DATA REVOLUTION AND MOOC: A NEW TECHNOLOGICAL PARADIGM Kristin Lord Catholic University of America.

GHANOUANE Karim [1] and BENKARAACHE Taoufik [2]

[1] Faculty of Law, Economics and Social Sciences Mohammedia, Hassan 2 University, Morocco
[2] Faculty of Law, Economics and Social Sciences Mohammedia, Hassan 2 University, Morocco

ABSTRACT

The technological euphoria has enabled organizations in all sectors to have access to large volumes of data. Data analysis has become an important area of study for practitioners and researchers in both social and economic fields. As such, the organization requires a specialized set of skills and resources for data collection and analysis. MOOCs; by being open online training courses that exploit new digital technologies; enhance the learning experience and reach a wide audience. MOOCs could be an alternative to fill the lack of data-centered skills. Through the combination of Web Scraping data extraction techniques and content analysis methodology, this paper aims to provide taxonomy of data-driven MOOC training offerings.

KEYWORDS

Big Data, MOOC, Skills, Content Analysis, Web Scraping

1. INTRODUCTION

There is a long history of humanity and data, with governments, businesses, researchers, and citizens producing and using data to monitor, regulate, benefit from, and make sense of the world. Traditionally, access to data has been a long and expensive process to generate, analyze, and interpret. Good quality data were scarce and valuable (Kitchin, 2014). In recent years, this trend has begun to change quite dramatically. While data have maintained their status as a scarce resource, their production and nature have undergone disruptive innovations in production, extraction, analysis, storage, and use (Christensen, 1997). Recent technological revolutions such as social media, mobile technology, the Internet of Things, allow us to generate large volumes of data instantaneously (McAfee et al. 2012). In addition, these technologies materially and discursively reconfigure the production, circulation, and interpretation of data, producing what we call "Big Data". The concept of Big Data and its application in business has attracted considerable attention in recent years because of its great potential to generate economic impacts and offer competitive advantages (Chen et al., 2012). Big Data refers to the amount of data that is beyond the capabilities of traditional technologies to effectively store, process, and analyze (Kaisler et al., 2013). Data analysis covers several applications such as e-commerce, business intelligence, e-government, health, and security (Chen et al., 2012).

In the age of Big Data, companies gain a competitive advantage in the marketplace based on data analysis and rapid decision making based on huge volumes of information. This decision

making is outside the information system function, otherwise dealing with the market falls within the scope of business units such as marketing, finance, and logistics (Chen et al., 2012). The need for specialized skills across the entire Big Data value chain may exceed the training supply. The need exceeded 1.5 million managers and analysts in Big Data processing in the United States (Manyika et al., 2011).

MOOCs represent a powerful new way of accessing knowledge, hence the growing interest in many discussions in higher education. Raising issues related to low pedagogical quality, high drop-out rates or research challenges triggered by their large-scale functionality (Margaryan et al., 2015; Onah et al., 2014; Dillenbourg et al., 2014; Manathunga et al., 2015). Most MOOCs follow a behaviorist pedagogical approach where instructors add educational content to the course and students self-assess their learning with questionnaires, limiting interaction between participants and instructors to discussion forums (Daniel, 2012).

This paper; on data-driven training through MOOCs (Massive Online Open Course); is intended to map the training offer conveyed by pioneering online platforms at the international level. We hope to contribute to future sources of knowledge and to enrich current discussions on the importance of online training. This paper will seek to answer the following questions: In which categories do MOOCs around data fit? In which languages? In which levels? What has been the evolution over the last 5 years? What disparity is observed between countries?

By analyzing and interpreting the statistical results of a text mining application on MOOC training offers, we develop taxonomy of information broken down by category, language, cost, duration... Our results can guide professionals and academic institutions to evaluate and improve their needs and distance learning offers respectively.

The rest of the article is organized as follows: Section 1 reviews the literature. Section 2 discusses the methodology adopted for this paper. Finally, Section 3 provides a detailed analysis and discussion of the extracted data.

