Research on Resource Integration Method of Civil Engineering Construction Course Based on Big Data Mining

Hong-wei Tian(✉)

School of Road Bridge and Architecture, Chongqing Vocational College of Transportation, Chongqing 400021, China

Abstract. Due to the characteristics of large amount of data, scattered storage, and complex data media in civil engineering construction course resources, the application of traditional resource integration methods has problems of poor integration effect and application performance. Therefore, civil engineering construction based on big data mining is proposed. Curriculum resource integration method. Set up association rules, and use big data mining technology to get the initial data results of civil engineering construction course resources. Respectively through the standardization, word segmentation and other steps to achieve the pre-processing of the civil engineering construction course resources. Extract the characteristics of civil engineering construction curriculum resource data, calculate the feature weights and get the comprehensive feature fusion result. According to the feature extraction results, the resource types are divided, and the curriculum resources of the same type are integrated to obtain the final integration result. Through comparison with traditional resource integration methods, it is found that the designed curriculum resource integration optimization method can reduce the packet loss rate of the integration result. The application of the resource integration method effectively improves the retrieval speed of construction curriculum resources, so it has application and promotion value.

Keywords: Civil engineering · Curriculum integration · Construction resources · Big data mining technology · Feature weight

1 Introduction

Civil engineering is the general term of science and technology for the construction of various engineering facilities. It refers to the applied materials, equipment, survey, design, construction, maintenance, repair and other technical activities, and also refers to the object of engineering construction. In order to provide sufficient application-oriented talents for civil engineering construction industry, major universities have established civil engineering specialty [1]. The major of civil engineering cultivates senior engineering and technical personnel who master the basic theory and knowledge of various civil engineering disciplines, and can engage in planning, design, construction, management

W. Fu et al. (Eds.): eLEOT 2021, LNICST 390, pp. 325–337, 2021.
https://doi.org/10.1007/978-3-030-84386-1_27

and research in the fields of housing construction, underground building, road, tunnel, bridge construction, hydropower station, port and offshore structure and facilities, water supply and drainage and foundation treatment.

Civil engineering teaching is a synthesis of traditional and modern technical knowledge. The content of traditional construction technology is still being followed, but modern new technologies, new materials, new techniques, and new methods have greatly changed the construction technology, making it a collection of modern high-tech A compulsory course for applied undergraduate students closely integrated with traditional technology, construction technology and construction management, theory and practice. The characteristics of the civil engineering construction course are strong practicality, strong comprehensiveness, wide range of knowledge involved, rapid development, etc., in the application-oriented undergraduate professional courses, the status of this course is second only to reinforced concrete courses. The reason for this is that the applied undergraduate program focuses on application. The future goal of cultivated college students is to become engineers. The biggest feature of engineers is to make engineering. The goal of professional course teaching is to train students in practical engineering. Therefore, civil engineering construction occupies a relatively important position in similar courses. In the teaching process of civil engineering, architectural teaching needs to be combined with engineering examples to cultivate students' practical ability.

In order to maximize the application of civil engineering construction curriculum resources, the integration and sharing of construction curriculum resources are realized under the network environment. At present, the commonly used integration methods include metadata based integration method, cloud computing based integration method and structured P2P network based resource integration method. However, the traditional integration methods have the problems of low application performance and high packet loss rate. Therefore, data mining technology is used to optimize the design of traditional integration methods. Data mining generally refers to the process of automatically searching hidden information with special relationship from a large number of data. Firstly, the real-time data of the actual civil engineering construction site are collected and stored in the data warehouse. Then, using data mining technology, combined with the curriculum needs to develop association rules, we get the mining results of civil engineering construction curriculum resources, and preprocess the mining data of civil engineering construction curriculum resources. According to the preprocessing results, we extract the data characteristics of civil engineering construction curriculum resources, and calculate the feature weight of civil engineering construction curriculum resources. This paper divides the types of civil engineering construction curriculum resources and realizes the integration of civil engineering construction curriculum resources. The application of this technology provides basic data guarantee for the integration of resources, so as to improve the integration effect of curriculum resources.

2 Design of Resource Integration Method for Civil Engineering Construction Course

The integration of civil engineering construction curriculum resources is a systematic work, which needs to be orderly characterized by the characteristics and content characteristics of civil engineering construction curriculum resources in accordance with fixed

steps and procedures under certain principles, so as to achieve the purpose of resource orderly processing, so as to realize the co construction and sharing of civil engineering construction curriculum resources. The integration of civil engineering construction curriculum resources can be divided into four steps: unified planning of resources, collection of resources, processing and management of resources, and integration and release of resources.

