Análisis de la oferta educativa en el ámbito de los MOOCs

D. G. Reina, M. R. Martinez-Torres, S. L. Toral, F. Barrero
Universidad de Sevilla

***Abstract*—Los MOOCS, acrónimo del término Massive Open Online Courses, han experimentado un extraordinario auge en los últimos años y se postulan como una de las herramientas más populares y eficientes en el ámbito educativo universitario. Múltiples plataformas educativas han proliferado a través de internet y la oferta educativa crece día a día. El análisis de esta amplia oferta educativa requiere de nuevas metodologías capaces de analizar de manera automática los contenidos ofertados. Este trabajo propone la utilización de técnicas de análisis semántico latente para analizar la oferta educativa actual dentro del área de Computer Science, que es una de las más demandas en el ámbito de los MOOCs. Los resultados obtenidos aportan información acerca de las temáticas emergentes en el área de Computer Science y pueden ser útiles para actualizar los contenidos ofertados en los cursos tradicionales.**

***Index Terms*—Open courses, MOOCs, oferta educativa, análisis semántico latente.**

I. Resumen

Según McAuley et al. [1], un MOOC es un curso en línea que da la opción de registro libre y abierto, un plan de estudios público y resultados sin plazos definidos. Los MOOCs se basan en la participación de los estudiantes, quienes organizan su participación según sus objetivos de aprendizaje, sus conocimientos y habilidades previas y sus intereses comunes. Su potencia radica en la capacidad de atracción de un elevado número de estudiantes en todo el mundo que auto-organizan su aprendizaje, y en las conexiones que, a través de los cursos y sus herramientas sociales asociadas, pueden conseguir con otros estudiantes (Daradoumis et al., 2013). Los cursos son abiertos, permiten un acceso a muy bajo coste o gratuito, y cada estudiante puede elegir su nivel de involucración en el curso. En este sentido, los MOOCs constituyen una forma de socialización del conocimiento hacia colectivos que por motivos económicos, geográficos o formales se verían excluidos de esta formación.

Por otro lado, el amplio público al que estos cursos van dirigidos permite maximizar las posibilidades del efecto de "cola larga", de modo que es posible encontrar una amplia audiencia para cursos muy específicos que difícilmente podrían impartirse a nivel local. Es precisamente esta especificidad de la oferta educativa de los MOOCs la que permite utilizar sus contenidos para realizar un análisis de las tendencias y contenidos emergentes. Para ello se utilizará la plataforma Coursera, que junto a EdX, Miríada X y Udacity constituyen las cuatro plataformas más importantes que en la actualidad ofertan cursos abiertos online masivos (Severance, 2012). En particular, se analizarán los cursos catalogados dentro de las áreas *Computer Science, Artificial Intelligence* y *Computer Science, Software Engineering*. Para cada uno de ellos se extraerá su título y resumen, así como los contenidos específicos ofertados. Sobre la colección de documentos obtenidos de esta manera se llevará a cabo un análisis semántico latente o LDA (Latent Semanatic Analyis) con el fin de atraer los principales tópicos o temas tratados por los cursos ofertados. Las principales contribuciones del trabajo son: (i) realización de un análisis de contenido sobre los MOOCs ofertados en Coursera (ii) identificación de los principales temas de los cursos ofertados y (iii) análisis de las tendencias formativas emergentes en el campo de Computer Science. Los resultados obtenidos tienen implicaciones importantes tanto para las plataformas de enseñanza de cursos online masivos como para los planes de estudio de las enseñanzas Universitarias. En el primer caso, proporciona un metodología válida de análisis de las principales áreas de interés, lo que permite ampliar la oferta en esas áreas específicas. Con respecto a la enseñanza reglada tradicional, proporciona interesantes resultados acerca de los huecos formativa que existen en los planes de estudio tradicionales, mucho menos dinámicos que las plataformas online a la hora de incorporar nuevos contenidos formativos.

El artículo se estructura de la siguiente forma: la sección siguiente detalla algunos trabajos previos sobre el éxito de los MOOCs así como las limitaciones detectadas. La sección III introduce la metodología utilizada en este estudio. La sección IV contiene los resultados obtenidos para la plataforma Coursera y, finalmente, la sección V concluye el trabajo.