2. THEORETICAL FRAMEWORK: LITERATURE REVIEW

2.1. Data Revolution: From Data to Big Data

Data refer to captured elements extracted through observations, calculations, experiments, or record-keeping (Borgman et al., 2007). For Rosenberg (2013), data exist before the interpretation that converts them into facts, evidence, and information. Similarly, data are abstract, discrete, aggregated, and independent of format, medium, language, producer, and context. According to Floridi (2008), data are dependent on three types of neutrality:

- Taxonomic: the data are relationally defined concerning other specific data
- Typological: data can take different forms, e.g. primary, secondary, metadata, operational, etc.
- Genetics: data can have semantics independent of their understanding.

As a data-centric approach, Big Data analysis has its roots in the field of date language database management. It relies heavily on various data collection, extraction, and analysis technologies (Chaudhuri et al., 2011).

Originally, data were mainly structured, collected by companies through various existing systems, and often stored in relational database management systems (RDBMS). Analytical techniques, popularized in the 1990s, were based on statistical methods developed in the 1970s and data mining techniques developed in the 1980s. Data management and warehousing were considered the basis for data analysis (Chen et al., 2012).

In the early 2000s, the Internet and the World Wide Web began to provide opportunities for data collection, research, and analytical development. Search engines and e-commerce platforms allowed organizations to have an online presence and interact directly with their customers (Chen et al., 2012). With the advent of Web 2.0, website design, product placement optimization, customer transaction analysis, market structure analysis, and product recommendations could be achieved through the analysis of web data (Doan et al., 2011; O'Reilly, 2005). Web 2.0 in turn required evolving skills and techniques in text mining (e.g., information retrieval, topic identification, opinion mining, question answering), web crawling, social network analysis, and Spatio-temporal analysis with existing DBMS-based systems (Chen et al., 2012).

Mobile devices such as the iPad, iPhone, and other smartphones and their complete ecosystems: from downloadable applications, travel advice, to multiplayer games, are transforming different facets of society, education, health, and entertainment. The Internet of Things, which covers sensor-based devices equipped with RFID, bar codes, and radio tags, opens up new opportunities for innovative applications (Chen et al., 2012). Underlying and contextual mobile analysis and location techniques for collecting, processing, analyzing, and visualizing mobile data are promising areas of academic research (Bitterer, 2011).

Laney in 2001 was the first to introduce the term Big Data into the enterprise and introduced the famous "3V" to refer to volume, speed, and variety. As companies have become increasingly attracted to Big Data, this interest has shifted from the constraints of storing this massive data to its analysis. Google, Amazon, and Facebook were the first to exploit Big Data, using data mining and machine learning techniques (Davenport, 2006). Unlike these predecessors in data analysis, Big Data applications focus on exploration, discovery, and prediction (Dhar, 2013). Data analysis includes enterprise-centric practices and methodologies that can be applied to a variety of high-impact applications such as e-commerce, business intelligence, e-government, health, and security (Chen et al., 2012).

The strategic analysis makes extensive use of data, statistical and quantitative analysis, and explanatory and predictive modeling to help managers make decisions and improve business operations (Chen, 2012). Big data also helps identify opportunities for future performance improvement (LaValle et al., 2011).

Fitzgerald (2014) points out that there are three types of career categories for graduates specializing in Big Data analysis: Business Consultants, Financial and Risk Analysts, and Data Scientists. Core competencies for business analysis include optimization analysis, descriptive analysis, and predictive analysis. For example, Watson, Wixom, and Ariyachandra (2013) suggest the following:

- Communication skills
- SQL skills and query
- Data mining and data warehousing
- Statistical skills
- Data visualization
- Text mining
- NoSQL skills
- Emerging functional knowledge

Several academic disciplines have contributed to the data revolution, including information systems, statistics, management, and marketing. Computer science programs, in particular, are well-positioned to develop a new generation of skills because of the emphasis on data and information technology management, business-oriented statistical analysis, and openness to other disciplines such as finance, accounting, marketing, and economics (Chen et al., 2012).
The impact of Big Data is already visible in science, business, government, and civil society, seeking to extract information and draw conclusions from a limited number of observations, and academic disciplines are now beginning to cope with this avalanche of data (H.J. Miller, 2010).