2.1 Mining Resource Data of Civil Engineering Construction Courses

The mining of civil engineering construction course resources is mainly divided into two steps. The first step is to collect real-time data in the actual civil engineering construction site and store it in the data warehouse [2]. Then, using the data mining technology, combined with the needs of the course, the association rules are established, and the mining results of civil engineering construction course resources are obtained.

The timeliness, reliability and completeness of data collection at the construction site are necessary for the dynamic management of the project. The specific construction process is shown in Fig. 1.

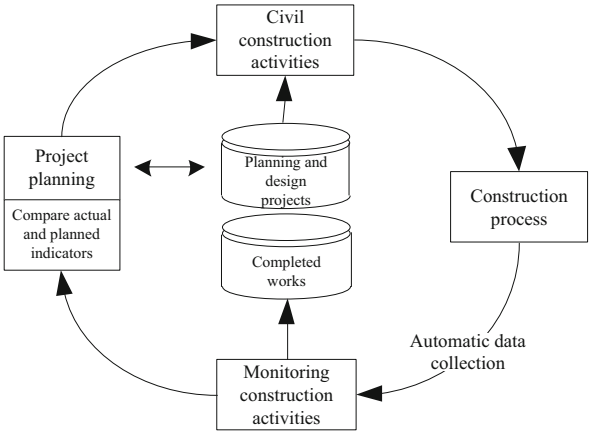

Fig. 1. Civil engineering construction flow chart

The data collected in the construction stage provides a good information platform for the safety monitoring and operation maintenance of the project. According to different uses, the automatic data acquisition technology of construction site can be divided into the following categories: automatic identification technology, positioning and tracking technology, image acquisition technology and sensor and intelligent monitoring technology.

On this basis, data mining technology is used to mine and collect the initial data of civil engineering construction course resources. The corresponding data mining model is shown in Fig. 2.

In the actual data mining process of civil engineering construction course resources, the mining model shown in Fig. 2 is used, and the association rules between the course

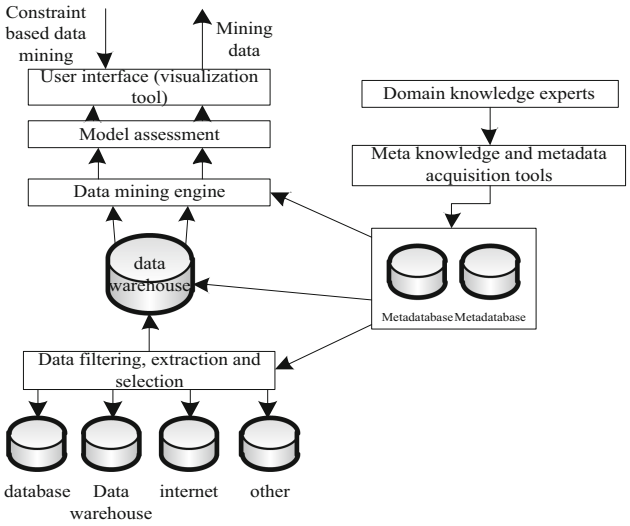

Fig. 2. Data mining model structure diagram

resources are used for specific data mining processing. Association rules are also called association patterns, which are logically implied expressions of the form $X \rightarrow Y$, where X and Y are judgments about the value of attributes in the database [3]. Let I be a collection of m different projects, D is a civil engineering construction database, and each transaction T is a collection of a group of projects in I. Each transaction or transaction T is associated with a unique identifier TID. For item set X, if X belongs to transaction T, then transaction or transaction T supports X. If there are k items in X, then X is also called k-items set, or the length of X is k. Association rules refer to an implicit data relationship in the following form. The specific relationship expressions and constraints are as follows:

$$\begin{cases} X \rightarrow Y \\ X \subseteq I, Y \subseteq I \\ X \cap Y \neq \varphi \end{cases} \tag{1}$$

In order to mine meaningful association rules, two thresholds should be given: minimum support and minimum confidence. The former represents the minimum requirement of a group of data sets in statistical sense, while the latter reflects the lowest confidence of users on association rules. In a given item set, the user support level is greater than or equal to the minimum support level of the item set. The task of association rule mining is to find all frequent association rules in D in a given transaction or transaction database D. Frequent association rules refer to those rules whose support is greater than or equal to the minimum support threshold given by users, and the confidence level is also greater than or equal to the minimum confidence threshold given by users.

Follow the set association rules to realize the mining of civil engineering construction course resource data. A complete data mining process includes the establishment of mining object data mart, the statistical work of small data samples, online analysis, the

establishment of data mining model and model optimization [4]. Each step is completed in a certain order, of course, there will be feedback between steps throughout the process. The process of data mining is not automatic, and many tasks need to be done manually. The business object studied in data mining is the foundation of the entire process, it drives the entire data mining process, and is also the basis for testing the final results and guiding analysts to complete data mining. The output result of the final data mining model is the final mining result of the civil engineering construction course resource data.

