A Hybrid Personalization Model for Searching Multiple MOOCs

Khadijah Alzahrani*[a], Maram Maccawy[b]
[a]Department of Computerized Information Systems, Faculty of Computers & Information Technology, King Abdulaziz University,
Jeddah, Saudi Arabia
*Corresponding Author: kalzahrani0193@stu.kau.edu.sa

Abstract

The gap between the Labor market needs and recent Higher education graduates continues to perpetually expand worldwide, including the Saudi labor market; where many employers complain that recent graduates lack some essential skills that would allow them to successfully compete in their workplace. On the other hand, online learning methods and technologies such as Massive Open Online Courses (MOOCs) are developing rapidly, providing an excellent mean for lifelong learning. However, due to the overwhelmingly large number of MOOCs, find it challenging to locate the most relevant courses to optimize their skill set. This research is inspired by the Saudi movement to improve education outcomes as an objective of *Saudi 2030 Vision*. This paper mainly inspects the *'lost in hyperspace'* phenomenon plaguing MOOCs platforms and content. It proposes a hybrid, personalized MOOCs search model known as the MOOC Recommender Search Engine (MRSE). This model aims to help learners search and reach suitable high-quality courses, which increases their chances of meeting the labor market requirements.

Keywords: MOOCs, Personalization, Search Engines, Recommender Systems, User Modeling (UM), MOOCs Recommendation Search Engine (MRSE).

Introduction

The skills gap continues to widen as many recent graduates lack the required skills needed in a continuously changing labor market. Hence, more graduates have been facing unemployment in the past few years. According to the Economic Policy Institute (EPI), worldwide unemployment is growing (ILO, 2018). Governments develop a sustainable local workforce by investing in skills, education, and training to facilitate careers for young people and boost their countries' competitiveness. However, the skills gap is still getting bigger in the Gulf Cooperation Council (GCC) and around the world (Ernst & Young, 2015). In particular, Saudi Arabia has the highest rate of unemployment in the GCC by 5.7 percent in 2017 according to the International Labor Office's (ILO) latest unemployment trend publication. Saudi Arabia and GCC countries need significant changes to improve their employment rates despite the fact that the unemployment in Saudi and GCC has been reduced by 0.1 percent in 2018 and the labor market outlook is stable (ILO, 2018). Therefore, it is a Saudi movement to balance the connections between education outcomes and the labor market requirements through the "2030 Vision" plans ("Vision 2030," 2017).

Enhancing education is a fundamental step towards employment growth, which needs a refocus to meet the changing needs of the labor market (ILO, 2018). Massive Open Online Courses (MOOCs) have gained immense acceptance amongst online learning communities. MOOCs present online degrees and high-quality courses authored by academics and experts (Jansen & Lizzie, 2018; Shah, 2013). As an online-learning platform, MOOCs offer open admission (anyone, anytime, and anywhere) to (open-sourced or not) courses regardless of the learner's background (Baturay, 2015). Some instances of such platforms are Edx (“edX”), Coursera (“Coursera”), Udemy (“Udemy,”) and Udacity (“Udacity Week”)

MOOCs have become a stand-alone educational platform. Through MOOCs, individuals obtain new knowledge and skills to attain professional advancement, and use this knowledge to access and compete in the labor market (Saltzman, 2014; Teaching, 2017). Universities and institutions focus on enhancing student capabilities to get better employment through MOOCs; by facilitating higher education, professional training and learners' experiences (Ossiannilsson, Altinay, Altinay, & Albright, 2016). The open education platform offered by MOOCs boosts students' innovation and skills to become more competitive in both education and the labor market (Zvacek, Restivo, Uhomoibhi, & Helfert, 2016). Hence, MOOCs are considered an excellent solution to reduce the skills gap; where skill level optimization increases students' employability (European Commission (DG EAC), 2013).

Over the years, 800 universities have launched a total of around 10,000 MOOCs (Shah, 2017). To date, the number of MOOCs courses reaches almost 9,400 as well as more than 500 MOOC-based credentials (see Figure 1) (Shah, 2018).