2.2. What is a MOOC?

Having issued the first MOOC in 2088, Siemens and Downes (2009) define it as a by-product of open teaching and learning. Since 2008, many MOOCs have been developed and delivered and several companies have been established by entrepreneurs at various prestigious universities to create and deliver these courses. The founders of the main MOOC platforms are educators who function as well-supported entrepreneurs, for example, edX launched by professors from the Massachusetts Institute of Technology (MIT) and Harvard University in 2012, Coursera founded by professors from Stanford University in 2012 (Deale, 2015). With the arrival of many for-profit MOOC providers such as Edx, Coursera, or Udacity, the "open to all" feature of MOOCs tends to lose ground to fee-based MOOCs (Masters, 2011).

Massive Open Online Courses (MOOCs) are open courses that are made available online by exploiting information and communication technologies to enhance the learning experience and reach a wide audience worldwide. MOOCs have the characteristics that they are open, free, easily accessible, participatory, and distributed (Pappano, 2012). Otherwise, MOOCs are the result of a combination of innovations and technologies that provide learning opportunities for a large number of people (Siemens, 2013). Glance, Forsey, and Riley (2012) provide a broad definition of the characteristics of MOOCs:

- A massive number of learners
- Online and open access
- Conferences transmitted in video format (8-12 minutes)
- Evaluations in the form of formative questionnaires, peer review, and/or self-assessment.
- Online forums for exchange between members or with instructors.

However, MOOCs are not free of limitations, given the massive number of learners, the management of the course is quite complex, hence the limited involvement of the instructors. Also, instructors are unable to review and score all tasks, so they are usually classified by

algorithms or evaluated by peers. Besides, MOOCs are characterized by high dropout rates and plagiarism and security vulnerabilities (Daradoumis et al., 2013; Migue et al., 2013).
Siemens (2013); cited by Aharony and Bar-Ilan (2016); identifies three models of MOOC:

- xMOOCs that replicate the traditional model of the online expert teacher and learners as knowledge consumers with video tutorials and graded tasks.
- The cMoocs focus on the pedagogical approach of connectivity that considers knowledge as a networked state and learning as the process of creating these networks, using online and social tools.
- Quasi-MOOCs refer to online courseware such as open educational resources and do not involve social interaction.

While some researchers argue that the MOOC craze could disrupt higher education by closing universities or replacing university degrees with certificates of completion (Gregory, 2012; Schierenbeck, 2012; Bayeck, 2016), others argue that MOOCs do not pose a threat to the education sector due to disparities in education, gender, and employment (Kalman, 2014; Bayeck, 2016).

3. Research Methodology

This study was designed according to a two-stage methodology:

In the first stage, we extracted a considerable number of training MOOCs around the data. In this sense, we chose "data", "data" and "data" as keywords. This step required the use of the technique of extracting website content, known as "Web Scraping", using free programs such as web scraper, parsehub or octoparse. Web Scraping, also known as web extraction or web harvesting, is a technique for extracting structured or unstructured data for later retrieval or analysis (Zhao et al., 2017). Typically, data are extracted using the HTTP protocol or via a web browser. This is done manually by a user or automatically by indexing robots. The data extraction technique is widely recognized as an efficient and powerful technique for collecting Big Data (Zhao et al., 2017; Mooney and Ackerman, 2015; Bar-Ilan, 2001).

Data collection has focused on course offerings from pioneering international platforms. These MOOCs are of the cMOOC type, over a period from 2015 to 2020. After removing all duplicate and incomplete data, we were left with a data set of 359 MOOCs, which we used as data for our analysis.

A second step using the content analysis methodology. This research technique is used for the purpose of objective, systematic and quantitative description of the content of the communication (Berelson, 1952). For deKolbe and Burnett (1991), content analysis is an observational research method adapted to mass communications, it allows us to systematically evaluate the symbolic content of all forms of communication and analyze it at several levels (image, word, roles, etc.).