II. Trabajos Previos

El rápido auge de los MOOC ha planteado varias cuestiones de investigación, muchas de ellas relacionadas con su desarrollo y su encaje dentro del modelo tradicional de enseñanza de la educación superior. Los MOOCs se basan en las mismas características que han popularizado el aprendizaje online: incrementar la disponibilidad de experiencias de aprendizaje salvando las barreras espacio temporales del aprendizaje cara a cara, mejorar la difusión del contenido formativo de manera más eficiente y facilitar a los instructores la gestión de un elevado número de estudiantes sin bajar la calidad formativa. No obstante, los MOOCs representan un campo diferenciado dentro de los denominados OER (Open educational resources) y comienzan a ser estudiados desde una perspectiva investigadora.

Uno de los motivos aducidos como causas de su éxito es la alta multidisciplinariedad de los cursos ofertados [5]. Este es el caso de los cursos ofertados por la plataforma Coursera, que incluye una amplia gama de cursos organizados por áreas de conocimiento y que en algunos casos llegan a cubrir más de un área. Estos cursos son mucho más flexibles que los ofertados por la educación reglada. Son impartidos por profesores y especialistas

pertenecientes a prestigiosas Universidades por todo el mundo (la mayoría por el momento americanas), que previamente han establecido un acuerdo de colaboración con dicha plataforma. No importa lo específico o multidisciplinar que sea el curso, la economía de escala que proporciona una potencial audiencia de miles de estudiantes posibilita el efecto de cola larga, de modo que cursos que a nivel local no conseguirían un mínimo número de estudiantes pueden ser viables en forma de MOOC [1], [6].

Otra de las características que distingue a los MOOCs de otras posibilidades educativas online es que funcionan de una manera muy parecida a una clase real, con una fecha de comienzo y final, lecturas, asignaciones de trabajos, test y exámenes y evaluaciones de los mismos bien mediante una herramienta automática o mediante una evaluación por pares [7]. Frente a otros esquemas de aprendizaje online donde el usuario sigue su propio ritmo, los MOOCs definen una planificación en tiempo real, con fechas de entrega definidas y penalizaciones en caso de incumplirlas.

Los MOOCs también prestan especial atención a las interacciones entre los usuarios, normalmente a través de foros. A través de ellos, los usuarios comentan temas, ejercicios, proponen nuevos tópicos de discusión o incluso interactúan con los instructores para resolver algunas cuestiones relacionadas con la asignación de trabajos. Muchos cursos también incluyen la posibilidad de evaluación por pares. El elevado número de alumnos hace inviable una evaluación personalizada, por lo que lo que habitualmente las evaluaciones son, o bien automáticas, mediante una herramienta software que compruebe las respuestas a un test o evalúe el resultado de un programa, o mediante un sistema de evaluación por pares donde son los mismos estudiantes matriculados en el curso los que evalúan los trabajos enviados por otros estudiantes, siguiendo unas determinadas pautas previamente indicadas en el curso. La ventaja de la evaluación por pares es que permite evaluar trabajos abiertos, donde los alumnos pueden desarrollar más su creatividad [8], aunque a costa de pedirles una mayor carga de trabajo. Actualmente, varias de las plataformas de MOOCs trabajan en los denominados AES (Automated Essay Scoring), que mediante modelos estadísticos permiten predecir la puntuación de trabajos escritos por los estudiantes [9].

Finalmente, los MOOCs están desempeñado un importante papel en la socialización del conocimiento. Su carácter abierto, a distancia y, en muchos casos, gratuito, permite que muchos colectivos tradicionalmente excluidos de una educación superior tengan acceso a ella. Al mismo tiempo, constituyen una herramienta de promoción de las propias Universidades, departamentos e instructores que ofertan el curso, que pueden darse a conocer de una manera muy efectiva entre un amplio público interesado en el tema. Este contacto a través de los MOOC ofertados permite atraer estudiantes de todo el mundo e incluso futuros investigadores, que mediante la realización del curso pueden adquirir nuevas inquietudes sobre determinados temas y solicitar estancias o becas de investigación en los centros responsables.

Aunque estas plataformas cuentan con cientos de miles de estudiantes, todavía tienen importante retos que resolver, como la evaluación, los altos índices de abandono [10], o su estructura pedagógica [11]. Si bien se cuentan por miles los alumnos matriculados en los cursos, sólo un pequeño porcentaje lo completa en su totalidad. También se plantean dudas acerca de los sistemas de evaluación o la detección de plagio.