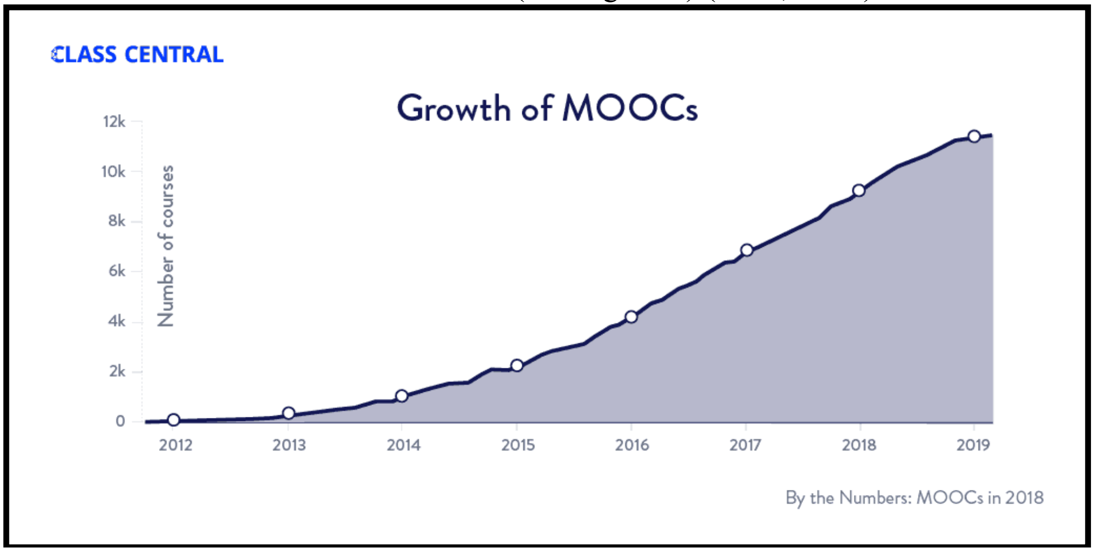

Figure 1. The Number of MOOCs in 2018 (Shah, 2018)

Consequently, some learners suffer from a problem known as ‘lost in hyperspace’ (MOOC in this case), while trying to search and reach the suitable material. According to Otter (Otter & Johnson, 2000), being ‘lost in hyperspace’ is one of the essential difficulties which users face when trying to search through hypertext systems. To ensure better performance and to attract and engage more learners, some MOOCs adopted personalization to improve individual learning experiences through implementing personalized services. For instance personalized learning paths, personalized assessments and feedback, as well as personalized forum threads and recommendation services for related learning materials or learning tasks (Sunar, Abdullah, White, & Davis, 2016; Xiao, Wang, Jiang, & Li, 2018).

In order to provide adequate personalization, the user needs to be understood and modeled. User modeling involves collecting all the required data about the user that would enable the system to provide him with a tailored user experience (Brusilovsky & Millán, 2007). Even though the personalization of MOOCs may reduce the number of materials available to the user, it does not overcome the 'lost in hyperspace' issue completely.

This paper explores the challenge of solving the 'lost in hyperspace' problem in MOOCs. It proposes the MOOC Recommender Search Engine (MRSE); a hybrid search model that consists of a recommender system and a search engine. The following sections will further introduce and clarify the related work as well as the proposed model.

Over the years, researchers adopted similar practices to this research approach, such as in the field of news (De Pessemier, & Vanhecke, 2015) and music (Cohen & Fan, 2000); through combining information filtering and information retrieval. However, the personalization research is still ongoing for MOOCs. This research is aimed towards the following objectives:

1. To explore the experiments, results and recommendations for similar implementations.
2. Identify best practices to support the implementation of the research approach.

Several search engines and online platforms have adopted the concept of retrieving multiple courses from multiple MOOCs, but most of them lack personalization. This may adversely affect the student improvement and the educational process. For this purpose, we have formulated the following research question:

1. How could we support the learners to navigate multiple MOOCs and retrieve the most relevant and suitable content?

The following sections will further elaborate on and clarify the proposed model and the related work.

Related Work

According to Jansen, Web search engines are an important tool for many people to locate online information or services (B. J. Jansen & Spink, 2006.). As the Web search engines are used to filter results from the Web, recommender systems are used to filter results from smaller-scale data mainly from a website. A recommender system is a software tool and technique that provides suggestions for items or content to a user to achieve personalization (Ricci, Rokach, Shapira, & Kantor, 2011). Popular websites such as Amazon and Netflix employ recommender systems to facilitate personalization for each user (Kembellec, Chartron, & Saleh, 2014).