4. Theoretical Framework: Literature Review

4.1. Annual evolution

In five years, the evolution has been 3375%, in 2020, 278 MOOCs related to the data were offered compared to 8 in 2015. The hypothesis of containment due to the COVID19 pandemic could explain this explosion.

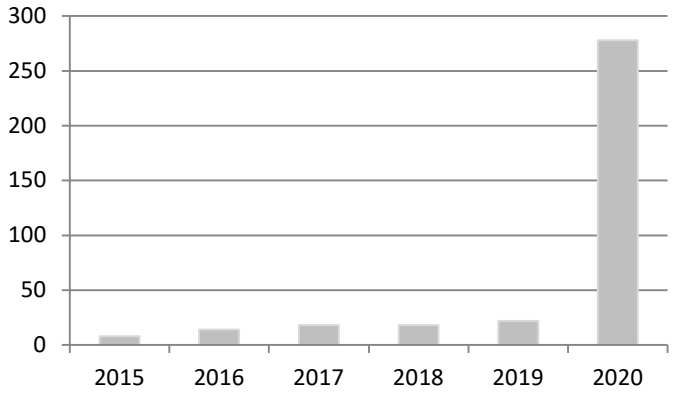

Figure 1. Number of MOOCs from 2015 to 2020

4.2. Distribution by platform

Market leader in the cMOOC market, Coursera and edX account for 84.5% of MOOC offers in connection with "Data". This can be explained by the fact that these platforms have evolved towards partnerships with public universities (Lewin, 2013). For example, by May 2013, Coursera had signed agreements with state universities that led to the enrolment of more than one million students, thus moving MOOCs further into the higher education mainstream (Lewin, 2013; Deale, 2015). Also, MOOCs, such as those offered by edX, now offer supervised examinations through the Pearson service with over 450 test centers in 110 countries, which increases the credibility of MOOCs (Parry, 2012; Deale, 2015).

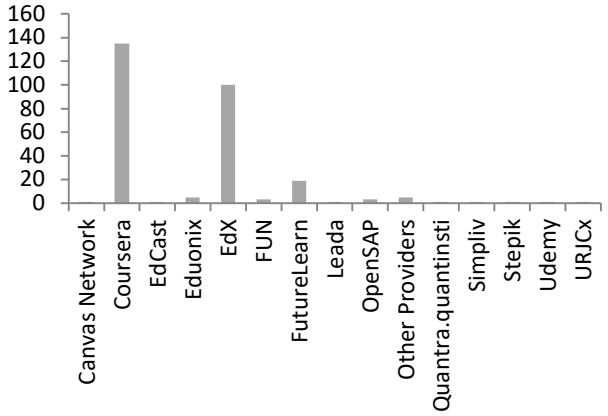

Figure 2. Distribution by platform

4.3. Distribution by categories

The Distribution of MOOCs by category is 51.8% and 24.5% in data science and computer science respectively. This finding is confirmed by Chen (2012), computer science programs, in

particular, are well-positioned to train a new generation of skills because of the emphasis on data management and information technology.

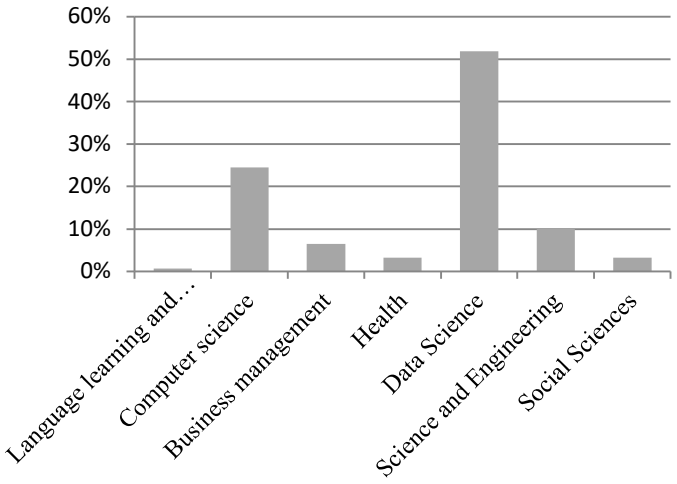

Figure 3. Distribution by categories

4.4. Distribution by the level of difficulty

49% and 39% are beginners and intermediate level MOOCs respectively. Usually, a MOOC has no pre-requisites, making it accessible to everyone on the internet (DeWaard et al., 2011). Researchers have found that learners remain more engaged if MOOCs are of medium difficulty (Adamopoulos, 2013; Deale, 2015).

