

## Supporting Students through a Recommendation System for Knowledge Acquisition in MOOCs Ecosystems

Rodrigo Campos<sup>1,3</sup>, Rodrigo Pereira dos Santos<sup>2</sup> (coorientador),  
Jonice Oliveira<sup>1</sup> (orientadora)

<sup>1</sup>PPGI – Universidade Federal do Rio de Janeiro (UFRJ)  
Av. Athos da Silveira Ramos, 274 – Cidade Universitária – Rio de Janeiro, RJ – Brasil

<sup>2</sup>PPGI – Universidade Federal do Estado do Rio de Janeiro (UNIRIO)  
Av. Pasteur, 458 – Urca – Rio de Janeiro, RJ – Brasil

<sup>3</sup>Instituto Federal de Educação, Ciência e Tecnologia do Rio de Janeiro (IFRJ)  
R. Buenos Aires, 256 – Centro – Rio de Janeiro, RJ – Brasil

rodrigo.campos@ufrj.br, rps@uniriotec.br, jonice@dcc.ufrj.br

**Abstract.** *With the growth of MOOCs (Massive Open Online Courses), students face difficulties in choosing suitable courses to meet a knowledge demand. Some recommendation systems have proposed solutions, but not exploring the student's prior knowledge. In this context, this work contributes to identifying and reducing the students' knowledge gap in MOOCs. To do so, we model and analyze the MOOCs ecosystems and propose a solution for recommending parts of courses to students, exploring the software ecosystem approach in the educational domain. After evaluating our solution based on three experiments, we observed that our recommendations present new content to fill users' knowledge gaps, being accurate, useful, and reliable.*

### 1. Introduction

In recent years, the open online education has been widely adopted due to the emergence of Massive Open Online Courses (MOOCs). These courses offer facilities for students and universities since they have no limit on the number of participants and students can control the course time. In this scenario, MOOCs can be analyzed in an ecological context, composed of a learning environment and a learning community, highlighting interactions between the system and the environment. Moreover, the allowance for opening new courses in different cities considers the quality of physical facilities, faculty, and institution goals, and they often depend on internal laws. However, public and private institutions may not be so fast to adapt or create new courses. MOOCs reduce the limitations of formal courses, as the entire learning process happens via the Internet, reaching several territorial regions not served by governments.

Because of these benefits, in recent years, there has been an increase in the number of MOOCs with a plethora of courses distributed in different providers. Consequently, students face difficulties in choosing the best courses to meet knowledge gaps. They want to acquire specific knowledge, but there is no guidance on which courses are the most appropriate. Some recommendation systems have emerged to address this problem. However, as shown in Section 2, they have several limitations in considering the users' prior knowledge and options of courses' modules to reach their learning needs. As such,

the problem investigated in this work arises: how to help users to achieve their own specific goals reducing their knowledge gaps, i.e., acquire new knowledge according to their interests. In this Master's thesis [Campos, 2019], we propose solutions to this problem based on the following main research question (RQ):

*RQ1*: How to identify and reduce knowledge gaps in the MOOCs ecosystems?

Moreover, this work investigates the following alternative research questions:

*RQ2*: What are the existing works about recommendation systems applied to MOOCs?

*RQ3*: What are the main challenges in recommendation systems applied to MOOCs?

*RQ4*: What are the main actors in the MOOCs ecosystems, and how they relate to each other?

The main objective of this work is to reduce users' knowledge gaps. Therefore, it is investigated how to support MOOCs' users, combining recommendation systems with data from MOOCs providers based on course modules (or full courses, if providers do not break them into learning units). We also investigate the MOOCs ecosystems characteristics, exploring how this perspective supports the basic processes of providers. Given the above objective, the methodology consists of the following steps:

- a) Conducting a literature review based on a Systematic Mapping Study (SMS) to map tools, techniques and works that address recommendation systems in the MOOCs scenario;
- b) Modeling the MOOCs domain based on the software ecosystem approach and then specifying and implementing a system to recommend courses, or course modules, to users. In this step, we define the techniques used in each process of the conceptual model of our solution, such as data treatments (e.g., extraction, storage, unification), topic modeling and labeling, and recommendation;
- c) Evaluating the solution from a quali-quantitative comprising three experiments. The first is focused on verifying the coherence of the proposed topic modeling technique, the second evaluates the applied topic labeling techniques, and the third evaluates the recommendation system in a simulation of the real-world.