A number of researchers have adopted this concept by embracing recommender systems in MOOCs, where users select courses that better fit their interests in the MOOC itself. Bousbahi and Chorfi (Bousbahi & Chorfi, 2015) propose a MOOC RS (MOOC-Re) by using the CBR system (Cased-Based Recommender System). Yunchou and Hongkun (Li & Li, 2017) proposed a recommender system for MOOC websites based on user behavior and built on course-based collaborative recommendation. Moreover, they developed a correlated pattern-based adaptation for the recommender system recommendation by combining both user-based clustering and course-based clustering. In addition, other researchers further proposed that MOOCs need to consider the time factor; by empowering real-time adaptivity (Beel, 2017; Pardos, Tang, Davis, & Le, 2017).

Proposing a recommender system for MOOCs may achieve personalization for each MOOC and even reduce the flood of materials but would not overcome the 'lost in hyperspace' issue. Many online tools have already adopted the concept of a unified MOOCs search engine such as: Class Central, MOOCs List and MOOC.org.

However, the obtained results lack personalization since these approaches rely on users' manual customization to achieve customized service. Although Class Central offers personalized and up-to-date results, better personalization is still needed.

Research efforts in this area have led to Courserush (Lee, Girish, & Kim, 2017) and Courducate (Cheng & Gao, n.d.). These are both MOOC search engines that aim to crawl the Web to list a set of different MOOCs in one place.

Courserush focuses on building search models based on open-source tools, and then comparing them with their proposed model results. The main struggle was gathering data and building data sets as well as finding the proper indexing and ranking function. Courserush indexing is based on URLs, tags, and descriptions.

Courducate, on the other hand, uses a specifically developed search engine and algorithms to achieve high-ranking results. The Courducate system gathered two functionalities, multi-site search and multi-field search that includes the university and tutor information; which achieved customized results. Both of these projects are yet to introduce any further personalized outputs.

On the other hand, MoocRec (Symeonidis & Malakoudis, 2016) adopts a similar hybrid approach to the one proposed by this research. This approach integrates a search engine that automatically retrieves data using web content mining techniques from MOOCs. It also implements a recommender system based on content filtering and a Matrix Factorization model to adopt personalization. Although MoocRec achieves personalization, it still needs testing, and it needs to improve accuracy in its recommendations. This would ensure that each user gets the correct tailored recommendations and fully personalized experience (see Table 1).

Table 1
Similar Hybrid MOOCs Retrieving Approaches

Academic Research	Proposed model	Used tools	Obstacles and Shortcomings
Courserush (2017)	Focuses on using and comparing different open-source tools.	• Manual data scraping. • Whoosh. • Haystack on Django. • BM25 ranking function.	• User query was not perfected, with incapability of further improvement (Lee, Girish, & Kim, 2017) • Manual data scraping (at a time), resulting expired data set. • Personalized outputs. • Retrieve information only from Edx, Udemy, and Coursera. Including only course titles and URLs.
Courducate	Uses a specifically developed search engine and algorithms to achieve high-ranking results.	• PhantomJS crawling toolkit ("PhantomJS - Scriptable Headless Browser," n.d.) & manual data scraping. • BM25 ranking function. • Self-built ranking function. • Apache Lucene for indexing.	• Personalized outputs. • Retrieve information only from Edx, Udemy, Udacity, Coursera, and Khan Academy.
MoocRec (2016)	Based on Matrix Factorization model for searching, combined with content-based filtering algorithm for personalization.	• Matrix factorization technique (Koren & Yehuda, 2008) that uses information from several resources/matrices.	• Testing the accuracy of recommendations (Symeonidis & Malakoudis, 2016). • Up to date results (since 2016) • Retrieve information only from Edx and Coursera. Moreover, results are not matching the user query. • Poor and non-friendly user interface.

Therefore, the MRSE approach proposed in this paper aims to consider and overcome the lacks and challenges of the previously mentioned MOOCs research work including personalization, recommendation accuracy and real-time factor.