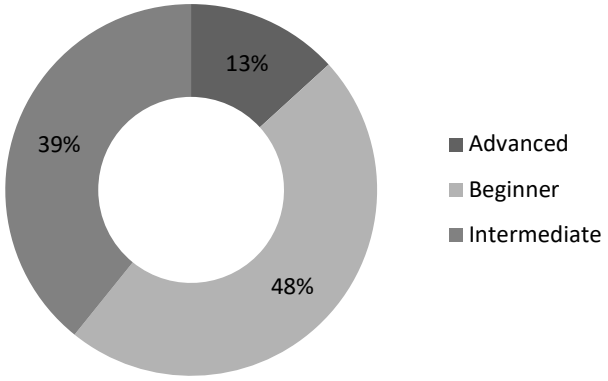

Figure 4. Distribution by the level of difficulty

4.5. Distribution by duration

39.6% of the MOOCs related to the data vary between 4 and 5 weeks, the estimated load per week in the number of hours is less than 8 hours for more than 78% of the extracted MOOCs. For the majority of studies, learners remained more engaged if the MOOCs had a moderate workload (about six hours per week), contained projects and final exams, peer review, and lasted less than eight weeks (Adamopoulos, 2013; Deale, 2015). Jordan's (2015) study found that the shortest MOOCs had the highest completion rates.

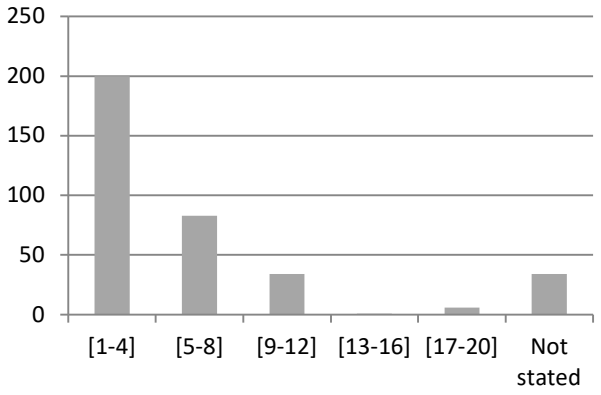

Figure 5. Distribution by number of hours/week

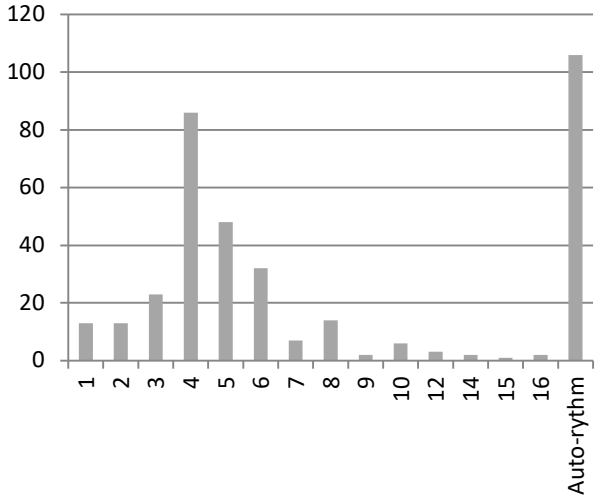

Figure 6. Distribution by number of weeks

4.6. Distribution by duration

Figure 7 clearly shows that almost all MOOCs are in English. This can be argued by a breakdown of MOOCs by a country where 85% of MOOCs are concentrated in countries where the language of instruction is English. Furthermore, as we have explained in the case of Coursera and edX, two American platforms where partnerships have been signed with public universities.

Furthermore, the distribution of MOOCs by country (see Figure. 8) shows a concentration of training offers in English-speaking countries, the majority of which are in the United States (64% of MOOCs).

