# A Systematic Mapping of the Classification of Open Educational Resources for Computer Science Education in Digital Sources

William Simão de Deus and Ellen Francine Barbosa

*Abstract*—***Contribution:*** **This study presents the results of a systematic mapping on the classification and organization of Open Educational Resources (OERs) focused on Computer Science Education (CSEd) in digital sources.**

***Background:*** **The number of open resources (e.g., images, videos, and websites) available on the Internet for the teaching of Computer Science has increased in recent years. As a consequence, several authors have proposed different ways of classifying them, which include lists of disciplines, classification algorithms, and semantic enrichment. However, such studies are dispersed and fragmented in the literature, making the current state-of-the-art unclear.**

***Research Questions:*** **1) What digital sources have been used to classify OERs for CSEd? 2) What CSEd domains have been explored for the classification of OERs? 3) What approaches have been adopted for the classification of OERs for CSEd?**

***Methodology:*** **Three search strategies were performed: automatic search in the five major digital libraries of CSEd, manual search in the major CSEd conferences and journals, and snowballing on the references and citations. As a result, 22 relevant studies were selected.**

***Findings:*** **The main findings were systematization of the main OER sources for CSEd, identification of 64 CSEd domains covered by OERs organized in a taxonomy, the immaturity level of investigations and tools, and lack of mechanisms that organize OERs for CSEd.**

*Index Terms*—**Open Educational Resources; Computer Science Education; OER Digital Source; OER Classification.**

## I. Introduction

AMONG the current digital materials for the teaching and learning of Computer Science Education (CSEd), such as websites, applications, and tools [1], many are Open Educational Resources (OERs), due to their permissive licenses and educational purposes.

However, two challenges have been faced. Firstly, OERs for CSEd can assume any digital format (e.g., videos, images, texts, among others), leading to no consensus on the best way to classify them digitally. Some researchers have used fixed CSEd domains as a technique (see Tovar et al. [2] or Rathod and Cassel [3]), which requires the application of lists with terms (e.g., Algorithms, Programming, and Database), and others have employed dynamic classifications provided by Machine Learning, such as Mouriño-García et al. [4], Hameed and El-Ameer [5] or Kastrati et al. [6]) Classifications repeat patterns such as name of technologies (Java, Python or C) and educational level (K-12, Computer Science 1 or Computer Science 2). Some complex techniques that involve semantic enrichment [7] and hybrid classification strategies [8] have also been applied. Nonetheless, all such studies are scattered and fragmented in the literature, leading to an unclear state-of-the-art due to the lack of a systematic mapping of the approaches.

Secondly, despite the large number of studies and approaches analyzing OER sources, similar mistakes are repeated [9] (e.g., unclear organizational structures, bad use of metadata, lack of documentation, among others). The lack of a mapping of the OERs classification has increased the complexity of the scenario, since such mistakes occur in novel approaches Leake and Lewis [10]. Therefore, the detection of an OER for CSEd is challenging, and students and teachers must perform several searches in different OER sources for finding a relevant result.

Considering the aforementioned context, this study reports on a Systematic Mapping (SM) based on Kitchenham and Charters [11] guidelines conducted towards identifying OER sources with CSEd and CSEd domains, as well as approaches that classify OERs for CSEd.

The remainder of this study is organized as follows: Section II presents the background and related work; Section III summarizes the SM design; Section IV provides the solution to each Research Question (RQ); Section V discusses the findings and the threats to validity; finally, Section VI is devoted to the conclusions and suggestions to future studies.

## II. Theoretical background

### A. Open Educational Resources for Computer Science

Term OER was defined by Unesco [12] as *"the open provision of educational resources, enabled by information and communication technologies, for consultation, use and adaptation by a community of users for non-commercial purposes"*. "Open" is related to rights and costs. On the one hand, permissive licenses, such as public domain and open licenses [13] are adopted. On the other hand, OER is free for users to access, use or adapt, and minimal restrictions (e.g., use of the open format) are applied. "Educational" refers to purpose: OER must aim at teaching, learning or research. "Resources" encompass media types. In summary, any digital format can be

Manuscript received March 16, 2021; revised October 19, 2021; accepted November 9, 2021. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Suprior - Brasil (CAPES) - Finance Code 001 and the Fundação de Amparo a Pesquisa do Estado de São Paulo (FAPESP 2018/26636-2). *(Corresponding author: William Simão de Deus.)*

The authors are with the Institute of Mathematics and Computer Science, University of São Paulo, São Carlos 13560970, Brazil (e-mail: williamsimao@usp.br; francine@icmc.usp.br).

considered an OER, as long as it has an educational purpose and is open.

In a related perspective, CSEd is the term that represents the teaching and learning of Computer Science. In recent years, several studies have focused on users' interest in OERs for CSEd (Combéfis et al. [1], Staubitz et al. [14], Tovar et al. [2], Rathod and Cassel [3]) aroused by the breadth of available resources. OERs can assist in the teaching and learning of several CSEd domains, such as Algorithms, Database, Programming, etc., in different digital types and educational levels. However, a serious challenge concerns their classification, since, due to their heterogeneity, several approaches can be adopted. Moreover, both gratuity and quality make OERs essential materials for students and teachers of CSEd.

### B. Related Work

The literature reports studies with similar objectives, but with severe limitations. Silva and Silveira [15] conducted a review of open educational games for programming education; however, they did not follow a well-established guideline, using a base with preprint studies (Research Gate) and performing general searches. This study differs from theirs, since the systematic research process was based on Kitchenham and Charters [11], and scientific libraries with peer-reviewed studies were adopted. Moreover, the search encompassed years 2002 to 2020, whereas Silva and Silveira [15] analyzed the 2014-2019 period.

Wang et al. [16] conducted a bibliometric analysis of 15 year's research on OERs, and the main limitations were use of unique digital library and language. The difference between their study and this SM are the total scientific bases used for the automatic search (5) and a more comprehensive analysis of CSEd in contrast to their inclusion of only sources of CSEd resources.

Ren [17] provided a short review of open and interactive publications towards technological innovations. Although relevant to the OER community, from the CSEd perspective, the contributions were focused on a generic analysis from the STEM (Science, Technology, Engineering e Mathematics) context, whereas this mapping presents the current state-of-the-art on CSEd conducted by a well-established methodology.

## III. Systematic Mapping Design

### A. Goal and Research Questions

The guideline proposed by Kitchenham and Charters [11] was adopted for this SM. Initially, the goal and the RQs were defined. The main goal was the way OERs for CSEd were classified, identifying the OER sources used, the domains of CSEd covered by OERs, and the approaches used for classification. The following RQs were then raised:

- $RQ_1$ What digital sources have been used to classify OERs for CSEd?
  **Rationale:** OER digital sources are dispersed and fragmented, thus leading to a gap on how and what sources contain them for CSEd. To the best of our knowledge, the literature reports no list of sources with CSEd content, and this RQ aims at identifying the OER sources for CSEd.
- $RQ_2$ What CSEd domains have been explored for the classification of OERs?
  **Rationale:** CSEd domains are also dispersed and fragmented, and several can be used (e.g., academic disciplines (i.e., algorithms), educational level (i.e., Computer Science 1 or Computer Science 2), subjects (i.e., programming, databases), among others). This RQ aims to identify the domains used in OER classification.
- $RQ_3$ What approaches have been adopted for the classification of OER for CSEd?
  **Rationale:** Although several approaches can organize OER sources, the literature lacks a mapping of the body of knowledge generated. This RQ aims to summarize the approaches that classify OER for CSEd.