The MRSE Model and Methodology

In order to solve the 'lost in hyperspace' problem in MOOCs, this research proposes a hybrid personalized search model which is called MOOCs Recommendation Search Engine (MRSE). It aims to utilize the recommender system and the search engine to provide a personalized MOOC service. This MRSE model has two main goals: *firstly*, to help learners reach MOOC courses in one place (a one-stop-shop for MOOCs). *Secondly*, to recommend users the courses that suit their needs based on their user profile resulting from the UM process. The MRSE model design comprises of two stages:

1. Deploying the Search Engine: The search engine retrieves information by ***a)*** capturing, or crawling the data on the Web, and providing results in minimal response time to the submitter, and subsequently ***b)*** indexing the Web pages as a data structure and retrieving the required content (Croft, 2015).
2. Deploying the Recommender System: where filtering algorithms, collaborative filtering in this case; are applied to the retrieved information to employ optimal personalization.

Figure 2 below provides a detailed view of the MRSE model.

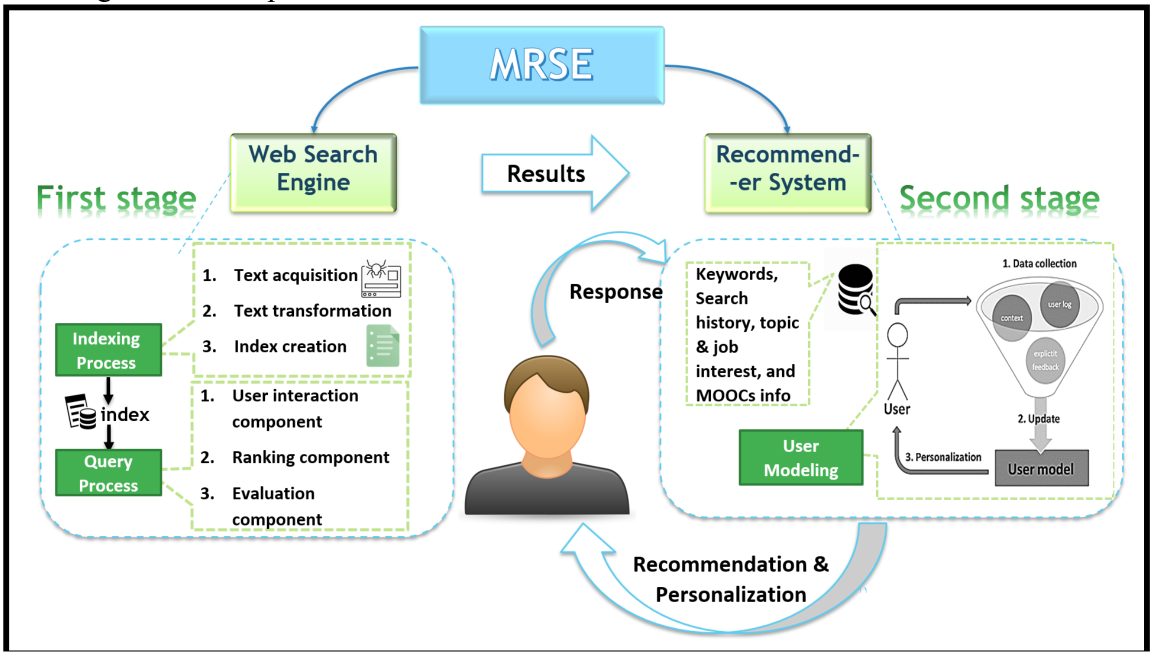

Figure 2 The MRSE Model

The MRSE Architecture

Based on the literature review limitations and the search aims, the model architecture design considers the following:

- Availability (open-sourced tools).
- Integration point of view.

Table 2 and Figure 3 illustrates the chosen tools and their capability to achieve the model goals.

Table 2
MRSE Tools and Capabilities

Tools	Description	Capabilities
1 Apache Nutch (“Apache Nutch TM,”)	Apache Web crawler relies on Apache Hadoop data structures.	• Open-source. • Enables fine-grained configuration. • Crawl the web in an automated manner. • Offers indexing for Apache Solr.

Tools		Description	Capabilities
2	Apache Solr (“Apache Solr,”)	Apache full-text searching and indexing tool.	• Open-source. • Real-time indexing. • Dynamic Clustering. • Database integration.
3	Apache Mahout (“Apache Mahout,”)	Apache project to produce machine learning algorithms. It can offer Recommendation, Classification, and Clustering.	• Open-source. • Offers Collaborative filtering.