## 2. Related Work

The first phase of our methodology contemplates the execution of an SMS to analyze how researches have evolved over the years and to identify opportunities that remain open in this field. We summarized features of works that addressed recommendations in MOOCs scenarios applying topic modeling (also applied in our solution), as shown in Table 1; we referred to our solution as “RS” as the acronym for “Recommendation System”. Given the fact that Apaza et al. (2014) and Bhatt et al. (2018) – works that are directly related to our solution – do not allow access to the source code or datasets, it was not possible to perform a comparison on efficiency and efficacy. Nonetheless, the comparative analysis in Table 1 allows us to highlight different aspects between them, as detailed in Campos (2019). Some works are inserted in scenarios where interaction between users can be extract and analyzed, being different from our solution. In this context, our recommendation system uses the content-based approach.

From the previous works, we observe advances in our work regarding the treatment of student's profiles from multiple MOOCs platforms aiming at the reduction of their knowledge gap. This is possible because our new method extracts providers' data from each Application Programming Interface (API). Another novelty of our solution is the recommendation itself and adopted techniques, besides the recommendation output, i.e., what is recommended to users, since our recommendation system considers parts of course and delivers packages of personalized modules based on the users' knowledge gap. In addition, it is relevant to highlight the fact that we model and analyze the target context as MOOCs ecosystems based on the software ecosystems approach to balance the MOOCs environment and strengthen interactions.

**Table 1. Comparison between related work and the recommendation system (RS) proposed in our research. Source: [Campos, 2019]**

		[Apaza et al., 2014]	[Wang et al., 2015]	[Jing and Tang, 2017]	[Song et al., 2017]	[Bhatt et al., 2018]	[Li et al., 2018]	RS
<b>Recommendation Approach</b>	Collaborative Filtering			X	X		X	
	Content-Based with LDA <sup>1</sup>	X				X		
	Topic Specific		X					
	Content-Based with NMF <sup>2</sup>							X
<b>Output</b>	Threads				X			
	Open Education Resources (OER)		X					
	Videos					X		
	Courses	X		X				
	Tags and Courses						X	
	Courses and its parts (module, relevant content)							X
<b>Type</b>	System	X	X			X		X
	Network						X	
	Algorithm Framework			X				
	Learning Assistance				X			
<b>Provider</b>	XuetangX			X			X	
	Coursera	X			X			
	Multiple		X			X		X

<sup>1</sup> Latent Dirichlet Allocation

<sup>2</sup> Non-negative Matrix Factorization

### 3. Proposed Recommendation System

We define the conceptual model of our solution in Figure 1, raising the MOOCs stakeholders and relating them to the software ecosystem's actors and roles [Barbosa et al., 2013]. The process starts when a user (1) accesses the Web-Based Recommendation System (2). To make the recommendation, the system requests access to user data through the authentication layer (3). Next, it is possible to access the user Knowledge Base (4) that holds information from the ecosystem's Input Data (5). Input data (5) contain authorized information (e.g., user's competencies), curriculum and other data from MOOCs providers. It represents possible interactions that such user has in the ecosystem, which contemplates the actors that interact with students.

Next, the system selects the history of the user and create a set of documents (6). This data is input to the user-item matrix (7). In this matrix (7), the "user" is composed of documents from the user's knowledge base (4), while the "item" is created based on data from the Background Data (8) layer. This layer has data from different MOOCs providers (9) to allow a broader recommendation. With the providers' data selected, the topic modeling method (10) creates the item topics (11) which complete the "item" in the matrix.

