Supporting the Retrieval of Open Educational Resources for Programming Education

William Simão de Deus and Ellen Francine Barbosa

[1] Institute of Mathematics and Computer Science - University of São Paulo
ICMC-USP - São Carlos, SP, Brazil

williamsimao@usp.br, francine@icmc.usp.br

Abstract. *In this study, we summarize our ongoing PhD research dedicated to the intersection between Open Educational Resources (OER) and Programming Education. More specifically, we have been researching how programming education can be democratized by using open and free materials, such as open books, open courses, among others materials. For this purpose, we have been adopting the Design Science Research (DSR) in order to investigate and create artifacts for students and teachers of programming to use and reuse OER.*

1. Introduction

Computer programming is an essential skill for many professions. In computing courses, for example, about 40% of the core hours are devoted to teaching and learning of programming and software development [ACM/IEEE 2020]. At the same time, other undergraduate programs have allocated space for introductory programming courses, specially in STEM (Science, Technology, Engineering, and Mathematics) area [Medeiros et al. 2019].

In this sense, it is important to investigate innovative ways to support the teaching and learning of programming. Open Educational Resources (OER) can provide several contributions to this purpose. OER are learning, teaching and research materials in any format and medium [UNESCO 2019]. These resources are released under open licenses, and they are free of charge, allowing the use, distribution, and adaption by others.

According to recent studies, programming OER can promote equity and inclusion, removing costs of teaching materials [Bahamón 2022]. In addition, these materials can make teaching easier and more effective by using resources already available on the Internet [McGill et al. 2022]. In fact, in recent years we have noticed an increase in the number of OER for programming education, such as video lessons, courses, books, websites, tools, applications, among other types of materials [Tovar et al. 2017].

However, there is a challenge entitled *"the discovery problem"* [Wiley et al. 2014]. This problem is caused because we have produced a lot of programming OER, but these resources are "hidden" in digital collections because users have used confusing, ambiguous or irrelevant metadata. As a consequence, current search engines cannot identify relevant programming OER. Students and teachers need to make several attempts to collect a valid result, but most of the time, no OER is found.

Based on the aforementioned context, we have investigated the discovery problem in our PhD work. More specifically, the implications for teachers and students of programming. Next, we characterize the investigated problem and its importance.

DOI: 10.5753/cbie_estendido.2022.226720

1.1. Research Motivation and Relevance

The discovery problem of programming OER is a relevant subject for many reasons. The first one is related to its adherence to achieve a sustainable world [UNESCO 2019]. OER encompasses the sustainable agenda proposed by UNESCO to democratize the access to education. Brazil, as a member state, should promote actions and policies to establish the use and reuse of these materials. Thus, our PhD is aligned with this purpose.

The second reason is the usage barrier. For example, in some areas, OER are considered as being of low quality because they are free [Wiley et al. 2014]. This PhD intends to reduce this barrier by contrasting the use/reuse of OER with traditional teaching materials. At the same time, we intend to evidence the quality of OER by performing empirical validations with students and teachers.

The third reason is summarized in the equity and inclusion provided by OER, as discussed by [Bahamón 2022]. As computer programming is an essential skill for the future, and we need to facilitate the access to free materials by underrepresented or socioeconomically disadvantaged groups.

With regard to innovation, in this PhD work we have proposed innovative strategies to identify and retrieve programming OER. In our PhD, we are investigating practices adopted by teachers, such as the use of relevant filters, user experience, and skills to be developed. In this sense, our PhD differs from similar works because we are reducing the dependence of metadata and manual classifications [Tovar et al. 2017]. In addition, our scope is dedicated to Portuguese and English users, unlike most initiatives [McGill et al. 2022].

Another intended innovation is the establishment of mechanisms for retrieving OER. Currently, OER are available in digital repositories with an outdated interface and simple searches. Instead, our goal is to generate web tools for teachers and students, providing mechanisms based on syntax, semantics, and synonyms.