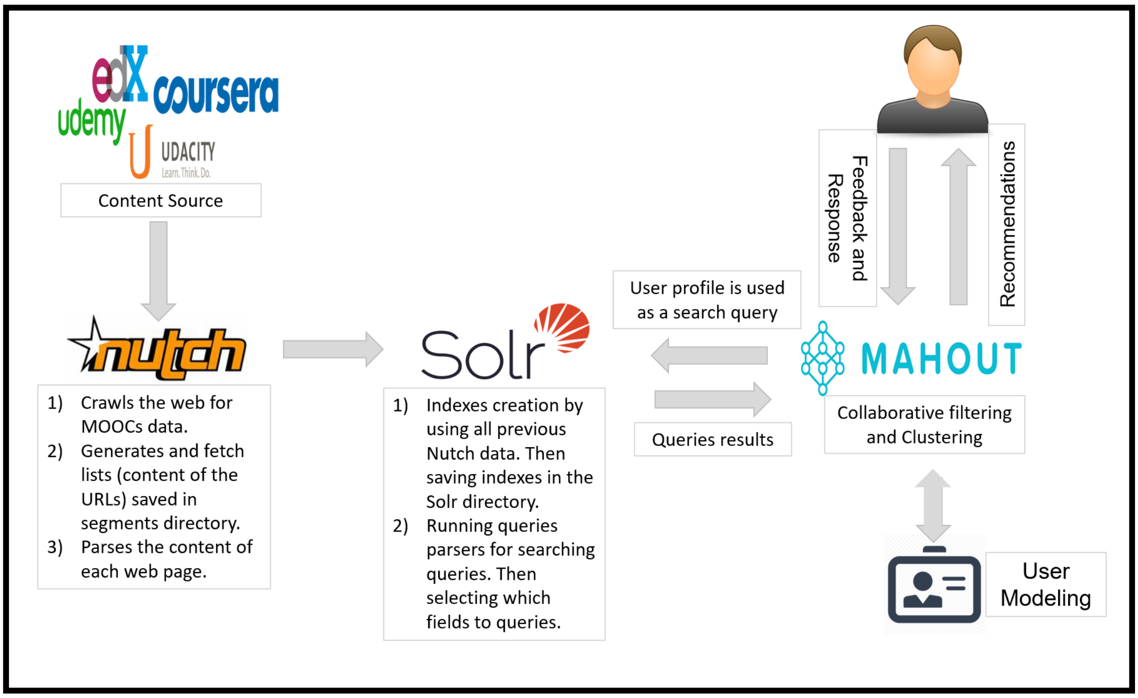

Figure 3. The MRSE Architecture

Discussion

Use of a search engine to find courses and a recommender system to produce personalized recommendations achieves short response time and fast processing, where they are powerfully optimized to retrieve new and relevant content quickly. Therefore, Solr in this case, along with Nutch fast fetching and parsers; will index "real-time indexes" and run queries on large sets of data. These will then crawl multiple MOOCs and deal with over-growing datasets. Mahout's collaborative filtering and clustering will help Solr to recommend the optimum results for users based on their user queries.

Currently, the search leans towards building the search engine (Nutch-Solr platform) where the parsing phase is still undergoing and the scraped data are examined to create clusters later.
Extracting data should be done under "site-specific inspection" (Cheng & Gao; Lee & Girish, 2017). Hence, the main goal in this phase is ensuring that Nutch scarps and crawls automatically only the MOOCs sites, in order to overcome the manual data scraping and manual data sets creation.

The research still needs to further investigate the clustering phase and user profile creation. Primarily, user modeling, the user model and personalization, will be constructed based on users' keywords, search history, topic & job interests, course info, rates, reviews, and MOOCs info to present a tailored user experience.

Challenges

The research must consider some challenges to ensure the system performance and accuracy. Those challenges are illustrated based on the literature review bottlenecks and obstacles:

- Ensure that Solr indexing and ranking is highly efficient, compared with the used BM25 ranking function.
- Build Mahout over Nutch and Solr platform.
- Create optimum clusters and user profiles.
- Ensure the accuracy of recommendations and the system performance.
- Find the optimum testing methods to measure the system.

Primarily, the system's performance will be measured based on (Croft, 2015):

  - Response Time: The delay between submitting a query and receiving the results.
  - Coverage: The speed of incorporating new data into the indexes.
  - Recency: The age of the stored information.