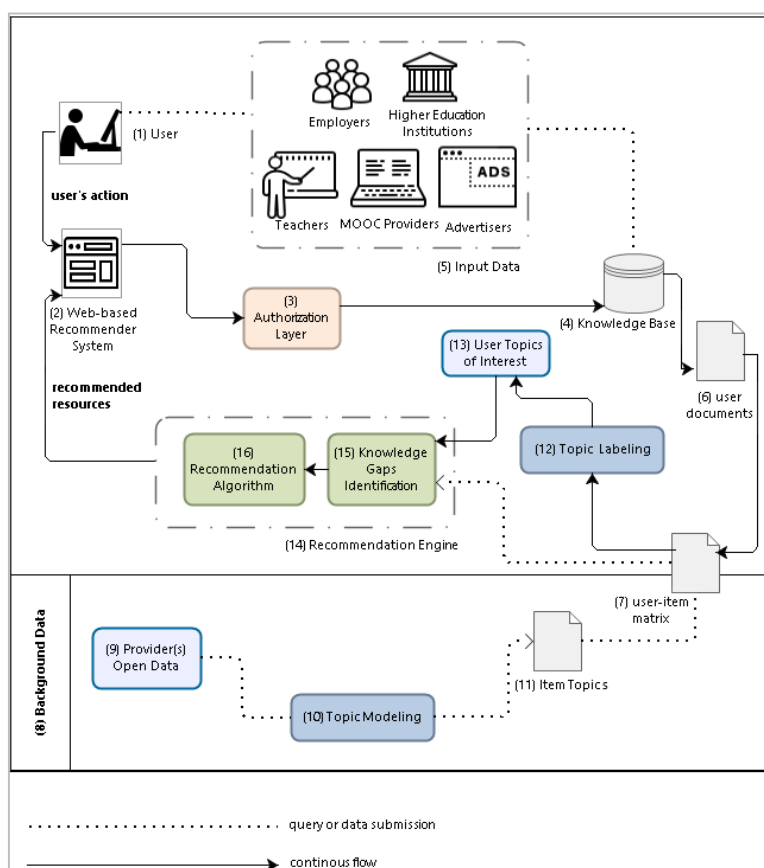


Figure 1. Conceptual model of the proposed solution. Source: [Campos, 2019]

This matrix (7) is input to the techniques for labeling topics (12). In our approach, labels help to define item topics (11), but also to create the user's topics of interest (13), i.e., the most related topics to the user. After creating the topics and labels, the recommendation engine layer (14) applies two procedures based on the user-item matrix. First, the Knowledge Gaps Identification (15) collects the user's existing and desired

skills and identifies the knowledge gap. Then, the Recommendation Algorithm (16) is applied to find similarities among item's topics (11) and user's documents (6). Finally, it ranks results and sends resources (recommendations/user's topics of interest) back to the system (2).

To implement our solution, we extracted data from three providers selected by open data availability and other factors: Khan Academy, Udemy, and edX. The cooperation between these independent entities is possible since we use the software ecosystem approach [Barbosa et al., 2013]. In this scenario, our conceptual model can be extended by other developers who can add other providers. After the extraction, we stored such data (documents) in our recommendation system using MongoDB (i.e., a document-oriented database). We chose this database since it integrates data in JSON format, finding the BSON – binary that takes up less space and is faster.

Next, we implemented other two algorithms. The first one models the topics using the Non-negative Matrix Factorization (NMF). Instead of manually indicating the number of topics in the model, as in the most used NMF, we implemented the stability analysis proposed by Greene et al. (2014) apud Nolasco (2016) to automatically find the ideal number of topics  $k$ . The models are tested with a minimum ( $kmin$ ) and maximum ( $kmax$ ) range for  $k$  until the identification of which value has the best topic coherence. Therefore, there is no need for a human decision. Topic coherence semantically quantifies the relatedness between terms in a topic, i.e., how much those terms are related. Despite the modeling, these topics are not always easy to understand. Therefore, the second algorithm implements the topic labeling that seeks to select a word (called as label) to express the theme or topic area. We selected two techniques for topic labeling: Text Selection (TS) and Keywords Selection (KS). In the evaluation phase (Section 4), we compared which one better represents topics and ensures a better labeling process.