### B. Search and Selection Process

The following inclusion (IC) and exclusion (EC) criteria were defined for solving the RQs:

- **Inclusion:**
  - $IC_1$: The study should investigate the classification of OER for CSEd;
  - $IC_2$: The study is a primary study;
- **Exclusion:**
  - $EC_1$: The study does not investigate the classification of OER for CSEd;
  - $EC_2$: The study is not a primary study;
  - $EC_3$: The study is not evaluated by peers;
  - $EC_4$: The study is not written in English;
  - $EC_5$: The study is an old version of another study already added;

Towards an accurate execution, the following three iterations were defined: first, read the main identification fields of studies (title, abstract, and keywords) to check the criteria briefly established. Second, read the potential studies selected in the previous step in a guided form (at this step, both Introduction and Conclusion sections have been read for the application of the criterion). Finally, extract data after a full reading of the selected studies.

*1) Automatic search:* Initially, two studies Mouriño-García et al. [4] and Chicaiza et al. [18] were selected, since they provided answers to the RQs and were considered a control group. The terms and synonyms were then extracted from them according to the domains of interest. The terms were unified with logical operators (OR/AND), and the following string base was obtained: *(categorization OR category OR classification OR schema OR scheme OR taxonomy) AND ("educational content" OR "educational resource" OR OER) AND (collection OR dataset OR repository)*. Pilot tests were conducted in Scopus library towards searching for studies based on title, abstract, and keywords, assuring the control group studies had been returned and the most meaningful terms had been adopted.

The automatic search involved the most common digital libraries in the CSEd area, namely Scopus, IEEE Xplore, ACM DL, Science Direct, and Web Of Science. It was based on title, abstract, and keywords through a variation in the string base, which was customized and applied to each digital library. Duplicated studies were removed, and 282 potential studies were retrieved. Table I shows a summary of the results.

| Source | Total |
|---|---|
| Scopus | 125 |
| IEEE Xplore | 85 |
| ACM DL | 16 |
| Science Direct | 59 |
| Web of Science | 62 |
| **SUB-TOTAL** | **347** |
| Duplications | 65 |
| **TOTAL** | **282** |

TABLE I
AUTOMATIC SEARCH RESULTS

*2) Manual search:* Manual searches were conducted for ensuring potential studies not indexed in automatic search would be identified. The searches detected studies published in relevant CSEd sources (Table II).

| Type | Source | Abbr. |
|---|---|---|
| Conference | Technical Symposium on Computer Science Education | SIGCSE |
| Conference | ACM Conference on Computer-Supported Cooperative Work & Social Computing | CSCW |
| Conference | IEEE/ASE Frontiers in Education Conference | FIE |
| Conference | International Conference on Learning Analytics & Knowledge | LAK |
| Journal | Computers & Education | C&E |
| Journal | IEEE Transactions on Education | ToE |
| Journal | IEEE Transactions on Learning Technologies | TLT |
| Journal | ACM Transactions on Computing Education | TOCE |
| Journal | Journal of Education and Information Technologies | EAIT |

TABLE II
MANUAL SOURCES

Conferences were selected according to the h5 index and the h5 median metrics provided by Google Scholar, and Journal Citations Report (JCR) was used for the journals selection. After the selection, manual searches were performed towards identifying potential studies in proceedings and issues.

*3) Snowballing search:* Snowballing was applied to identify studies as a complementary search. The citations received (forward) and the reference list (backward) of each selected study were checked, and the criteria were then applied for the identification of potentials studies.

Snowballing is iterative ($si$) and ends when no study has been identified. The first snowballing iteration ($si_1$) used the initial set of selected studies produced by both automatic and manual searches, and the second ($si_2$) employed the studies selected in ($si_1$). After ($si_2$), no relevant study was identified and snowballing was ended. Figure 1 shows the final result of the automatic, manual, and snowballing searches.

### C. The Data Extraction Framework

A framework organized into four facets was generated for the data extraction. The first facet (General) showed fields for the identification of the study, whereas the second (RQ1 - Sources) focused on data extraction for RQ solution. The third facet (RQ2 - Domain) solved RQ, and the fourth (RQ3 - Approaches) focused on data extraction for RQ. Figure 2 illustrates the framework design. The dataset generated is provided in Appendix A.

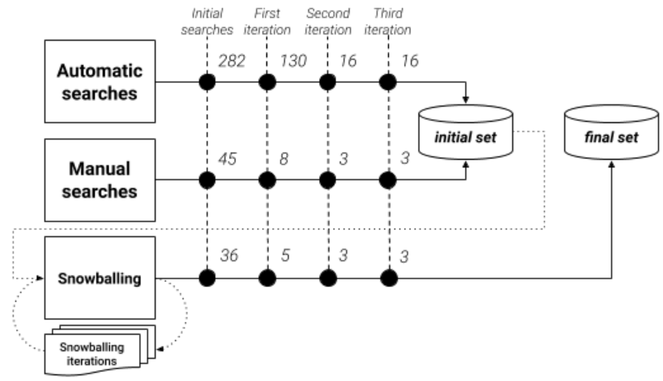


Fig. 1. Final result of searches

### D. Data Synthesis

The data were synthesized in the following steps: first, sentences were extracted from each study according to the data field in the framework. A cross-checking was then performed, and the data extracted from each study were compared to the others. For instance, some studies adopted term *"automatic annotation process"*, while others adopted *"automatic tagging"* to describe the classification strategy used. However, the analysis of the approaches revealed both studies used the same strategy with distinct terms. Whenever possible, well-recognized guidelines and relevant studies (e.g., Computer Science Draft Competencies [19], Wiley et al. [20] and Yuan et al. [21]) were adopted for the classification of data extracted and organized from the studies.

## IV. RESULTS

22 studies were selected, of which 54% (12) had been published over the past five years and 36% (8) had been published recently (2 years or less). The remainder were distributed in the 2006-2014 period. Year 2019 showed the largest number of publications, i.e., 27% (6). A list with the selected studies is provided in Appendix B.

### A. The Sources ($RQ_1$)

Initially, 24 sources were identified and organized as follows: repositories (19), which encompass any type of OER; platforms (8), which support only 1 type of OER (such as video or book); and dataset (1), used for algorithm training. 14 studies were considered invalid, since they were unavailable for access. Figure 3 shows the valid sources identified.