III. METODOLOGÍA

El procesamiento del lenguaje natural es un conjunto de técnicas informáticas y lingüísticas que analiza el lenguaje humano mediante algoritmos computacionales. La aproximación más simple consiste en obtener una matriz de incidencia términos-documentos, donde cada celda de la matriz contiene el número de veces que cada palabra aparece en cada documento. Concretamente, el modelo de espacio de vectores considera un espacio de dimensión V donde las palabras son los ejes de este espacio V-dimensional y los documentos son puntos en dicho espacio, siendo V el número de palabras del vocabulario [12]. Usando esta matriz, las similitudes entre documentos pueden evaluarse como las distancia entre vectores en este espacio V-dimensional, por ejemplo utilizando el coseno del ángulo entre ellos. Un análisis similar podría realizarse entre los términos considerando en este caso a los documentos como ejes del espacio. El principal problema de este modelo es que en grandes colecciones de documentos con muchos términos en el vocabulario, el elevado número de dimensiones dificulta la interpretación de los resultados. Una solución a este problema consiste en proyectar primero los documentos en un sub-espacio de menor número de dimensiones donde la estructura semántica de los documentos sea más clara [13]. En este espacio de dimensiones reducido se pueden aplicar las mismas medidas de similitud o bien algoritmos de clustering [14]. Una de las técnicas más utilizadas en este ámbito es el análisis semántico latente (LSA, Latent Semantic Analysis), que descompone la matriz de incidencia términos-documentos mediante una descomposición en valores singulares para construir un nuevo espacio como combinación de las dimensiones originales, donde sea posible elegir un espacio de dimensiones reducidas [15].

Desde un punto de vista formal, el modelo de espacio de vectores considera n documentos que contienen un vocabulario de m términos, siendo $n<<m$. La matriz de incidencia términos-documentos A es una matriz mxn en el que el elemento a_{ij} representa el número de veces que la palabra i aparece en el documento j. Si dos columnas de A son similares eso significa que los correspondientes documentos tienen palabras similares y que, por tanto, están semánticamente relacionados. Lo mismo puede decirse de las filas de A, que en este caso representarían palabras. Si aparecen conjuntamente en muchos documentos, las filas tendrán cierta similitud y las palabras estarán entonces también semánticamente relacionadas. Tanto en el espacio de n dimensiones de las columnas como en el de m dimensiones de las filas, la similitud puede medirse por el coseno de los vectores columna o fila [16].

No obstante, este modelo presenta algunos problemas. Por ejemplo, las palabras parecidas o sinónimos se tratan igual que si fueran palabras completamente distintas. Incluso aunque dos documentos usen exactamente las mismas palabras, existirán otras muchas no compartidas y que son periféricas o poco relacionadas con el tópico que tratan los documentos. Al comparar los vectores que representan los documentos, el efecto negativo de las

palabras no compartidas puede atenuar o casi eliminar el efecto de las palabras compartidas, que pueden ser pocas en número aunque semánticamente relevantes.

El modelo LSA evita estos problemas proyectando los vectores columnas y filas de A en un espacio de dimensiones reducidas, donde las comparaciones de estos vectores sean menos susceptibles a los efectos de las palabras seleccionadas como vocabulario. Para ello se lleva a cabo en primer lugar una descomposición en valores singulares de A, que se expresa como el producto de otras tres matrices:

$$A=U\Sigma V^T \qquad (1)$$

donde U es una matriz mxm ortogonal, $\Sigma=diag(\sigma_1,...,\sigma_n)$ es una matriz diagonal mxn que contiene los valores singulares σ_i de A, y V es una matriz ortogonal nxn. U es realmente la matriz de autovectores de AA^T mientras que V es la matriz de autovectores de A^TA [15]. Para reducir el número de dimensiones, LSA aproxima la matriz A utilizando los k autovalores más altos, dando lugar a la matriz diagonal Σ_k, siendo $k<<n$. Esto trunca la matriz U a una matriz mxk y la matriz V^T a una matriz kxn. Como resultado se obtiene la matriz A_k dada por $U\Sigma_k V^T$ y que es la aproximación de rango k más cercana a A [17]. Del mismo modo que cada fila de A corresponde a una palabra y cada columna a un documento, cada fila de U corresponde a una palabra y cada columna de V^T a un documento. La similitud de dos palabras en el nuevo espacio reducido de k dimensiones se obtiene comparando los primero k elementos de las filas de la matriz U y, del mismo modo, la similitud de dos documentos se obtiene a partir de los primeros k elementos de la matriz V^T. A diferencia del modelo de espacio de vectores, LSA no calcula el número exacto de ocurrencias de los términos en los documentos sino que estima el número de ocurrencias en el espacio reducido de k dimensiones. En este espacio de k dimensiones, el significado de las palabras se infiere del contexto en el que ocurren. Esto significa que LSA evita los problemas de sinonimia, ya que los sinónimos se usan habitualmente en el mismo contexto [18].