In the content-based recommendation method implementation, we apply NMF to cross-reference the user's documents (Figure 1, see 6) and the item topics (Figure 1, see 11), creating the document-terms matrix and the “user topics of interest”. Next, we joined the user documents into a single search string that represents the knowledge obtained or enrolled by the user. With this information, we use the Euclidean distance between this search string and the other documents in the item model, so it is possible to recommend other documents related to the user. The closest documents to the user profile are the most recommended ones; to identify them, we sorted the Euclidean distance results ordering by the smaller distances. A set of documents  $s$  related to the topic and represented by merged information (module title, module URL, and others) is extracted. Next, we apply an algorithm to verify what documents of the user layer are in the set of extracted documents  $s$  excluding them from the final list (verify 'novelty'). Finally, the top-6 of the list is displayed back to the user.

#### 4. Evaluation and Results

We conducted a set of experiments with data collected from MOOCs providers (a total of 106,574 modules). In the first experiment to verify the effectiveness of the topic modeling technique, such dataset was applied in two scenarios: a) the first with our topic modeling approach using the NMF technique; and b) the second that is a more traditional approach based on LDA (our baseline). Next, we compared results by checking the topic coherence values for a stability analysis ranging from 5 to 30 topics in both techniques

(LDA and NMF). The steps to model the item topics were followed so that this data can be clustered according to the similarity between them. The results showed that the best coherence was 0.32 in the baseline using LDA, i.e., less than the best coherence of 0.36 using our approach, indicating that our solution is relevant to be applied in the recommendation as it better represents the item layer.

In the second experiment, we evaluated the representativeness of the labeling technique. We generated labels using our approach (with TS, its variations, and KS) and compared them with MOOCs providers' labels. Considering the cosine distance between strings in each defined technique and the provider's manual labels, it is possible to compare and select which technique gets better results, i.e., within the closest proximity. The TS approach is chosen by the reason that the extracted modules do not have keywords. In this context, we selected the fast keyword extract algorithm. Nonetheless, we also generated labels using the KS approach by selecting the keywords from the courses in which the modules are allocated. In our algorithm to generate labels, we extract primitive labels by selecting the top-30 documents linked to each topic in which each document is equivalent to a module. After applying some text treatments, we extracted the top-10 terms for each topic and follow different steps to TS and KS.

The comparison was made by the selection of providers labels, more specifically from the top-30 document of each topic. We calculated the cosine distance between the TS (top-1), TS (top-3), and KS strings relative to the provider strings. Results pointed to the closer proximity of our TS (top-3) to provider labels (in 12 of 14 tested topics), stating that our approach can automatically generate labels for the topics modeled in our scenario. The biggest advantage with this technique is the automatic identification for describing item layer topics, but we can also consider that it is useful to the "user topics of interest" of the recommendation process (Figure 1).

Lastly, the third evaluation was an experiment involving the user's perspective. We collect explicit feedback from participants about recommendations using the metrics Main Average Precision (MAP), Utility, Novel, and Confidence based on the properties specified in Ricci et al. (2015) to evaluate the quality/performance of recommendations. When analyzing the metrics, it is important to highlight that the average precision (AP) is part of the precision-oriented metrics that consider the ranked list of items and the concept of precision, which in recommendation systems apply as the fraction of TOP N recommended items that are relevant to the user. AP is higher as the recommendation has more hits. Moreover, the higher these hits are, the higher AP is. Differently, the utility indicates the value that the user or system gets from the recommendation [Ricci et al., 2015], which in our specific case is the usefulness of the recommendations for users (students).

The novelty metric indicates recommendations for items that the user does not yet know and that can be filtered in several ways [Ricci et al., 2015]. In our case, the collection is made explicitly by asking the user, but the information is also confirmed in the user's ecosystem. Given that these modules belong to the providers adopted in this evaluation, it is possible to check the existence of the participant's enrollment in this module through the database. The last metric is confidence, which refers to how reliable the system can be in its recommendations. In this case, the calculation is given by the ratio between all positive evaluations and the total number of valid evaluated modules. So, we defined the hypotheses of the experiment:

- a) **Null hypothesis (H0):** The proposed recommendation method achieved efficacy of less than 50% in the properties of MAP, Utility, Novel, or Confidence.