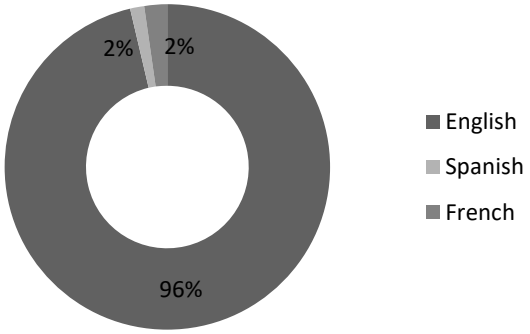

Figure 7. Distribution by language

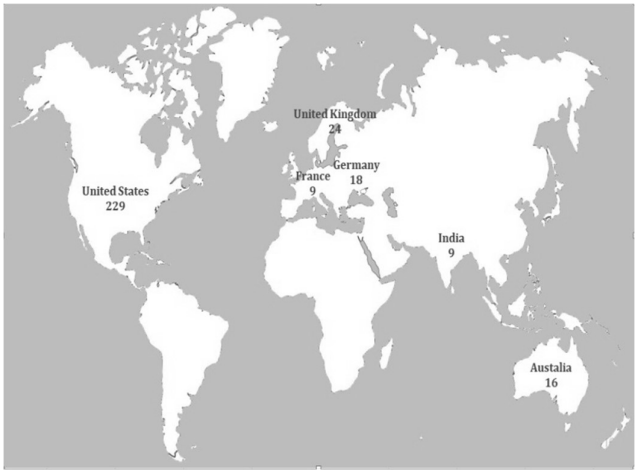

Figure 8. Distribution of MOOCs by country (more than 9 MOOCs)

4.7. Distribution by duration

The analysis of the graph above shows that from 2018 onwards, the trend has been reversed in favor of paying MOOCs for data-linked training. This is only the culmination of partnerships between MOOC platforms and universities in the USA to grant credits to students who have successfully completed a MOOC at lower prices than those needed to follow the same courses on campus (Kolowich, 2013).

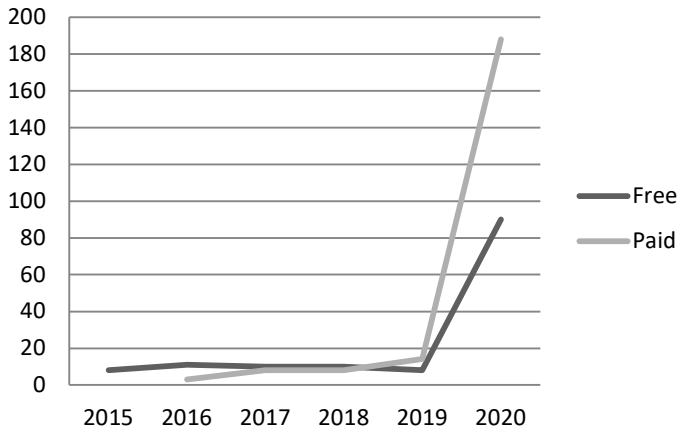

Figure 9. Distribution by Cost

5. CONCLUSIONS

This promises to be an exciting decade for research and development of high-impact data analysis for industry and academia. While business and industry have already taken important steps towards adopting Big Data for their needs. The academic community faces unique challenges and opportunities to produce relevant and sustainable scientific and societal impacts (Chen et al., 2012). Research and training programs need to carefully assess future directions, programs, and action plans.

For the scientific community, the analysis of data-driven training provision in the form of MOOC presents a unique opportunity for universities to position themselves as a viable option for training professionals with the depth and academic rigor necessary to address the needs and increasing complexity of data analysis. Business schools will need to focus on MOOCs designed to improve communication and reporting skills. MOOCs can fill this interdisciplinary need and cover analytical and computer skills, knowledge of the business world, as well as communication skills, all with a focus on data.

For practitioners, this article should serve as a reference based on the current state of course offerings, with an emphasis on popularizing the concept of Big Data for all business units of the company or even integrating the MOOC concept into the framework of intra-company training.