IV. Resultados

Dentro de la plataforma Coursera, se han analizado un total de 25 cursos pertenecientes a la categoría 'Artificial Intelligence' y otros 25 de la categoría 'Software Engineering'. Para cada uno de ellos se ha extraído el título así como la breve descripción del curso y sus contenidos que se muestra a los usuarios para que decidan o no cursarlo. El total de 50 documentos constituye la colección sobre la que se ha aplicado el algoritmo LSA descrito en la parte de metodología, considerando un total de 100 términos de indexación elegidos en base a sus frecuencias de aparición. Una vez construida la matriz de incidencia términos-documentos, se lleva a cabo la descomposición en valores singulares descrita. Como resultado se consigue una reducción del número de dimensiones. En particular, se han eliminado el máximo número de valores singulares que permitan preservar un 80% de la varianza total de los datos. En el nuevo espacio reducido se ha aplicado un algoritmo de clustering basado en la distancia de Jaccard, cuyo resultado se muestra en la Figura 1.

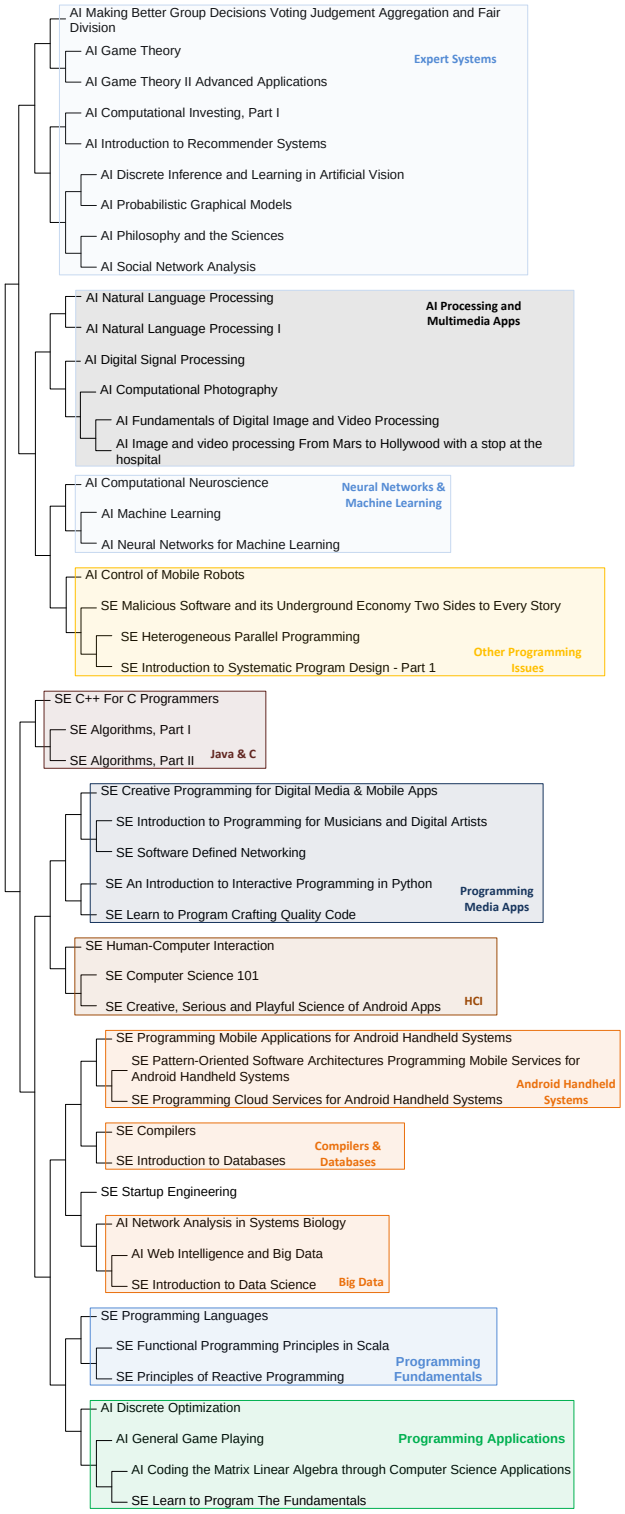

Figura 1. Agrupación de los cursos seleccionados a partir del algoritmo de clustering.

La mitad inferior del dendrograma incluye la mayoría de los cursos relacionados con 'Software Engineering' en tanto que la mitad superior incluye los relacionados con 'Artifical Intelligence'. Aunque en la mayoría de los casos se agrupan juntos los cursos pertenecientes a cada área, existen casos con un cierto solape donde se mezclan

cursos de ambas áreas. Las principales categorías dentro de cada área así como el solape entre ellas se detallan en la Tabla 1.

TABLA 1. CATEGORÍAS DE LAS ÁREAS ' SOFTWARE ENGINEERING', 'ARTIFICAL INTELLIGENCE' Y SOLAPAMIENTO ENTRE ELLAS.

Software Engineering	Artifical Intelligence	Solapamiento
Fundamentos de programación	Sistemas Expertos	Big Data
Compiladores y bases de datos	Técnicas de procesamiento y aplicaciones multimedia	Aplicaciones de programación
Sistemas Android Handheld	Redes neuronales y aprendizaje automático	Otros aspectos de programación
Interacción hombre-máquina		
Programación aplicaciones multimedia		
Java y C		