The valid sources were accessed for the identification of the way OERs were organized and classified. The procedure analyzed whether the source provided manual search (a user browsing a collection by clicking on categories) and automatic search (a user searching for a term), and required login for accessing the content. It also checked the OERs features, the type of license (OER availability under open licenses and individually or in the entire collection), and type of metadata (metadata available either in the content, or in an embedded way). Table III shows a summary of the analysis.

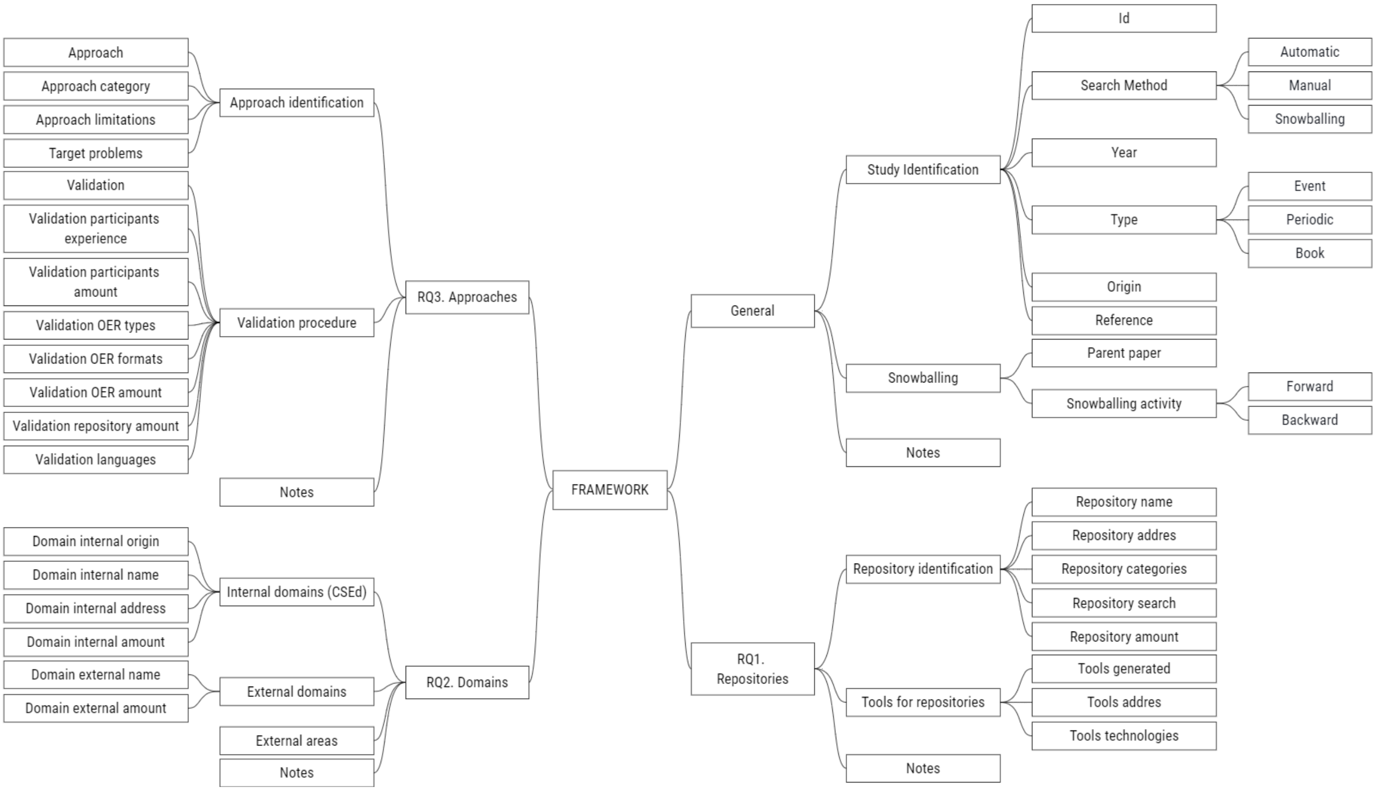


Fig. 2. Framework design for perform data extraction

| Source name | Link | Manual | Automatic | Login | Licence | Metadata |
|---|---|---|---|---|---|---|
| Cornell Virtual Workshop | https://cvw.cac.cornell.edu/default | ✓ | | | Individually | |
| Coursera | https://www.coursera.org/ | ✓ | ✓ | ✓ | | Embedded |
| CrowdComp | https://dl.acm.org/doi-num | | | | | |
| DBPedia | https://wiki.dbpedia.org/ | ✓ | ✓ | | | |
| edX | https://www.edx.org/ | ✓ | ✓ | ✓ | | Embedded |
| GLOBE | http://www.globe-info.org/ | | | | | |
| Google Books | https://books.google.com/ | | ✓ | | | Embedded |
| LOCWD | https://www.emerald.com | | | | | |
| Merlot | https://www.merlot.org/merlot/ | ✓ | ✓ | | Individually | Embedded |
| OERCommons | https://www.oercommons.org/ | ✓ | ✓ | | Individually | Embedded |
| Open Stax CNX/Connexions | https://cnx.org/ | ✓ | ✓ | | Collection | Embedded |
| Slideshare Channel | https://www.slideshare.net/eccutpl | ✓ | ✓ | | | Embedded |
| Udacity | https://www.udacity.com/ | ✓ | ✓ | ✓ | | |
| Videolectures.net | http://videolectures.net/ | ✓ | ✓ | | Collection | Embedded |
| Youtube | https://www.youtube.com/ | ✓ | ✓ | | Individually | |
| Youtube Channel | https://www.youtube.com/user/eccutpl | ✓ | ✓ | | Individually | |
| Mit OpenCourseWare | https://ocw.mit.edu/index.htm | ✓ | ✓ | | Collection | Embedded |
| NSDL | https://nsdl.oercommons.org/ | ✓ | ✓ | | Individually | Embedded |
| OpenLearn | https://www.open.edu/openlearn/ | ✓ | ✓ | | Collection | Embedded |
| UCI | http://open.uci.edu/ | ✓ | ✓ | | Individually | Embedded |
| Wikibooks | https://www.wikibooks.org/ | ✓ | | | Collection | |
| RUCM | https://biblioteca.ucm.es/en | ✓ | ✓ | Variable | Individually | Variable |
| Serendipity | http://serendipity.utpl.edu.ec/ | ✓ | ✓ | | Individually | Embedded |

TABLE III

SOURCES OF OERS USED IN THE SELECTED STUDIES. OBS: CONNEXIONS HAS BEEN UPDATED FOR OPEN STAX CNX.

However, a problematic scenario was identified: 81% of sources do not support manual search for OERs and 77% provide no automatic search. Concomitantly, 18% impose some login restriction for access to the content of the resources, 40% provide no information on the types of OER licenses, and 22% do not identify any type of metadata.

The auxiliary identification mechanisms were then analyzed. References and citations to CSEd topics were collected, and the documentation that supported the identification and classification of resources, such as features of the collection, tips for searches and languages were investigated. 54% of sources cited CSEd domains explicitly on their pages. From this perspective, 4 sources (18%) provided some type of documentation to support the user and 2 (9%) were multi-language. Figure 4 shows a summary of the results.