There were limits to this study. It involved collecting data, using a Web Scraping technique, to draw up taxonomy of courses concerning the data. This study was descriptive in nature and needs to be followed by further research on MOOCs, how MOOCs are implemented, particularly about Big Data and its relationship to the various disciplines, and then further research with instructors and learners to determine their value as tools to enhance student learning.

DOI: 10.5121/ije.2021.03601

REFERENCES

[1] Adamopoulos, P. (2013). What makes a great MOOC? An interdisciplinary analysis of student retention in online courses.

[2] Aharony, N., & Bar-Ilan, J. (2016). Students' perceptions on MOOCs: An exploratory study. Interdisciplinary Journal of e-Skills and Life Long Learning, 12, 145-162.

[3] Bar-Ilan, J. (2001). Data collection methods on the Web for infometric purposes—A review and analysis. Scientometrics, 50(1), 7-32.

[4] Bayeck, R. (2016). Exploratory study of MOOC learners' demographics and motivation: The case of students involved in groups. Open Praxis, 8(3), 223-233.

[5] Berelson, B. (1952). Content analysis in communication research.

[6] Bitterer, A. (2011). Hype cycle for business intelligence. Gartner, Inc., Stamford, CT.

[7] Borgman, C. L., Wallis, J. C., & Enyedy, N. (2007). Little science confronts the data deluge: habitat ecology, embedded sensor networks, and digital libraries. International Journal on Digital Libraries, 7(1), 17-30.

[8] Chaudhuri, S., Dayal, U., & Narasayya, V. (2011). An overview of business intelligence technology. Communications of the ACM, 54(8), 88-98.

[9] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. MIS quarterly, 1165-1188.

[10] Christensen, C.M. (1997) The Innovator's Dilemma. Harvard Business Review Press, Cambridge, MA.

[11] Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. Journal of interactive Media in education, 2012(3).

[12] Daradoumis, T., Bassi, R., Xhafa, F., & Caballé, S. (2013, October). A review on massive e-learning (MOOC) design, delivery and assessment. In 2013 eighth international conference on P2P, parallel, grid, cloud and internet computing (pp. 208-213). IEEE.

[13] Davenport, T. H. (2006). Competing on analytics. Harvard business review, 84(1), 98.

[14] Deale, C. S. (2015). An exploratory study of hospitality and tourism educators' use and perceptions of MOOCs. Journal of Teaching in Travel & Tourism, 15(2), 150-165.

[15] DeWaard, I., Abajian, S., Gallagher, M. S., Hogue, R., Keskin, N., Koutropoulos, A., & Rodriguez, O. C. (2011). Using mLearning and MOOCs to understand chaos, emergence, and complexity in education. International Review of Research in Open and Distributed Learning, 12(7), 94-115.

[16] Dhar, V. (2013). Data science and prediction. Communications of the ACM, 56(12), 64-73.

[17] Dillenbourg, P., Fox, A., Kirchner, C., Mitchell, J., & Wirsing, M. (2014). Massive open online courses: current state and perspectives (Dagstuhl perspectives workshop 14112). In Dagstuhl Manifestos (Vol. 4, No. 1). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[18] Doan, A., Ramakrishnan, R., & Halevy, A. Y. (2011). Crowdsourcing systems on the world-wide web. Communications of the ACM, 54(4), 86-96.

[19] Fitzgerald, M., Kruschwitz, N., Bonnet, D., & Welch, M. (2014). Embracing digital technology: A new strategic imperative. MIT sloan management review, 55(2), 1.

DOI: 10.5121/ije.2021.03601

[20] Floridi, L. (2008) 'Data', in W.A. Darity (ed.), International Encyclopedia of the Social Sciences, 2nd edition. Detroit: Macmillan. Preprint online: http://www.philosophyofinformation.net/publications/pdf/data.pdf

[21] Glance, D. G., Forsey, M., & Riley, M. (2013). The pedagogical foundations of massive open online courses. First monday.

[22] Gregory, M. A. (2012). MOOC and you're out of a job: Uni business models in danger. The Conversation, 11.