**H0:** ( $\mu\text{MAP\_OurApproach} < 50\%$ ) OR ( $\mu\text{Utility\_OurApproach} < 50\%$ ) OR ( $\mu\text{Novelty\_OurApproach} < 50\%$ ) OR ( $\mu\text{Confidence\_OurApproach} < 50\%$ ), where:

$\mu\text{MAP\_OurApproach}$  = MAP of the feedback collected for our recommendation system

$\mu\text{Utility\_OurApproach}$  = Utility of our recommendation system collected through user feedback

$\mu\text{Novelty\_OurApproach}$  = Novelty of our recommendation system collected through user feedback

$\mu\text{Confidence\_OurApproach}$  = Confidence of our recommendation system collected through user feedback

- b) **Alternative hypothesis (H1):** The proposed recommendation method achieved efficacy greater than 50% in the properties of MAP, Utility, Novel, and Confidence.

**H1:** ( $\mu\text{MAP\_OurApproach} \geq 50\%$ ) AND ( $\mu\text{Utility\_OurApproach} \geq 50\%$ ) AND ( $\mu\text{Novel\_OurApproach} \geq 50\%$ ) AND ( $\mu\text{Confidence\_OurApproach} \geq 50\%$ ).

We developed a web system based on Python 3.6 and a set of web technologies to collect answers of each participant about the top-6 recommended modules or courses, as well as information to characterize the participants. A total of 27 people was invited to the study, of which 19 completed all the required steps. We consider that each participant corresponds to a specific recommendation subset. As such, each participant had to answer the question “Would you find it relevant to learn this content?” to each module through the web system to calculate MAP and confidence. The options were Yes (relevant content) or No (no relevant content).

The question “How useful would this content be for you?” was asked to participants to calculate utility. Five levels of responses are used as options (Likert scale). To calculate novelty, participants had to answer the question “Have you learned this course before?” with possible answers “Yes” or “No”. Results indicated MAP = 62.24%, Utility = 68.89%, Novelty = 99.12%, and Confidence = 72.81%, i.e., efficacy greater than the level established in the alternative hypothesis. Therefore, it was possible to verify that our proposed recommendation system for MOOCs ecosystems is an accurate, useful, reliable tool that presents new content to fill the knowledge gap of users within such ecosystem.

From the results described in this section, it was possible to verify that the proposed process answer our main research question RQ1 since it demonstrates that: a) our data integration method and the software ecosystem approach allow the unification of data from multiple MOOCs providers that serve as input to our recommendation system; b) the proposal of a modified NMF approach allows the detection of similarity between these data for the recommendation and it also contributes to the system scalability by automatically defining the ideal number of topics; c) topic labeling techniques can be used in the MOOCs ecosystems scenario to identify labels automatically (instead of manual labels from providers) and it also helps in reducing the

students' knowledge gap by identifying their interest; and d) when applying these structures/components in our web-based recommendation system in an integrated way, it was possible to support students in reducing their knowledge gaps by identifying and proposing parts of courses that correspond to what are expected from them.

## 5. Contributions

This Master's thesis covers relevant contributions for Computers in Education, as we investigate the context of technologies applied to MOOCs. Regarding the challenges in the field, this Master's thesis contributes to "Connected Data in Education" (translated by the authors), more specifically about "publishing data" [Pereira et al., 2017], as proposed in DesafIE!<sup>3</sup>. Pereira et al. (2017) highlight the need to address the specificities of different target domains based on the integration/interoperability between different vocabularies. This work is also related to "Challenges for Finding and Retrieving Information in Audiovisual Educational Content" [Silveira and Borges, 2017], which suggests the creation of models that represent the user's style and preferences in educational audiovisual content.

Besides the contributions mentioned in Section 2, it is possible to include the reduction in the knowledge gap and the content-based recommendation method itself. In addition, we point the method for extracting data from multiple providers that can be reused in other applications and the results of the secondary study (SMS). Our solution can be applied in the process of structuring and boosting the social and economic dynamics of a city, helping to identify knowledge demands. In the current context in which education in several countries has been affected by COVID-19, our solution becomes even more relevant.