*1) Tools identified:* Tools developed for facilitating the discovery/classification of resources were identified. The first was CROERA system, presented in study S01, which automates the classification of OER, including resources of CSEd area. Data were extracted from three repositories and the tool created a

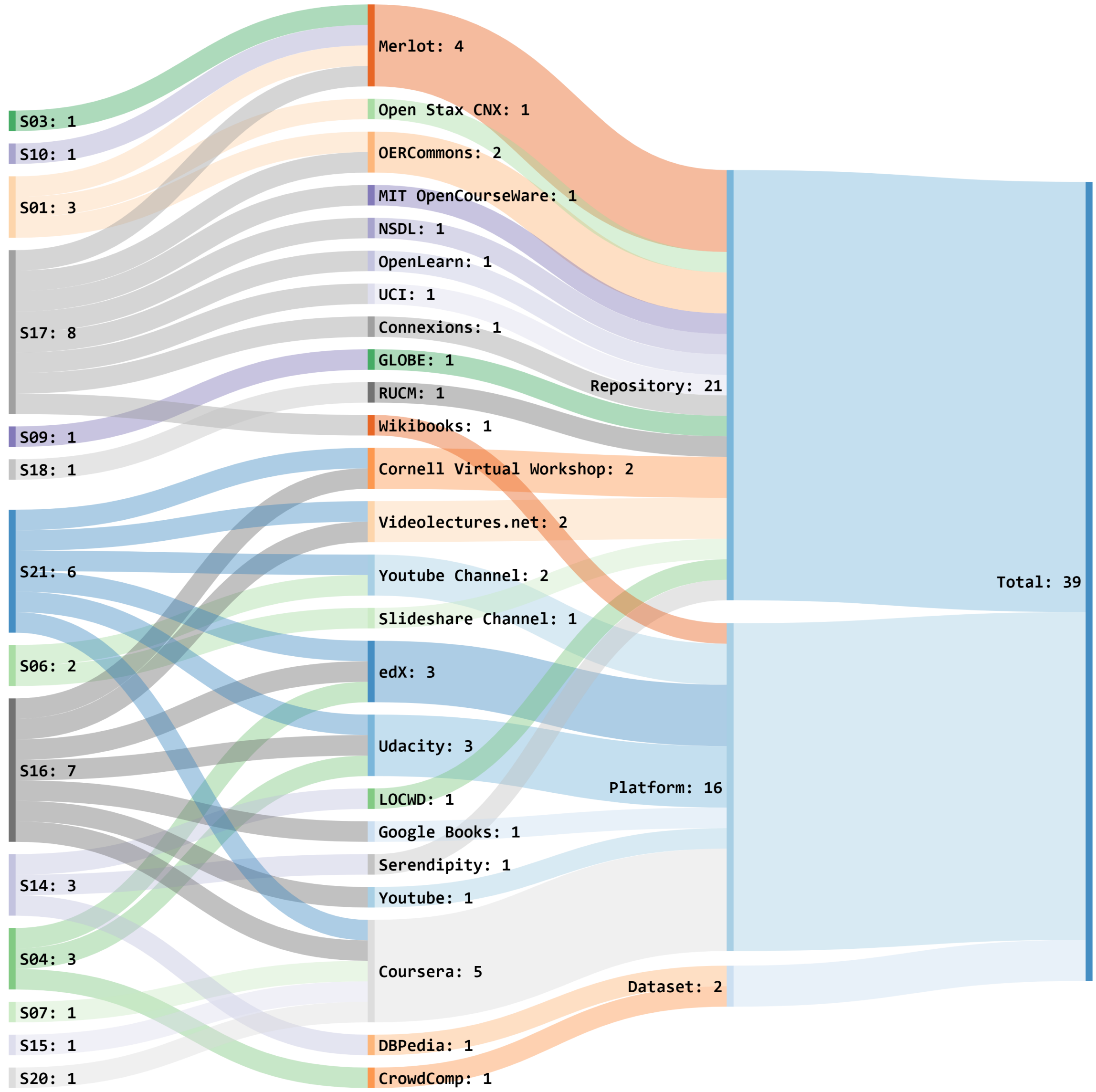


Fig. 3. Sources of OERs used. The number shows the total citations received by each source according to the selected studies

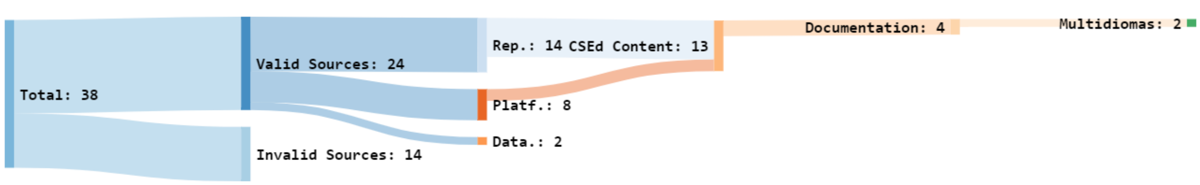


Fig. 4. Lack of support to users

new classification model, regardless of the original scheme.

Another tool was introduced in S10. The authors provided a graphic classification model that enables users to search for a term and navigate through a tree of previously classified options. 10 CSEd topics are currently available for users.

Two studies (S18 and S19) discussed the use of ontologies for future adaptations and repository creation. Each ontology was designed for organizing and systematizing the content of OERs repositories. While the ontology presented in study S18 focuses on CSEd, the one in study S19 is general and can be applied to organize OERs for CSEd.

Finally, study S05 addressed the planning and creation of a repository for the teaching and learning of free and open-source software and contained materials to support educators, students, and interested users. Similarly, the authors of study S21 presented a repository called ERuDIte for Data Science resources. However, the sites of the cases cannot be accessed.

### *B. CSEd Domains ($RQ_2$)*

The CSEd domains presented in the selected studies were mapped for solving $RQ_2$, and Computer Science Draft Competencies [19] was adopted for their systematization. The document provides the competencies that Computer Science students should have and the possibilities for their development. The domains were compared with each competency, and 64 domains were classified at the end of the process (Figure 5). Next is a list of CSEd competences supported by the domains identified in OERs:

- **AL-Algorithms and Complexity:** Domains related to both complexity and analysis of algorithms. The ones identified assist in the implementation of data structures for the solution to problems.
- **AR-Architecture and Organization:** Domains related to the implementation and organization of high-level computers.
- **CN-Computational Science:** Domains for a simple and mathematical creation of real-world situations.
- **DS-Discrete Structures:** Only the Discreet Structures competence was identified; its specific domains were not considered.
- **GV-Graphics and Visualization:** Domains of the use of graphic and visual techniques in Computer Science.
- **HCI-Human-Computer Interaction:** Similarly to the Discreet Structures, only the Human-Computer Interaction competence was identified; its domains were not discussed.
- **IAS-Information Assurance and Security:** Domains related to cybersecurity and computer security.
- **IM-Information Management:** One of the main competences investigated in the studies analyzed. The domains discussed by the resources are organized into two distinct categories: one for the use of query languages towards the recovery of information from a database, and the other for concepts of modelling and notations for models of relational data.
- **IS-Intelligent Systems:** Domains related to the use of interpretation and translation of natural languages. Domains on inferences were also identified from real-world problems.
- **NC-Networking and Communication:** All domains of this competence, namely design and development of client-server applications, design and implementation of network protocols, and contrast between fixed and dynamic techniques were explored.
- **OS-Operating System:** The domains presented were application of theoretical knowledge for problem-solving and Implementation of programming solutions for operating systems.
- **PBD-Platform-based Development:** Domains related to the development of mobile apps. The competence itself was also investigated.
- **PD-Parallel and Distributed Computing:** High-performance computation and the competence itself were examined.
- **PL-Programming Languages:** One of the main competencies explored by OERs. Domains that support the programming-language competence and assist both introductory aspects and conceptual domains were identified, and those related to design and implementation of classes, encapsulation concepts, and programming languages were mapped.
- **SDF-Software Development Fundamentals:** Domains related to development and algorithms that illustrate iterative processes. The decomposition of programs into components and the competence itself were introduced.
- **SE-Software Engineering:** Domains related to system software design. Only the competence was investigated.
- **SF-Systems Fundamentals:** Only the competence was explored by OERs, and no specific domains were identified.
- **SP-Social Issues and Professional Practice:** Similarly to the other competences, no specific domain was explored, and only the competence itself was exposed.

The main competencies supported by OERs are programming language and information management. Technologies such as programming languages and concepts are available in both cases, which favors the teaching and learning process. However, some competencies are explored only at a high level (“Parallel and Distributed Computing”, “Discrete Structures”, “Human-Computer Interaction”, “Software Engineering”, “Systems Fundamentals", and “Social Issues and Professional Practice”). Next, the origins of these domains are identified.

*1) CSEd domains origins:* The selected studies adopted different origins to choose the CSEd domains to be used. For example, S03 used the curriculum recommendations provided by the Computer Sciences Accreditation Board, and S17 and S19 used only ACM guidelines. S10 selected domains without using guides or references and based only on the authors' experience, whereas S05 adopted the domains of Open Source Education proposed by Sasikumar[1] as the domain origin. S18 used the research of Hong-yan et al. [22] as a source, and S01, S02, S04, S06, S07, S08, S09, S12, S14, S15, S16, and S20 employed the original domains provided by the sources showed in Section IV-A. The origin of the CSEd domains could not be identified in S11, S13, S18, S21, and S22. Figure 6 shows the origins detected.

*2) CSEd domains problems:* Unfortunately, the criteria for the selection of CSEd domains were immature in six selected studies. Several approaches only provide definitions, but do not support their decisions. For instance, S13 presented the *10 major disciplines of CSEd* as a selection criterion, but no technical reference that supported the assumption was identified. Similarly, S10 provided 10 keywords to validate the proposed approach. However, keywords largely derived from computing domains are confusing and involve technologies (such as CSS and JSP), operating systems (Linux), programming languages (Python, SQL), among other computing domains. S11, S13, S18, S21, and S22 enabled no identification of the domain origin.

[1] http://thelittlesasi.wikidot.com/foss-edu

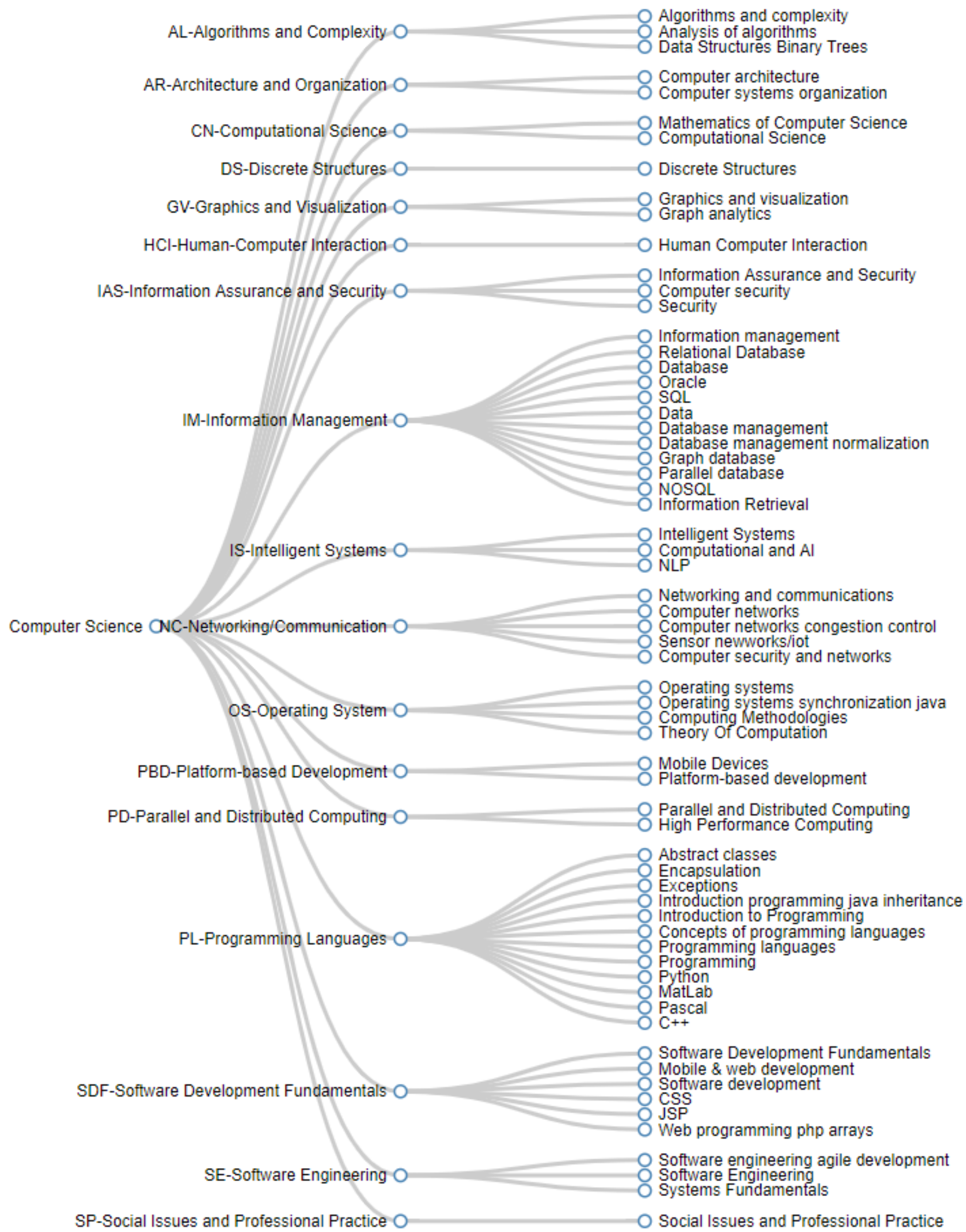


Fig. 5. A taxonomy of CSEd domains based on the selected studies

Granularity also appears to be a current problem for OER classification, since some approaches focus on general classifications, such as CSEd disciplines, whereas others are devoted to specific ones (e.g., educational level). The synchronization of the data is complex; consequently, a gap has been created between general and specific classifications, hampering the connection of different OER sources. As an example, a repository can adopt a general classification system based on programming disciplines (algorithms, databases, networks, etc…), whereas another adopts educational levels (beginning, intermediate, and advanced). However, several disciplines are distributed at different levels (e.g., "Introduction to Algorithms", "Algorithms", "Database I", and "Database II"). In such cases, how should the OERs of those two sources be classified, since their connection causes the classification of a repository to overlap the classification of another?