Dentro del área de 'Software Engineering' aparecen cursos relacionados con el ámbito de la programación básica (fundamentos, Java y C, compiladores y bases de datos), sistemas Android, aplicaciones multimedia e interacción hombre-máquina. Estas dos últimas categorías aparecen cercanas en el dendrograma de la Figura 1, dado que los cursos de interacción hombre máquina se centran mucho en la interfaz gráfica con el usuario.

En el área 'Artifical Intelligence' destaca la categoría sobre sistemas expertos, que a su vez se desglosa en dos: la toma de decisiones basadas en la teoría de juegos y las basadas en modelos gráficos probabilistas y análisis de redes sociales. Una segunda categoría se centra en las técnicas de procesamiento con aplicación al lenguaje natral, imágenes digitales y video. La tercera agrupación trata sobre redes neuronales y aprendizaje supervisado.

Finalmente, se han obtenido tres categorías transversales entre ambas áreas de conocimiento. La primera es sobre el área emergente de los Big Data. Desde el punto de vista del 'Software Engineering' se tratan aspectos tales como bases de datos paralelas, ciencias de datos, técnicas de procesamiento del lenguaje natural, análisis de redes sociales y técnicas de visualización de resultados. Los cursos en el área de 'Artifical Intelligence' inciden más en su aplicación a la inteligencia en la web o en los sistemas biológicos, incluyendo la aplicación de técnicas estadísticas y de aprendizaje supervisado, análisis de redes sociales y análisis semántico. La segunda categoría trata sobre aspectos aplicados de la programación para resolver problemas de optimización o de álgebra lineal, y la última abarca otros aspectos de la programación relacionados con robótica, virus informáticos o programación paralela.

Para cada una de las dos áreas consideradas se han obtenido además las principales palabras clave clasificadas según su frecuencia de aparición. El resultado se muestra en forma de tag cloud o nube de etiquetas para cada una de las áreas consideradas, Figura 2 y Figura 3. La nube de etiquetas es una lista de etiquetas populares que se muestran por orden alfabético, con un tamaño de letra proporcional a su popularidad.