Based on the aforementioned context, our PhD work intends to solve the following Research Question (RQ): **How can we support the identification and retrieval of programming OER?**

1.2. Research context

This research has been performed under the supervision of Dr. Ellen Francine Barbosa at the Institute of Mathematics and Computer Science - University of São Paulo (ICMC-USP), in São Carlos, Brazil. The research started in August/2019 and the deadline for its completion is September/2023. According to this schedule, we have 14 months to complete our investigation.

This PhD research is supported by the São Paulo Research Foundation (FAPESP - Scholarship #19/26871-4). In addition, our work was approved in qualifying exam and it generated five publications.

In the remainder of this paper we address the main steps of the PhD research: In Section 2 we present the research methodology adopted. In Section 3 we summarize the results. Finally, in Section 4 we synthesize the final remarks.

2. Research methodology

In this PhD, the Design Science Research (DSR) is adopted as research methodology [Pimentel et al. 2019, Dresch et al. 2015]. In short, DSR investigates artificial problems and creates artifacts to solve them. In our PhD, we have investigated an artificial problem (the discovery problem) and we intend to create solutions (mechanisms) to reduce it.

DSR is a type of design-based research method. Therefore, it adopts an iterative and sequential process of six activities. These activities define what is to be done, but the researcher must define how to conduct each one. To structure this process, we have followed the assumptions of [Pimentel et al. 2019] and [Dresch et al. 2015], summarized in Figure 1.

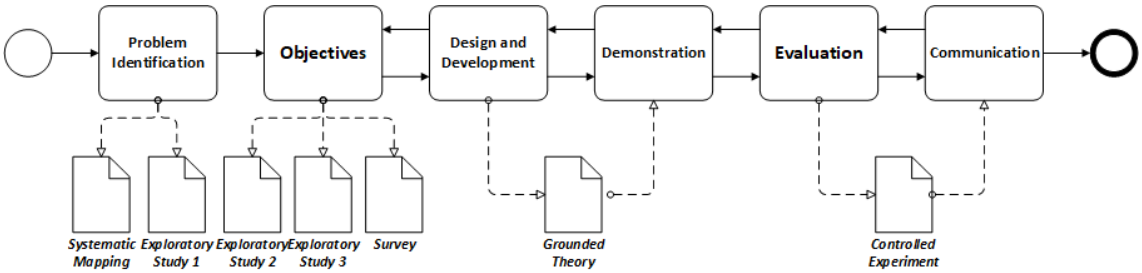

Figure 1. Research Methodology
Source: Authors

Activity 1: Problem identification and motivation

The purpose of this activity is to define our investigation scope. For this, we conducted a set of studies in the scientific literature and OER communities. We elaborated an exploratory study on OER in order to map how digital communities have adopted OER. After, we conducted a systematic mapping to identify the current state-of-the-art of OER. Based on the results obtained, we identified a gap in the intersection between OER and programming education. More specifically, the identification and retrieving of OER by students and teachers of programming. Our investigation generated two published studies:

- Deus, W. S. de; Barbosa, E. F. *"A Systematic Mapping of the Classification of Open Educational Resources for Computer Science Education in Digital Sources"*. In: **IEEE Transactions on Education**, 2021. (ToE'2021). Available at: https://doi.org/10.1109/TE.2021.3 128019
- Deus, W. S. de; Barbosa, E. F. *"The Use of Metadata in Open Educational Resources Repositories: An Exploratory Study"*. In: **44th Annual Computers, Software, and Applications Conference**, 2020, Spanish, Madrid. (COMPSAC'2020). Available at: https://doi.org/10.1109/COMPSAC48688.2020.00025

Activity 2: Objectives

The aim of this activity is to refine the research goals. For this purpose, we collected as much data as possible to create a large base of information on programming OER. We carried out two exploratory studies, generating a dataset with 4,027 OER for programming education. Also, we surveyed teachers of programming. The obtained results were summarized in following publications:

- Deus, W. S. de; Barbosa, E. F. *"Recursos Educacionais Abertos para o Ensino e Aprendizado de Programação no Brasil: Primeiros Referenciais"*. In: **XXXII Simpósio Brasileiro de Informática na Educação**, 2021, Brasil. (SBIE'2021). Available at: https://doi.org/10.57 53/sbie.2021.218 391

- Deus, W. S. de; Barbosa, E. F. *"An Exploratory Study on the Availability of Open Educational Resources to Support the Teaching and Learning of Programming"*. In: **50th Frontiers in Education Conference**, 2020, Uppsala, Sweden. (FIE'2020). Available at: https://doi.org/10.1109/FIE44824.2020.9274202
- Deus, W. S. de; Fioravanti, M. L.; Oliveira, C. D. de; Barbosa, E. F. *"Emergency Remote Computer Science Education in Brazil during the COVID-19 pandemic: Impacts and Strategies"*. In: **Brazilian Journal of Computers in Education**, 2020. (RBIE'2020). Available at: https://doi.org/10.5753/RBIE.2020.28.0.1032

Activity 3: Design and Development

The purpose of this activity is to systematize the collected data to identify challenges that hinder the identification and retrieval of programming OER. We have investigated the collected data looking for patterns; for example, we noticed that programming OER often have technical terms (such as *loops, variables*, or *recursion*). Therefor, our emerging theory stated that we should propose a vocabulary to identify programming OER more easily. In this activity, we have adopted the Grounded Theory [Pandit 1996] to elaborate an emerging theory on the challenges and barriers to be overcome. At the same time, we are prototyping potential solutions.

Activity 4: Demonstration

The objective of this activity is to perform proof of concepts to support our emerging theory. Currently, we have already developed two mechanisms following our emerging theory: (1) IPAVL - An Academic Vocabulary List for Introductory Programming; and (2) TAGGER – A new strategy to identify OER.

The IPAVL has 230 terms related to the teaching and learning of programming, such as abbreviations, technical words, programming languages names, among others. These terms can be adopted in OER collections to identify relevant resources for teachers and students of programming. The study describing the IPAVL will be submitted to an international journal.

TAGGER is an innovative strategy using weights and HyperText Markup Language tags instead of metadata to facilitate the identification of programming OER. We intend to submit the manuscript describing the TAGGER to an international journal.

Activity 5: Evaluation

The purpose of this activity is to evaluate our emerging theory. For this, we intend to perform an empirical investigation (e.g., focal group, study case or controlled experiment) with teachers and students of programming. Such participants will access the mechanisms produced, providing feedbacks related to usability and search efficiency.

Activity 6: Communication

In this activity, outputs, results, and findings are disseminated to the community. For this purpose, we have adopted scientific publications in events/periodicals. Also, we have been creating a web portal to support the use and reuse of OER by teachers and students of programming.

3. Results

Publications

Currently, five studies were published in our PhD work and two studies are under evaluation. We are working on two new studies to be submitted by 2023 and new collaborations with other researchers.

Contributions

We investigated the current state-of-the-art and practice around OER and programming education. Also, we created an OER dataset for programming education with 463 resources. In addition, we are identifying and solving problems faced by teachers and students of programming (i.e., outdated interfaces and lack of specific filters).

Mechanisms

In our PhD, mechanisms are web solutions for teachers and students of programming. Current, we are planning the development of a web portal unifying the dataset created and TAGGER strategy. Also, we are creating a web tool to make available the IPAVL.

Talks

During the PhD, the author participated in two events as speaker: (1) *Roda de Debates: Educação Aberta. Recursos Educacionais Abertos: Desafios e Oportunidades.* University of São Paulo. (Invited speaker); and (2) *Recursos Educacionais Abertos - Definições e uso.* Federal Rural University of Pernambuco. (Main speaker). Also, our papers "A Systematic Mapping of the Classification of Open Educational Resources for Computer Science Education in Digital Sources" and "The Use of Metadata in Open Educational Resources Repositories: An Exploratory Study" were presented at Open Education Week 2022 by Dr. Ellen Francine Barbosa. These papers were presented in the Webinar celebrated by the UNESCO[1].