### *C. Approaches Adopted (RQ$_3$)*

Sentences were extracted from the selected studies towards providing a summary of each approach and, therefore, solving RQ$_3$. This step aimed at identifying the main features of each study and the main limitations of each approach. The approaches were grouped according to their features, which led to the following four types:

**Overlap classification:** refers to an original classification of the OER replaced by a new one. It was the most common

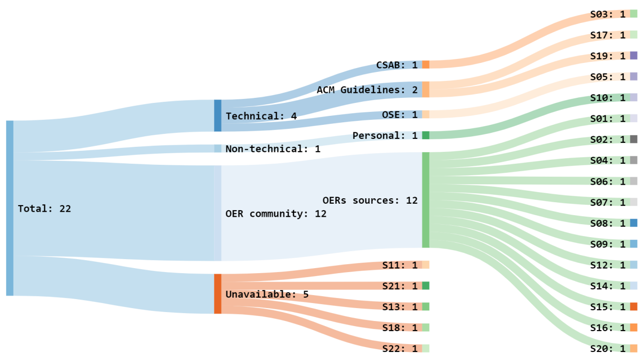


Fig. 6. Studies and CSEd domain origins

approach in the selected studies and almost always used classifying algorithms to generate (new) categorizations.

- Main features: content and/or metadata of the resource are used for the generation of a new classification; ii) the new classification overlaps the original one; iii) the new classification can be random or previously defined in fixed categories.
- When to use? When a new source must be created, two or more different sources must be interconnected, or resources available on the web must be classified.
- Problems: The approaches can be biased if the dataset used for training has a predominant domain.
- Studies: The strategy was adopted by the following 12 studies (54%): S01, S05, S10, S13, S14, S11, S06, S08, S12, S19, S20, and S22.

**Fixed classification:** refers to approaches that classify a resource during registration in the source by means of different strategies.

- Main features: i) only one source is used; ii) no comparison is made with other classification approaches.
- When to use? When the aim is to analyze the collection of an OER source or propose an improvement for the categorization of an existing source.
- Problems: difficulty for a correct classification of resources, mainly when OERs for CSEd can be classified into two or more categories.
- Studies: The following seven studies (31%) presented the approach: S03, S09, S02, S06, S18, S17, and S21.

**External annotation:** it uses an external knowledge base (such as Wikipedia[2] or DBPedia[3]) to extract semi-structured data. A certain resource is then enriched and classified according to the data extracted from the external base.

- Main features: i) an external base is used for annotation; ii) a source (such as OER repositories, dataset, or web resources) is used for annotation.
- When to use? When the aim is to verify whether the resources were classified correctly or when the classification is expected to be improved.
- Problems: its main limitation is the use of an external base – a small base leads to a poor classification.
- Studies: Only S04 (4%) introduced the strategy.

**Hybrid:** two or more approaches are used to classify OER.

- Main features: i) use of mixed approaches, thus combining different characteristics.
- When to use? When the aim is a robust classification merging two or more strategies, avoiding weaknesses when used individually.
- Problems: Hybrid approaches are costly and commonly involve people and machines.
- Studies: Only S16 (4%) introduced the approach.

Approaches often use metadata to generate new classifications, which is challenging for two reasons: metadata do not show the full content of an OER, thus ignoring relevant data about the material, and each OER source has a structure and format for listing metadata. Such a diversity hampers an easy insertion of new OER sources in the approach, and several studies have cited high processing cost as a challenge. Consequently, most approaches presented in the SM are complex, and no useful tool is provided to users. Experimental tests have been conducted with specialists (e.g., teachers and researchers in the field), who have almost always tagged and classified resources with no collaboration of students, which can be considered a threat. The target audience of OER sources is highly heterogeneous, and the work of only specialists does not reflect the real scenario. Such implications are commented next.

*1) The maturity of approaches:* The research type facet proposed by Petersen et al. [23] was applied for the establishment of the maturity degree of each approach in a real scenario. In what follows is a summary of the results.

**Evaluation Research:** The approach was proposed, implemented, and validated. During validation, either it was compared with other classification techniques, or human users participated in experiments. Seven (31%) studies (S01, S04, S07, S13, S15, S18, and S20) reached those premises in their validations and were classified as high-level maturity.

**Solution Proposal:** A new or reviewed solution is proposed and its applicability is discussed through either an example, or arguments. However, no empirical validation of its effectiveness is performed. Eleven (54%) studies (S02, S05, S06, S08, S10, S11, S12, S14, S16, S19, S21, and S22) have such a characteristic and were classified as intermediate level, since they lack empirical validation.

**Opinion Paper:** Provides opinions on the benefits or challenges of a certain approach. One study (S09) (4%) has such a characteristic and describes the experience adopting a classification approach supported by metadata for improving collaborative learning.

**Experience Paper:** Explains the implementation of a certain approach and is considered of low maturity, since its both impact and relevance on the users' routine cannot be measured. Two studies (S03 and S19) on this level were identified.

Figure 7 shows an overview of participants, sources, and resources involved in the validation. In general, the number of participants is reduced – 5, on average. Validations with larger groups are not common, and the sources used are also few, i.e., between 1 and 3. Approaches that adopt more than 10 sources

[2]https://en.wikipedia.org/wiki/Main_Page
[3]https://wiki.dbpedia.org/

were considered outliers due to their occurrence in few studies. On the other hand, OERs are largely used in validations. However, besides validations with several resources, a highest number of sources and participants is required, thus approximating the reality of validations to the real OERs scenario.

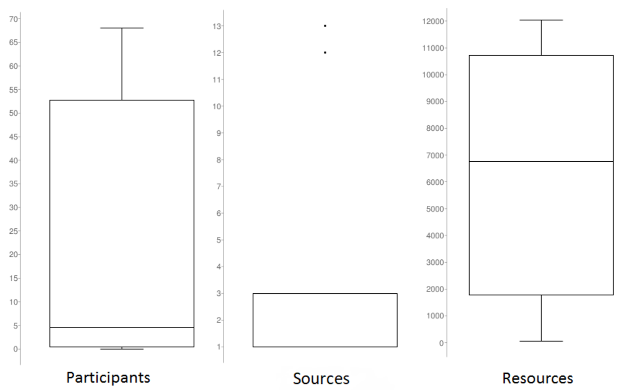


Fig. 7. Maturity level of the selected studies based on total number of participants, sources, and resources in each validation.