Conclusion

The connection between education and the labor market in Saudi Arabia needs balancing through continuous knowledge enhancement via lifelong learning. MOOCs present a significant contribution in enhancing knowledge, skills and expertise. However, users need more guidance to be able to reach the most suitable courses for them to avoid the 'lost in hyperspace' issue between all the different MOOCs. Having too many MOOCs with a large number of courses and too many results is a limiting issue, especially for recent graduates. Hence, there is a need for a unified searching tool that provides the best of both worlds; one that presents results from different sources (MOOCs) and personalizes these results as well. Therefore, this research proposes MOOCs Recommender Search Engine (MRSE) which combines in the search engine, both content-based filtering and a recommender system. Hopefully, this approach will make it easier for learners to use the MOOC courses efficiently and have better employment opportunities in the Saudi labor market. The current challenge of this model is gathering internal course data from different MOOCs, notably data pertaining to computer science and IT courses from Edx ("edX"), Coursera ("Coursera"), and Udacity ("Udacity Week"). Currently, the model proposed in this research is under implementation. Moreover, it is yet to be employed to assess and report its effectiveness and efficiency.

References

Apache Mahout. (n.d.). Retrieved December 28, 2018, from https://mahout.apache.org/

Apache Nutch™ -. (n.d.). Retrieved December 28, 2018, from http://nutch.apache.org/

Apache Solr -. (n.d.). Retrieved December 29, 2018, from http://lucene.apache.org/solr/

Baturay, M. H. (2015). An Overview of the World of MOOCs. *Procedia - Social and Behavioral Sciences*, *174*, 427–433. https://doi.org/10.1016/J.SBSPRO.2015.01.685

Beel, J. (2017). It's Time to Consider "Time" when Evaluating Recommender- System Algorithms [Proposal], 1–5. Retrieved from https://arxiv.org/ftp/arxiv/papers/1708/1708.08447.pdf

Bousbahi, F., & Chorfi, H. (2015). MOOC-Rec: A Case Based Recommender System for MOOCs. *Procedia - Social and Behavioral Sciences*, *195*, 1813–1822. https://doi.org/10.1016/j.sbspro.2015.06.395

Brusilovsky, P., & Millán, E. (n.d.). *User Models for Adaptive Hypermedia and Adaptive Educational Systems*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.5703&rep=rep1&type=pdf

Cheng, Q., & Gao, Y. (n.d.). *Courducate-An MOOC Search and Recommendation System*. Retrieved from http://www.cs.virginia.edu/~hw5x/Course/IR2015/docs/Projects/Samples/2.pdf

Cohen, W. W., & Fan, W. (2000). Web-collaborative filtering: recommending music by crawling the Web. *Computer Networks*, *33*(1–6), 685–698. https://doi.org/10.1016/S1389-1286(00)00057-8

Croft, W. B. (2015). Search Engine book.

De Pessemier, T., Vanhecke, K., Leroux, S., & Martens, L. (2015). Combining collaborative filtering and search engine into hybrid news recommendations. *CEUR Workshop Proceedings*, *1542*, 14–19. https://doi.org/10.1109/IVS.2002.1188024

Ernst & Young. (2015). How will the GCC close the skills gap?, 30.

European Commission (DG EAC). (2013). Commission launches ' Opening up Education ' to boost innovation and digital skills in schools and universities, (September), 90–92.

International Labour Organization. (2018). *World Employment Social Outlooks: trends for 2018*. https://doi.org/ISBN 978-92-2-129260-9

Jansen, B. J., & Spink, A. (n.d.). How are we searching the World Wide Web? A comparison of nine search engine transaction logs. https://doi.org/10.1016/j.ipm.2004.10.007

Jansen, D. ;, & Lizzie, K. (2018). *The 2018 Open up Education Trend Report on MOOCs*. The Netherlands . Retrieved from www.eadtu.eu%7Cwww.openuped.eu

Kembellec, G., Chartron, G., & Saleh, I. (2014). Recommender systems. *Recommender Systems*, 1–232. https://doi.org/10.1002/9781119054252

Koren, Y., & Yehuda. (2008). Factorization meets the neighborhood. In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD 08* (p. 426). New York, New York, USA: ACM Press. https://doi.org/10.1145/1401890.1401944

Lee, Seungchul; Girish, Rishi; Kim, Y. U. (2017). COURSERUSH A MOOC SEARCH ENGINE My, *91*, 399–404.