Moreover, the problem is included in the challenges "Information Ecosystem Development" and "Open and Collaborative Processes Ecosystems" from the Grand Research Challenges in Information Systems in Brazil [Araujo, 2016]. It is also related to the challenge "Participative and universal access to knowledge for the Brazilian citizen" in Grand Challenges in Computer Science Research in Brazil [Salgado et al., 2015]. Moreover, this Master's thesis brings some relevant contributions to Software Engineering, especially to the field of Software Ecosystems, on domain-specific ecosystems and open platform architecture. Our solution involves the topic modeling technique, which is an unsupervised machine learning technique. It is important to highlight that machine learning is aligned to the Artificial Intelligence, which is one of the priorities of MCTIC concerning "research projects, development of technologies and innovations for the period 2020 to 2023" (Art. 4)<sup>4</sup>.

The knowledge produced in this Master's thesis was disseminated in the national conferences: [Campos et al., 2018a] and [Campos et al., 2020a] (Capes Qualis CC B2), and in a short course (and book chapter) at the ERSI [Marinho et al., 2019]. In addition, it was published in international venues [Campos et al., 2018b] (Capes Qualis CC B1), [Campos et al., 2019] and [Campos et al., 2018c]. Finally, this work was approved for the second phase of the Information Systems Doctoral and Master's Thesis Competition

---

<sup>3</sup> Computer Challenges Workshop on Education

<sup>4</sup> [http://www.mctic.gov.br/mctic/opencms/legislacao/portarias/Portaria\\_MCTIC\\_n\\_1122\\_de\\_19032020.html](http://www.mctic.gov.br/mctic/opencms/legislacao/portarias/Portaria_MCTIC_n_1122_de_19032020.html)



(CTDSI) [Campos et al., 2020b] and contributed to the author being selected to France Excellence Programme launched by Campus France and the French Embassy in Brazil.

## 6. Conclusion

With the advance of MOOCs ecosystems, it might be difficult for students to choose the best courses among all the available options. Returning to our RQ1, we verify that our solution allowed us to support students in the identification and reduction of knowledge gaps. This Master's thesis opens some opportunities for future work, including an expansion of the SMS in order to cover the most recent work and the development of a web interface to the recommendation system based on the suggestions from the third experiment. This interface would trigger the recommendation and authorize access to private data from different providers where a user has accounts. An option would be the extension of the existing web system based on some adaptations.

## Acknowledgment

We thank IFRJ for supporting this research. We also thank CAPES, CNPq, and FAPERJ (Brazil) for their financial support to the research group.

## References

- Apaza, R. G., Cervantes E. V. and Quispe L. C. (2014). "Online Courses Recommendation based on LDA". In: *1st Symposium on Information Management and Big Data*, Cusco, Peru, pp. 42-48.
- Araujo, R. (2016). "Information Systems and the Open World Challenges". In: Boscarioli, C., Araujo, R. M. and Maciel, R. S. P. (eds.), *I GranDSI-BR - Grand Research Challenges in Information Systems in Brazil 2016-2026*. SBC, pp. 42-51.
- Barbosa, O. A. L. P., Santos, R. P., Alves, C. F., Werner, C. M. L. and Jansen, S. (2013) "A Systematic Mapping Study on Software Ecosystems from a Three-Dimensional Perspective". In: Jansen, S., Brinkkemper, S. and Cusumano, M. A. (eds.), *Software Ecosystems: Analyzing and Managing Business Networks in the Software Industry*. Edward Elgar Publishing, pp. 59-81.
- Bhatt, C., Cooper, M. and Zhao J. (2018). "SeqSense: Video Recommendation Using Topic Sequence Mining". In: *International Conference on Multimedia Modeling*, Cham, Switzerland, pp. 252-263.
- Campos, R.** (2019). *Recommendation System for Knowledge Acquisition in MOOCs Ecosystems*. Master's Thesis in Informatics. PPGI/UFRJ, Rio de Janeiro, Brazil.
- Campos, R.**, Santos, R. P. and Oliveira, J. (2018a). "Recommendation Systems for Knowledge Reuse Management in MOOCs Ecosystems". In: *Anais Estendidos - XIV Simpósio Brasileiro de Sistemas de Informação*, Caxias do Sul, Brasil, pp. 146-148.
- Campos, R.**, Santos, R. P. and Oliveira, J. (2018b). "Web-Based Recommendation System Architecture for Knowledge Reuse in MOOCs Ecosystems". In: *IEEE 19th International Conference on Information Reuse and Integration*, Salt Lake City, USA, pp. 193-200.
- Campos, R.**, Santos, R. P. and Oliveira, J. (2018c). "Using Multilayer Social Networks in an Analysis of Higher Education for Professional Demand". In: *I Workshop on Big*