*2) Problems solved:* Each approach was designed to solve specific problems of OER classification (45 in total), whose organization was based on the studies of Yuan et al. [21] and Wiley et al. [20]. 10 studies quoted the discovery problem 17 times, which facilitated the finding of OER, and another 10 referenced the localization problem, i.e., how to make an OER more useful in other contexts. Six studies cited the interoperability problem (resources must be "researchable" among digital sources), whereas five mentioned the sustainability problem (staff and financial problems). The quality problem, i.e., the perception OERs are low-quality because they are free, was referenced in one study.

Five studies citing problems not addressed by Yuan et al. [21] and Wiley et al. [20] were identified. They referred to the lack of metadata and ontologies, as well as specific mechanisms to solve classification problems. The problem was defined as "Lack of mechanisms", which refers to structural problems in OER collections that require specific solutions. One study cited another problem not related in the other categories presented.

## V. Discussions

### A. Highlights and Findings

The highlights of RQ$_1$ are the current status of the main sources of OERs for CSEd and the lack of support for the search of OERs for CSEd. Despite the large number of online resources for the learning of CSEd topics, the lack of a standard has created a modern Babel tower, as addressed by Queirós and Leal [24]. The first highlight in the mapping contributed to support this problem. The community of researchers can benefit from the systematization of the main sources for identifying the main OER sources used in the investigation of CSEd. Moreover, the lack of support for searches of OERs for CSED is also highlighted. The search for OER has shown more complex, since several OER sources face problems such as lack of documentation and license issues. New classification approaches create new OER sources (e.g., repositories or tools); therefore, the propagation of those errors must be avoided. Finally, only two digital sources analyzed provided documentation and multi-language support to users. The first finding of this SM refers to the lack of practical advancements for supporting users' routine during searches for OERs; consequently, the solution to such a challenge must be focused on, mainly regarding resources for CSEd.

RQ$_2$ highlights the identification of CSEd domains covered by OERs in the selected studies. In this context, the second finding is the identification of 64 CSEd domains explored by OERs organized in a taxonomy that presents the competences most focused on the teaching and learning of CSEd and those that require investigations and analyses. RQ also emphasizes the use of non-technical domains and immature decisions. In short, several studies showed no methodological rigor while selecting CSEd domains, and some evidences provided in Section II also suggest similar problems in current OER research. Moreover, the analysis of RQ$_1$ identified 14 sources (26%) unavailable for access. The third finding refers to the need for a higher rigor of both technical quality of the products generated and methodological quality of the studies conducted. Although the limitation of the pages of the studies hampers the explanation of some decision-making, the authors can provide additional resources such as data and protocols in open repositories, enabling the reuse of several solutions and artifacts by other users or researchers.

RQ$_3$ shows two highlights, namely maturity level of investigations and lack of a mechanism that organizes OERs for CSEd. Regarding the former, only four selected studies (18%) involved participants for validation (commonly specialists), showing a small number of participation (2-7) and few sources (1-3). However, OER sources and users, as well as their profiles are distinct. The fourth finding refers to the immaturity of validations in the OER scenario. The validation of approaches related to OERs must encompass heterogeneous publics whenever possible towards simulating the reality. Lastly, the SM revealed a new problem for OER sources. Four studies (18%) cited the need to build a specific solution to the problem of OER identification. Since no previous study had identified the problem, its discovery for OER repositories was called *lack of mechanism*, representing the fifth finding.

Regarding the implications of this investigation, although several approaches can classify OERs for CSEd, the impact of the resources on students and teachers has not been visible yet. The results revealed a gap in educational aspects. The investigated approaches are often products of algorithmic training or manual classification; however, educational aspects of teaching and learning CSEd are disregarded. Educational issues, such as the CSEd abilities to be learned by students using this OER, or the competences to be worked on by a CSEd teacher using this OER in a classroom are ignored by the classifications. Despite OER being an educational product, it has not been explored by most investigations, and, although this mapping has identified 64 CSEd domains, several are wide and generic. The educational practice of OERs must still be focused on so that the problem of discovery of relevant materials can be solved.

### B. Threats to Validity

**External validity:** the main external threat is the fact the selected studies are not relevant/representative for the analyzed area. Towards mitigating it, the following actions were taken: i) a specialist in SM was consulted for the development of the mapping; ii) five digital libraries identified relevant studies; and iii) complementary search techniques were applied.

**Internal validity:** towards avoiding the spread of internal threats in the SM, guidelines and studies rather than personal classifications were adopted, and data were collected according to the nature, characteristics, and information provided by the original studies. Two SM experts revised the SM research protocol and corrected and evaluated this investigation.

**Construct validity:** automatic searches into title, abstract, and keywords were used towards mitigating this bias, and five digital libraries were consulted. Several tests were applied to Scopus library for the construction of the search string.

**Conclusion validity:** the contents of the SM were constantly refined and updated in function of the results, and a specialist reviewed the results and conclusions after the selection of studies.

## VI. Conclusion and Future Work

This study provided the results of an SM on the classification of OERs for CSEd in digital sources. The purpose was to find answers to some RQs raised. $RQ_1$ focused on identifying the classifications used in the sources, which led to the mapping of the main OER sources, tools, and categorizations. Problems such as lack of documentation were also detected. $RQ_2$ regarded the main CSEd domains covered by OERs. 64 CSEd domains were identified, generating a taxonomy, and problems such as use of non-technical domains and immature decisions were detected. Lastly, $RQ_3$ investigated the classification approaches, which were organized into four groups according to their features, showing their maturity and problems solved. An overview of the current state-of-the-art has been provided from the results, and synchronous contributions have been made towards filling the literature gaps presented. As a result, contributions to the culture of OERs can become more solid from the CSEd perspective, providing possible lines of investigation to different OER communities.

Future research must involve the development of well-defined classification structures for teaching and learning, as well as tests and validations with different users' profiles. In this sense, classification models based on age and user profile should be investigated.

## Acknowledgment

The authors are grateful to Angela Cristina Pregnolato Giampedro for proofreading. Thanks also to the reviewers and specialists who have contributed to this research.

## Appendix A - Dataset

The data generated by this study are available at the following address: https://doi.org/10.5281/zenodo.5543943.