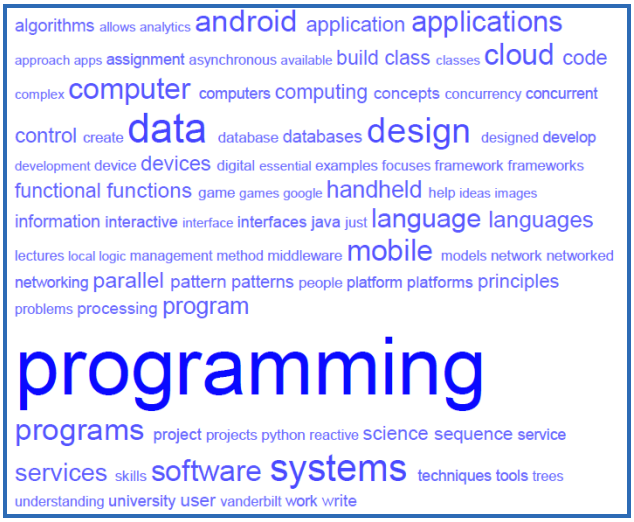

Figura 2. Tag cloud de los cursos pertenecientes a 'Software Engineering'.

Figura 3. Tag cloud de los cursos pertenecientes a 'Artifical Intelligence '.

En consonancia con la clasificación obtenida previamente, los cursos del área 'Software Engineering' afrontan esencialmente la perspectiva de la programación. Pero se observa un interés creciente en las aplicaciones para móviles en general y Android en particular, las aplicaciones en la nube y la gestión de datos.

Por su parte, los cursos del ámbito del 'Artifical Intelligence' se encuentran dominados por las aplicaciones de video e imágenes, así como los algoritmos de procesamiento y aprendizaje.

Los principales resultados que pueden derivarse del análisis anterior son los siguientes:

- El campo del 'Software Engineering' sigue estando dominado por una demanda elevada de cursos de programación, tanto básica como avanzada, y sobre distintos lenguajes de programación. No obstante, la popularidad de los dispositivos móviles y la nube hace que haya una demanda creciente de formación orientada específicamente a este tipo de dispositivos y a las aplicaciones en la nube.
- El campo de 'Artifical Intelligence' incluye elementos clásicos de procesamiento para aplicaciones de video e imágenes, y algoritmos de toma de decisiones. Destaca sin embargo un interés mayor en el procesamiento del lenguaje natural y sus aplicaciones en la web, así como en sistemas expertos basados en modelos gráficos.
- Existen campos emergentes que son transversales a las dos áreas, como los Big Data o la programación de aplicaciones robóticas.

Estos resultados muestran en primer lugar, que los MOOCs pueden actuar a modo de barómetro sobre las tendencias de demanda formativa. Los puntos anteriores destacan la existencia de demandas específicas formativas fuera de los campos tradicionales ofertados en ambas áreas de conocimiento. Se trata de campos formativos emergentes que, aunque en principio pueden tener una demanda local limitada, su demanda a nivel global es grande y anticipan lo que puede ser una demanda local futura más elevada. En segundo lugar, los resultados obtenidos reafirman la necesidad de abordar determinados campos desde una perspectiva multidisciplinar, mediante la integración de dos o más áreas de conocimiento.

Estos resultados también tienen implicaciones importantes sobre la enseñanza reglada dentro de los sistemas de educación superior. Las nuevas demandas formativas en campos emergentes sugieren la necesidad de planes de estudios más flexibles, capaces de adaptarse a una evolución tecnológica en constante cambio. La programación y aplicaciones sobre dispositivos móviles, o el análisis experto de portales web constituyen temas no suficientemente abordados en los nuevos grados y titulaciones ofertadas por las Universidades. Asimismo, es cada vez más imperativa la oferta formativa multidisciplinar, que salte las barreras en muchos casos estancas de las áreas de conocimiento. El campo emergente de los Big Data posee por ejemplo relación no sólo con las dos áreas objeto de este artículo, sino con varias otras adicionales. Sin esa muldisciplinariedad es imposible una oferta formativa completa en estos campos. Si bien la enseñanza reglada no puede funcionar a la misma velocidad que la ofertada por los MOOCs, sí que puede monitorizar las tendencias con objeto de anticiparse a las futuras demandas.

V. CONCLUSIONES

Este artículo analiza mediante técnicas de procesamiento del lenguaje natural la oferta formativa en el campo de los MOOCs. La hipótesis de partida es que la demanda formativa a nivel global actúa a modo de predictor de las demandas futuras a nivel local, al aprovechar las economías de escala que proporcionan potenciales usuarios en todo el mundo y sus efectos de "cola larga". Como caso de estudio, el artículo analiza los cursos ofertados en la plataforma coursera relativos a las áreas de 'Software Engineering' y de 'Artifical Intelligence'. Los resultados obtenidos aportan resultados importantes en cuanto a las nuevas demandas formativas en campos emergentes, así como sobre el carácter multidisciplinar de determinadas demandas formativas.

