autores

| Pos | Author               | Total AGR | h-index | Pos | Author | Total AGR         | h-index |
|-----|----------------------|-----------|---------|-----|--------|-------------------|---------|
| 1   | Meinel C             | 35        | 4.0     | 7   | 16     | Yang D            | 14 -3.0 |
| 2   | Alario-Hoyos C       | 26        | 2.0     | 9   | 17     | Fox A             | 14 -2.5 |
| 3   | Perez-Sanagustin M   | 26        | 4.5     | 9   | 18     | Garcia-Penalvo FJ | 14 0.0  |
| 4   | Kloos CD             | 24        | 0.0     | 6   | 19     | Wang J            | 14 -1.0 |
| 5   | Munoz-Merino PJ      | 23        | 1.0     | 7   | 20     | Reich J           | 13 -1.5 |
| 6   | Dillenbourg P        | 21        | 1.5     | 5   | 21     | Kim J             | 13 -1.5 |
| 7   | Rose CP              | 20        | -2.5    | 8   | 22     | Lujan-Mora S      | 13 0.5  |
| 8   | Wang Y               | 18        | 1.0     | 3   | 23     | Zhang Y           | 13 1.0  |
| 9   | Renz J               | 17        | 2.0     | 5   | 24     | Kalz M            | 13 4.0  |
| 10  | Marchetti F          | 16        | -1.5    | 6   | 25     | Zhang J           | 13 1.5  |
| 11  | Staubitz T           | 16        | 2.0     | 5   | 26     | Kizilcec RF       | 12 0.5  |
| 12  | Ruiperez-Valiente JA | 16        | 1.5     | 4   | 27     | Pritchard DE      | 12 0.5  |
| 13  | Pampaloni G          | 15        | -1.5    | 6   | 28     | Li X              | 12 3.0  |
| 14  | Zacchini S           | 15        | -1.5    | 6   | 29     | Gasevic D         | 11 -1.0 |
| 15  | White S              | 15        | -0.5    | 4   | 30     | Sein-Echaluce ML  | 11 0.0  |

#### 4.4 Análisis de instituciones

Cada autor que figura en las publicaciones tiene asociada una institución. ScientoPy también hace uso de esta información en su procesamiento. El poder identificar cuáles son las instituciones más representativas en las temáticas es una ayuda a los investigadores para seleccionar estancias de investigación o para unirse a alguno de sus programas académicos o proyectos de investigación. Este análisis le da prestigio a las instituciones e impulsa a otras a seguir publicando para alcanzar un puesto sobresaliente. La figura 6 muestra una gráfica con las principales instituciones que han publicado en temática MOOC. El tipo de grafica utilizada para este análisis, ScientoPy la denomina paramétrica. En ella se muestran dos gráficas, en la primera el número acumulativo de documentos vs el año de publicación y en la otra la AGR vs el h-index. De esta manera es más fácil visualizar mejor el crecimiento de cada institución y su comparación con los demás.