## Appendix B - Selected Studies

| Id | Title |
|---|---|
| S01 | Cross-repository aggregation of educational resources |
| S02 | Developing a metadata schema for CSERD: a computational science digital library |
| S03 | Promoting MERLOT Communities Based on OERs in Computer Science and Information Systems |
| S04 | Prerequisites between learning objects: Automatic extraction based on a machine learning approach |
| S05 | Building a knowledge repository of educational resources using dynamic harvesting |
| S06 | An approach for description of Open Educational Resources based on semantic technologies |
| S07 | Transfer Learning to Timed Text Based Video Classification Using CNN |
| S08 | Towards Automatic Competence Assignment of Learning Objects |
| S09 | Adopting the metadata approach to improve the search and analysis of educational resources for online learning |
| S10 | Web-Based Learning Object Search Engine Solution Together with Data Visualization: The Case of MERLOT II |
| S11 | Supporting Students to Find Relevant Learning Resources through Social Bookmarking and Recommendations |
| S12 | Keywords extraction for automatic indexing of e-learning resources |
| S13 | Building a Search Engine for Computer Science Course Syllabi |
| S14 | Domain Categorization of Open Educational Resources Based on Linked Data |
| S15 | Integrating word embeddings and document topics with deep learning in a video classification framework |
| S16 | Linking Educational Resources on Data Science |
| S17 | Open educational resources in computer science teaching |
| S18 | A semantically enriched context-aware oer recommendation strategy and its application to a computer science oer repository |
| S19 | Ontological modeling of educational resources: a proposed implementation for greek schools |
| S20 | Bridging learning analytics and cognitive computing for big data classification in micro-learning video collections |
| S21 | Bd2k training coordinating center's erudite: the educational resource discovery index for data science |
| S22 | Proposed recommender system for open educational resources for informatics institute for postgraduate studies |

TABLE IV
SELECTED STUDIES

## References

[1] S. Combéfis, G. de Moffarts, and M. Jonavov, "Tlcs: A digital library with resources to teach and learn computer science," in *Olympiads Informatics*, vol. 13, Online, 2019, pp. 3–20.

[2] E. Tovar, H. Chan, and S. Reisman, "Promoting merlot communities based on oers in computer science and information systems," in *Proc. of the IEEE 41st Annu. Computer Software and Applications Conf. (COMPSAC)*, vol. 2, Turin, 2017, pp. 700–706.

[3] N. Rathod and L. Cassel, "Building a search engine for computer science course syllabi," in *Proc. 13th ACM/IEEE-CS Joint Conf. Digital Libraries*, Indianapolis, 2013, p. 77–86.

[4] M. Mouriño-García, R. Pérez-Rodríguez, L. Anido-Rifón, M. J. Fernández-Iglesias, and V. M. Darriba-Bilbao, "Cross-repository aggregation of educational resources," *Comput. Educ.*, vol. 117, no. 1, pp. 31–49, Feb. 2018.

[5] M. R. Hameed and A. S. El-Ameer, "Proposed recommender system for open educational resources for informatics institute for postgraduate studies," in *Int. Eng. Conf. (IEC)*, Erbil, 2019, pp. 126–130.

[6] Z. Kastrati, A. S. Imran, and A. Kurti, "Integrating word embeddings and document topics with deep learning in a video classification framework," *Pattern Recognit. Lett.*, vol. 128, no. 1, pp. 85 – 92, Dec. 2019.

[7] F. Gasparetti, C. D. Medio, C. Limongelli, F. Sciarrone, and M. Temperini, "Prerequisites between learning objects: Automatic extraction based on a machine learning approach," *Telemat. Inform.*, vol. 35, no. 3, pp. 595 – 610, Jun. 2018.

[8] J. L. Ambite, J. Gordon, L. Fierro, G. A. P. C. Burns, and J. Mathew, "Linking educational resources on data science," in *Proc. AAAI Conf. Artificial Intell.*, Honolulu, 2019, pp. 9404–9409.

[9] M. Leake and C. M. Lewis, "Recommendations for designing cs resource sharing sites for all teachers," in *Proc. 2017 ACM SIGCSE Tech. Symp. Comput. Sci. Educ.*, Seattle, 2017, p. 357–362.

[10] M. Leake and C. Lewis, "Designing a new system for sharing computer science teaching resources," in *Proc. 19th ACM Conf. Comput. Supported Cooperative Work and Social Computing Companion*, San Francisco, 2016, pp. 321–324.

[11] B. Kitchenham and S. Charters, “Guidelines for performing systematic literature reviews in software engineering, version 2.3,” School Comput. Sci. Math., Keele Univ., Keele, U.K., Tech. Rep. EBSE-2007-01, 2007.

[12] Unesco, “Forum on the impact of open courseware for higher education in developing countries: Final report,” United Nations Educational, Scientific and Cultural Organization, Paris, France, Tech. Rep. CI.2002/CONF.803/CLD.1, 2002.

[13] O. V. Semenikhina, M. G. Drushlyak, Y. A. Bondarenko, S. M. Kondratiuk, and I. M. Ionova, “Open educational resources as a trend of modern education,” in *Proc. 42nd Int. Conv. Inform. and Commun. Technology, Electron. and Microelectronics*, Opatija, 2019, pp. 779–782.

[14] T. Staubitz, R. Teusner, and C. Meinel, “Towards a repository for open auto-gradable programming exercises,” in *2017 IEEE 6th Int. Conf. Teaching, Assessment, and Learning Eng.*, Hong Kong, 2017, pp. 66–73.

[15] J. Silva and I. Silveira, “A systematic review on open educational games for programming learning and teaching,” *Int. J. Emerg. Technol.*, vol. 15, no. 09, pp. 156–172, May 2020.

[16] X. Wang, M. Liu, Q. Li, and Y. Gao, “A bibliometric analysis of 15 years of research on open educational resources,” in *Proc. 25th Int. Conf. Comput. Educ*, Christchurch, 2017, pp. 296–305.

[17] X. Ren, “Open and interactive publishing as a catalyst for educational innovations,” in *Proc. 32nd nnu. Conf. Australasian Soc. for Comput. in Learning in Tertiary Educ.*, Melbourne, 2015, pp. 248–257.

[18] J. Chicaiza, N. Piedra, J. Lopez-Vargas, and E. Tovar-Caro, “Domain categorization of open educational resources based on linked data,” in *Int. Conf. Knowledge Eng. and the Semantic Web*, Cham, 2014, pp. 15–28.

[19] C. T. Force, “Computing curricula 2020: Paradigms for global computing education,” Assoc. for Computing Machinery (ACM) and IEEE Comput. Comput. (IEEE-CS), New York, Tech. Rep. 10.1145/3467967, 2020.

[20] D. Wiley, T. J. Bliss, and M. McEwen, “Open educational resources: A review of the literature,” in *Handbook of Research on Educational Commun. and Technology*. New York: Springer, 2014, pp. 781–789.

[21] L. Yuan, S. MacNeill, and W. G. Kraan, “Open educational resources-opportunities and challenges for higher education,” University of Bolton, Bolton, U.K., Tech. Rep. 2008-01, 2008.

[22] Y. Hong-yan, X. Jian-liang, W. Mo-ji, and J. Xiong, “Development of domain ontology for e-learning course,” in *IEEE Int. Symp. IT in Medicine & Educ.*

[23] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, “Systematic mapping studies in software engineering,” in *Proc. 12th Int. Conf. Evaluation and Assessment in Software Eng.*, Swindon, 2008, p. 68–77.

[24] R. Queirós and J. P. Leal, “Babelo — an extensible converter of programming exercises formats,” *IEEE Trans. Learn. Technol.*, vol. 6, no. 1, pp. 38–45, Jan-Mar. 2013.