*Social Data and Urban Computing (BIDU)*, 44th International Conference on Very Large Data Bases (VLDB), Rio de Janeiro, Brazil. Aachen: CEUR-WS, vol. 2247.

**Campos, R.**, Santos, R. P. and Oliveira, J. (2019). "A Recommendation System Enhanced by Topic Modeling for Knowledge Reuse in MOOCs Ecosystems". In: Rubin, S. H. and Bouzar-Benlabiod, L. (eds.), *Reuse in Intelligent Systems*. CRC Press, pp. 116-142.

**Campos, R.**, Santos, R. P. and Oliveira, J. (2020a). "A Recommendation System based on Knowledge Gap Identification in MOOCs Ecosystems". In: *XVI Brazilian Symposium on Information Systems (SBSI)*, São Bernardo do Campo, Brazil.

**Campos, R.**, Santos, R. P. and Oliveira, J. (2020b). "Recommendation System for Knowledge Acquisition in MOOCs Ecosystems". In: *Anais Estendidos - XVI Simpósio Brasileiro de Sistemas de Informação*. Porto Alegre: SBC.

Jing, X. and Tang, J. (2017). "Guess You Like: Course Recommendation in MOOCs". In: *International Conference on Web Intelligence*, Leipzig, Germany, pp. 783–789.

Li, C., Song, Z. and Tang, J. (2018). "User Tagging in MOOCs Through Network Embedding". In: *2018 IEEE Third International Conference on Data Science in Cyberspace*, Guangzhou, China, pp. 235–241.

Marinho, L. H., **Campos, R.**, Santos, R. P., Silva, M. F. and Oliveira, J. (2019). "Conceitos, Implementação e Dados Privados de Algoritmos de Recomendação". In: *VI Escola Regional de Sistemas de Informação do Rio de Janeiro (ERSI)*, Rio de Janeiro, Brasil, pp. 06-37.

Nolasco, D. (2016). *Identificação Automática de Áreas de Pesquisa em C&T*. Master's Thesis in Informatics. PPGI/UFRJ, Rio de Janeiro, Brazil.

Pereira, C. K., Siqueira, S. W. M. and Nunes, B. P. (2017). "Dados Conectados na Educação". In: *Workshop de Desafios da Computação aplicada à Educação (DesafIE!)*, São Paulo, Brasil, pp. 740-747.

Ricci, F., Rokach, L., Shapira, B. and Kantor, P. B. (2015). *Recommender Systems Handbook*. Springer.

Salgado, A. C., Motta, C. L. R. and Santoro, F. M. (2015). *Grandes Desafios da Computação no Brasil - Relatos do 3o Seminário*. SBC.

Silveira, I. F. and Borges, P. R. S. (2017). Desafios para a Localização e Recuperação de Informação em Conteúdos Educacionais Audiovisuais. In: *Workshop de Desafios da Computação aplicada à Educação (DesafIE!)*, São Paulo, Brasil, pp. 702-709.

Song, J., Zhang, Y., Duan, K. and Hossain, M. S. (2017). "TOLA: Topic-oriented learning assistance based on cyber-physical system and big data". *Future Generation Computer Systems*, v. 75, pp. 200-205.

Wang, J., Xiang, J. and Uchino, K. (2015). "Topic-Specific Recommendation for Open Education Resources". In: *International Conference on Web-Based Learning*, Cham, Switzerland, pp. 71-81.