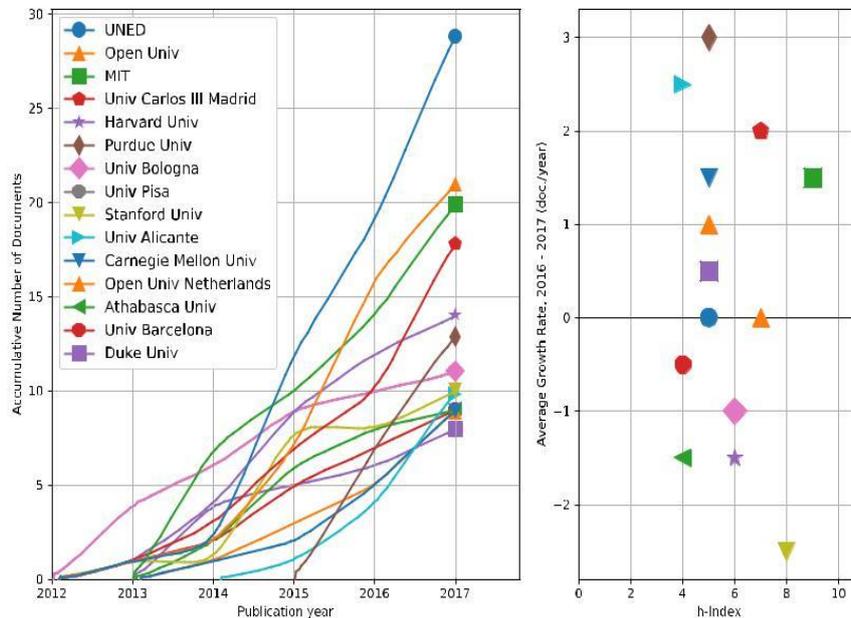


Figura 6. Top 10 de instituciones

## 5 Conclusiones

Los MOOC son una temática creciente que ha generado impacto de investigación en los últimos años principalmente en temas de pedagogía y tecnología. Gracias a la herramienta ScientoPy se ha podido lograr el objetivo a través de un análisis bibliométrico a partir de las bases de datos de WoS y Scopus.

El país con más aportes de investigación en MOOCs es Estados Unidos, la palabra clave definida por autor más usada es “Education”, la temática en tecnología más explorada es “Learning Analytics” y los Autores que más contribuciones han hecho son de Estados Unidos y España. Los MOOC tienen una tendencia importante en todos los campos, y se espera que la curva de investigación siga creciendo. Los análisis bibliométricos sí aportan a los estados de arte de nuevos trabajos, además de reconocer los diferentes aportes por parte de autores, instituciones y países.

**Agradecimientos.** El presente artículo ha contado con el apoyo del proyecto VRI 49694 MOOC-MenTES “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”, cofinanciado en el marco de alianzas rurales por el Ministerio de Educación Nacional de Colombia. Los autores agradecen además 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.

## **Referencias**

- [1] Hew, K. F. y Brush, T. Integrating technology into K-12 teaching and learning: current knowledge gaps and recommendations for future research. *Educational Technology Research and Development* 55, 223-252. issn: 1556-6501 (jun. de 2007).
- [2] Salinas, A., Nussbaum, M., Herrera, O., Solarte, M. y Aldunate, R. Factors affecting the adoption of information and communication technologies in teaching. *Education and Information Technologies* 22, 2175-2196 (2017).
- [3] Leydesdor, L., Milojevic, S. y col. *Scientometrics* (2015).
- [4] Jaramillo-Morillo, D., Sarasty, M. S., González, G. R. y Pérez-Sanagustín, M. Follow-Up of Learning Activities in Open edX: A Case Study at the University of Cauca en *European Conference on Massive Open Online Courses* (2017), 217-222.
- [5] Cormier, D., Stewart, B., Siemens, G. y McAuley, A. What is a MOOC? (2010).
- [6] Allen, I. E. y Seaman, J. Changing course: ten years of tracking online education in the United States. *Nursing standard (Royal College of Nursing (Great Britain))*: 1987) 26, 47. issn: 0029-6570 (2013).
- [7] Anderson, T. y McGreal, R. Disruptive Pedagogies and Technologies in Universities. *Journal of Educational Technology & Society* 15, 380-389. issn: 11763647, 14364522 (2012).
- [8] Stephen, B. Back to the Future with MOOCs. *Icicte* 77, 237-246 (2013).
- [9] Ruiz-rosero, J., Id, G. R.-g. y Williams, J. M. SS symmetry ScientoPy example for Internet of Things: A Scientometric Review, 1-34 (2017).
- [10] Burnham, J. F. Scopus database: a review. *Biomedical digital libraries* 3, 1 (2006).
- [11] Reuters, T. Web of Science. (2012).
- [12] Siemens, G. MOOCs for the win! *ElearnSpace* 2012.
- [13] Meinel, C. y Willems, C. *openHPI: the MOOC offer at Hasso Plattner Institute* (Universitätsverlag Potsdam, 2013).
- [14] Alario-Hoyos, C., Pérez-Sanagustín, M., Delgado-Kloos, C., Muñoz-Organero, M., Rodríguez-de-las-Heras, A. y col. Analysing the impact of built-in and external social tools in a MOOC on educational technologies en *European Conference on Technology Enhanced Learning* (2013), 5-18.