Li, Y., & Li, H. (2017). MOOC-FRS: A new fusion recommender system for MOOCs. *Proceedings of 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference, IAEAC 2017*, 1481–1488. https://doi.org/10.1109/IAEAC.2017.8054260

Nelson, T. (2006). Lost in hyperspace [4]. *New Scientist*, *191*(2561), 26. https://doi.org/10.1002/elsc.200620112

Online Courses - Learn Anything, On Your Schedule | Udemy. (n.d.). Retrieved January 11, 2019, from https://www.udemy.com/?utm_source=adwords-brand&utm_medium=udemyads&utm_campaign=NEW-AW-PROS-Branded-Search-World-EN-ENG_._ci__._sl_ENG_._vi__._sd_All_._la_EN_._&tabei=7&utm_term=_._ag_48933380294_._ad_279519253635_._de_c_._dm__._pl__._ti_kwd-310556426868

Ossiannilsson, E., Altinay, F., Altinay, Z., & Albright, J. (2016). MOOCs as Change Agents to Boost Innovation in Higher Education Learning Arenas. https://doi.org/10.3390/educsci6030025

Otter, M., & Johnson, H. (2000). Lost in hyperspace: metrics and mental models. *Interacting with Computers*, *13*(1), 1–40. https://doi.org/10.1016/S0953-5438(00)00030-8

Pardos, Z. A., Tang, S., Davis, D., & Le, C. V. (2017). Enabling Real-Time Adaptivity in MOOCs with a Personalized Next-Step Recommendation Framework. In *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale - L@S '17* (pp. 23–32). New York, New York, USA: ACM Press. https://doi.org/10.1145/3051457.3051471

PhantomJS - Scriptable Headless Browser. (n.d.). Retrieved December 29, 2018, from http://phantomjs.org/

Ricci, F., Rokach, L., & Shapira, B. (2011). Introduction to Recommender Systems Handbook. In *Recommender Systems Handbook* (pp. 1–35). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-85820-3_1

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Eds.). (2011). *Recommender Systems Handbook*. Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-85820-3

Saltzman, G. M. (2014). The Economics of MOOCs. *The NEA 2014 Almanac of Higher Education*, 19–29. https://doi.org/doi:10.1201/9781420051254.ch5

Shah, D. (n.d.-a). A Product at Every Price: A Review of MOOC Stats and Trends in 2017 | EdSurge News. Retrieved October 18, 2018, from https://www.edsurge.com/news/2018-01-22-a-product-at-every-price-a-review-of-mooc-stats-and-trends-in-2017

Shah, D. (n.d.-b). By The Numbers: MOOCS in 2017 — Class Central. Retrieved October 29, 2018, from https://www.class-central.com/report/mooc-stats-2017/

Shah, D. (n.d.-c). The Second Wave of MOOC Hype is Here and it's Online Degrees — Class Central. Retrieved December 29, 2018, from https://www.class-central.com/report/second-wave-of-mooc-hype/

Sunar, A. S., Abdullah, N. A., White, S., & Davis, H. (2016). Personalisation in MOOCs: A critical literature review. *Communications in Computer and Information Science*, *583*, 152–168. https://doi.org/10.1007/978-3-319-29585-5_9

Symeonidis, P., & Malakoudis, D. (2016). *MoocRec.com : Massive Open Online Courses Recommender System*. Retrieved from http://ceur-ws.org/Vol-1688/paper-01.pdf

Teaching, D. (2017). BizMOOC Discussion 01 - Existing MOOC initiatives in higher education and business sector and the distribution of MOOC ... BizMOOC Discussion 01 Existing MOOC initiatives in higher education, *2020*(February), 2014–2016.

Vision 2030. (2017), *91*, 399–404. Retrieved from www.vision2030.gov.sa

Xiao, J., Wang, M., Jiang, B., & Li, J. (2018). A personalized recommendation system with combinational algorithm for online learning. *Journal of Ambient Intelligence* *and Humanized Computing*, *9*(3), 667–677. https://doi.org/10.1007/s12652-017-0466-8

Zvacek, S., Restivo, M. T., Uhomoibhi, J., & Helfert, M. (2016). Computer Supported Education: 7th International Conference, CSEDU 2015 Lisbon, Portugal, May 23-25, 2015 Revised Selected Papers. *Communications in Computer and Information Science*, *583*(November 2017). https://doi.org/10.1007/978-3-319-29585-5