Open Educational Resources

In order to disseminate our results among teachers, students and researchers, we are disseminating our main results into EduCapes repository. These resources can be used or reused by others users in many contexts. Currently, the OER produced are available in the following address: https://educapes.capes.gov.br/OER-phd.

Future contributions

In the future, we intend to distribute our mechanisms as Open Source Software. For this, the source code will be distributed freely, under open licenses. Also, to contribute for teachers and students, we are developing a web portal. This web portal will provide instructions and information on the use and reuse of OER, providing a large OER collection for programming education.

In addition, we intend to finish our data analysis by adopting Grounded Theory. In other words, we will establish an emerging theory on how OER of introductory programming can be identified and retrieved. Next, we intend to evaluate our emerging theory using empirical strategies with real users, such as students, teachers and relevant stakeholders in programming education.

[1]https://www.unesco.org/en/articles/unesco-celebrates-open-education-week-identifying-quality-oer-using-emerging-technologies

4. Conclusion

One of the most relevant outcomes in our PhD is to democratize the access for programming education, promoting free and open educational materials. In this sense, the main artifact generated by this PhD will be a set of tools for teachers and students of programming to identify and retrieve OER for programming education.

Finally, it is essential to highlight some limitations of our PhD work. the main threat is related to the research methodology adopted. It may introduce several risks in our analysis due to multiple data collection strategies. To avoid the propagation of such risks, experts and collaborators are being consulted to contribute to our investigation.

Acknowledgment

This work was supported by grant #2019/26871-4, São Paulo Research Foundation (FAPESP).

References

ACM/IEEE, C. T. F. (2020). Computing curricula 2020: Paradigms for global computing education. Technical Report 10.1145/3467967, Assoc. for Computing Machinery (ACM) and IEEE Comput. Comput. (IEEE-CS), New York. Access date: 08/04/2022.

Bahamón, J. C. (2022). A case study on the adoption of open educational resources in a c programming course. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2*, SIGCSE 2022, page 1128, New York, USA. ACM.

Dresch, A., Lacerda, D. P., and Antunes, J. A. V. (2015). *Design Science Research*, pages 67–102. Springer International Publishing, Cham.

McGill, M. M., Baskin, J., Berry, M., Burke, Q., Delyser, L. A., Grover, S., Lewis, C., Morrison, B. B., and Pruitt-Mentle, D. (2022). K-12 computing education and education research resources. In *Proceedings of the 53rd ACM Technical Symposium on Computer Science Education V. 2*, SIGCSE 2022, page 1033, New York, USA.

Medeiros, R. P., Ramalho, G. L., and Falcão, T. P. (2019). A systematic literature review on teaching and learning introductory programming in higher education. *IEEE Transactions on Education*, 62(2):77–90.

Pandit, N. R. (1996). The creation of theory: A recent application of the grounded theory method. *The qualitative report*, 2(4):1–15.

Pimentel, M., Filippo, D., and Santoro, F. M. (2019). Design science research: fazendo pesquisas científicas rigorosas atreladas ao desenvolvimento de artefatos computacionais projetados para a educação. *Metodologia de Pesquisa em Informática na Educação: Concepção da Pesquisa. Porto Alegre: SBC*.

Tovar, E., Chan, H., and Reisman, S. (2017). Promoting merlot communities based on OERs in computer science and information systems. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC) V. 2*, pages 700–706.

UNESCO (2019). Recommendation on open educational resources (oer). Technical report, Unesco, Online. Access date: 24/01/2022.

Wiley, D., Bliss, T. J., and McEwen, M. (2014). *Open Educational Resources: A Review of the Literature*, pages 781–789. Springer New York, New York, NY.