ACKNOWLEDGEMENTS

Este trabajo ha sido financiado por la Consejería de Economía, Innovación, Ciencia y Empleo, Junta de Andalucía (Proyecto de Excelencia referencia P12-SEJ-328).

REFERENCES

[1] A. McAuley, B. Stewart, G. Siemens and D. Cormier, The MOOC model for digital practice. Disponible en: http://www.elearnspace.org/Articles/MOOC_Final.pdf. [Acceso 7 Enero 2014].

[2] T. Daradoumis, R. Bassi, F. Xhafa, S. Caballe, "A Review on Massive E-Learning (MOOC) Design, Delivery and Assessment", 2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Compiegne, France, pp. 208- 213, 2013.

[3] C. Severance, "Teaching the World: Daphne Koller and Coursera", Computer, Vol. 45, Iss. 8, pp. 8-9, 2012.

[4] R. C. de Boer, H. van Vliet, "Architectural knowledge discovery with latent semantic analysis: Constructing a reading guide for software product audits", Journal Systems and Software, Vol. 81, Iss. 9, pp. 1456–1469, 2008.

[5] S. Audsley, K. Fernando, B. Maxson, B. Robinson & K. Varney, "An Examination of Coursera as an Information Environment: Does Coursera Fulfill its Mission to Provide Open Education to All?", The Serials Librarian, Vol. 65, No. 2, pp. 136-166, 2013.

[6] I. S. Abeywardena, "Public opinion on OER and MOCC: a sentiment analysis of twitter data", Proceeding of International Conference on Open and Flexible Education (ICOFE 2014), Hong Kong SAR, China, 2014.

[7] C. Sandeen, "Assessment's place in the new MOOC world", Research & Practice in Assessment, Vol. 8, Iss. 1, pp. 5-12, 2013.

[8] N. J. Pelaez, Calibrated peer review in general education undergraduate human physiology. In P. A. Rubba, J. A. Rye, W. J. DiBiase, & B. A. Crawford (Eds.), Proceedings of the Annual International Conference of the Association for the Education of Teachers in Science, Costa Mesa, CA, pp. 1518-1530, 2001.

[9] S. P. Balfour, "Assessing Writing in MOOCs: Automated Essay Scoring and Calibrated Peer Review", Research & Practice in Assessment, Vol. 8, Iss.1, pp. 40-48, 2013.

[10] T. Clarke, "The advance of the MOOCs (massive open online courses): The impending globalisation of business education?", Education + Training, Vol. 55, Iss: 4/5, pp.403-413, 2013.

[11] M. Clara, E. Barbera, " Learning online: massive open online courses (MOOCs), connectivism, and cultural psychology", Distance Education, Vol. 34, Iss. 1, pp. 129-136, 2013.

[12] G. Salto, and M. J. McGill, An Introduction to Modern Information Retrieval, McGraw-Hill, New York, 1983.

[13] D. Cai, X. He, and J. Han, Document Clustering Using Locality Preserving Indexing, IEEE Transactions on Knowledge and Data Engineering, Vol. 17, Iss. 12, pp. 1624-1637, 2005.

[14] J. Shi, and J. Malik, Normalized Cuts and Image Segmentation, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 22, Iss. 8, pp. 888-905, 2000.

[15] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman, Indexing by latent semantic analysis, Journal of the American Society of Information Science, Vol. 41, Iss. 6, pp. 391–407, 1990.

[16] S. Kawaguchi, P. K. Garg, M. Matsushita, K. Inoue, MUDABlue: An automatic categorization system for Open Source repositories, The Journal of Systems and Software, Vol. 79, pp. 939–953, 2006.

[17] G. H. Golub, C.F. Van Loan, Matrix Computations, (Johns Hopkins University Press, 1989.

[18] R. C. De Boer, H. van Vliet, Architectural knowledge discovery with latent semantic analysis: Constructing a reading guide for software product audits, Journal Systems and Software, Vol. 81, Iss. 9, pp. 1456–1469, 2008.