[23] Miller, H. J. (2010). The data avalanche is here. Shouldn't we be digging?. Journal of Regional Science, 50(1), 181-201.

[24] Jordan, K. (2015). MOOC research literature browser.

[25] Kaisler, S., Armour, F., Espinosa, J. A., & Money, W. (2013, January). Big data: Issues and challenges moving forward. In 2013 46th Hawaii International Conference on System Sciences (pp. 995-1004). IEEE.

[26] Kalman, Y. M. (2014). A race to the bottom: MOOCs and higher education business models. Open Learning: The Journal of Open, Distance and e-Learning, 29(1), 5-14.

[27] Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. Sage.

[28] Kolbe, R. H., & Burnett, M. S. (1991). Content-analysis research: An examination of applications with directives for improving research reliability and objectivity. Journal of consumer research, 18(2), 243-250.

[29] Kolowich, S. (2013). The professors who make the MOOCs. The Chronicle of Higher Education, 18.

[30] Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META group research note, 6(70), 1.

[31] LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. MIT sloan management review, 52(2), 21-32.

[32] Lewin, T. (2013, June 19). Online classes fuel a campus debate. The New York Times, NYTimes.com. Retrieved February 2, 2014

[33] Manathunga, K., & Hernández-Leo, D. (2015). Has research on collaborative learning technologies addressed massiveness? A literature review. Educational Technology & Society, 18(4), 357-370.

[34] Manyika, J. (2011). Big data: The next frontier for innovation, competition, and productivity. http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation.

[35] Margaryan, A., Bianco, M., & Littlejohn, A. (2015). Instructional quality of massive open online courses (MOOCs). Computers & Education, 80, 77-83.

[36] Masters, K. (2011). A brief guide to understanding MOOCs. The Internet Journal of Medical Education, 1(2), 2.

[37] McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. Harvard business review, 90(10), 60-68.

[38] Miguel, J., Caballe, S., & Proeto, J. (2013). Providing information security to MOOC: Towards affecting student authentication. 5th International Conference on Intelligent Networking and Collaborative Systems.

DOI: 10.5121/ije.2021.03601

[39] Mooney, M. E., & Ackerman, S. A. (2015). Reaching Regional and Local Learners via a Great Lakes MOOC. AGUFM, 2015, ED23B-0858.

[40] Onah, D. F., Sinclair, J., & Boyatt, R. (2014). Dropout rates of massive open online courses: behavioural patterns. EDULEARN14 proceedings, 1, 5825-5834.

[41] O'reilly, T. (2005). Web 2.0: compact definition.

[42] Pappano, L. (2012). The year of the MOOC. New York Times. Retrieved July 15, 2013.

[43] Parry, M. (2012). edX offers proctored exams in open online courses. The Chronicle of Higher Education. Retrieved February 13, 2014.

[44] Rosenberg, D. (2013). Data before the fact. Raw data" is an oxymoron, 15, 40.

[45] Schierenbeck, C. (2012). Disruptive Innovations in higher education. Fixing Higher Education, 51–65.

[46] Downes, S., & Siemens, G. (2009). Connectivism and Connective Knowledge: Getting Started. MOOC course, University of Manitoba.

[47] Siemens, G. (2013). Learning analytics: The emergence of a discipline. American Behavioral Scientist, 57(10), 1380-1400.

[48] Watson, H. J., Wixom, B. H., & Ariyachandra, T. (2013). Insights on Hiring for BI and Analytics. Business Intelligence Journal, 18(2), 4-7.

[49] Zhao, R., Liu, Y., Zhang, N., & Huang, T. (2017). An optimization model for green supply chain management by using a big data analytic approach. Journal of Cleaner Production, 142, 1085-1097.

[50] Zeeshan, K., Watanabe, C., & Neittaanmäki, P. (2021). RENEWING MATHEMATICAL AND DIGITAL EDUCATION INITIATIVES IN THE ASIAN COUNTRIES: Areview. International Journal of Education (IJE), 2(1).

DOI: 10.5121/ije.2021.03601