2.2 Data Preprocessing of Civil Engineering Construction Course Resources

The pre-processing of civil engineering construction curriculum resources mainly aims at the processing of resources with different formats, including data resources, text resources, image resources and image resources. The preprocessing operation can not only remove a large number of noise data in the text, but also provide the required corpus for the subsequent classification process [5]. Taking the text resources of civil engineering construction course resources as an example, because the text is edited and sorted by different users, in order to avoid garbled code in the process of word segmentation, it is necessary to standardize the coding method before word segmentation. The conversion principle of unstructured data is shown in Fig. 3.

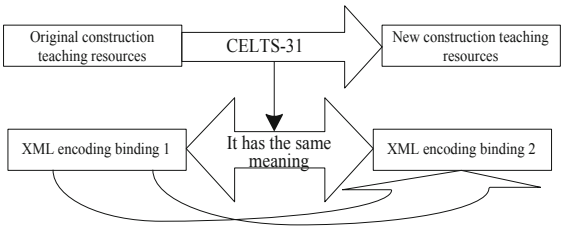

Fig. 3. Schematic diagram of unstructured data conversion

On this basis, Chinese word segmentation is carried out. The operation is to segment the whole text according to certain special rules, and form feature items with words, words or phrases as the basic unit. In the process of processing, the string is matched with the entries in a large enough machine dictionary according to certain rules. If the match is successful, the string is saved as a word segmentation result. In addition, stop words refer to meaningless features generated after word segmentation. The purpose of removing stop words is to reduce the spatial dimension of feature items to be processed in the process of text classification and improve the classification efficiency [6]. After word segmentation of the text, the obtained characters and words form a feature set, and then each feature item in the feature set is matched with the stop word in the stop word list. If the matching is successful, the feature is removed from the feature set; if the matching fails, the feature is retained.

Finally, the processed civil engineering construction course resource data is standardized, so that the processed data becomes data conforming to the standard normal distribution, that is, the average value is 0 and the standard deviation is 1. The standardization process is shown in formula 2.

$$x = \frac{X - \mu}{\sigma} \tag{2}$$

Where X is the original course resource data sample, μ and σ represent the mean value and standard deviation of all sample data respectively. After using the standardized method to process the data, the new training set data due to the square difference is standardized, each dimension of the feature vector value is equivalent processing, each dimension features are subject to the mean value of 0, variance 1 normal distribution, avoid the weight imbalance caused by the difference of each feature vector value.

2.3 Extracting Data Characteristics of Civil Engineering Construction Course Resources

Among the massive original feature items, most of them have little effect on the classification results. The purpose of feature selection is to filter out the most powerful feature items from these massive original feature items for retention, so as to achieve the purpose of reducing the dimension of the high-dimensional feature vector space, reduce the amount of calculation in the subsequent classification process, and avoid occurrences. "Dimensional disaster". The relevant knowledge of statistics and machine learning provides a good theoretical basis for the construction of feature selection methods [7]. The specific features that need to be extracted include document frequency, mutual information, and information contribution rate of the resource. The document frequency refers to the proportion of documents with a certain feature item in the total number of documents in the corpus. The calculation formula is as follows:

$$DF(T_i) = \frac{N_{T_i}}{N_{total}} \tag{3}$$

Where N_{T_i} and N_{total} represent the number of documents with feature T_i and the total number of documents respectively. Set a threshold M in advance, and calculate the document frequency value DF of all feature items. The relationship between the feature T_i and the document frequency threshold is judged. If it is greater than the threshold value, the feature term will be retained, otherwise the feature item will be eliminated. Mutual information is an index to measure the degree of correlation between resource data:

$$MI(T_i, C_i) = P(C_i) \sum_{i=1}^{n} \lg \frac{P(T_i|C_i)}{P(T_i)} \tag{4}$$

Among them, $MI(T_i, C_i)$ represents the mutual information between the feature item T_i and the text, n represents the text of n categories in the corpus, $P(C_i)$ represents the probability of belonging to category j, $P(T_i|C_i)$ represents the frequency of the document that the feature item appears in the category, and $P(T_i)$ represents the feature The document frequency of the item.

2.4 Calculating the Characteristic Weight of Civil Engineering Construction Course Resources

The feature extraction of civil engineering construction curriculum resources is related to its feature weight, so it is necessary to calculate its weight value. The weight calculation function based on feature contribution considers that the clustering of resources is mainly based on the similarity between resources, and the similarity calculation of resources mainly depends on the characteristics of resources. Therefore, the importance of the feature can be characterized by examining the contribution of features to document similarity calculation [8]. Generally, the point product of resource feature matrix is used to calculate the similarity of resources, as shown in formula 5.

$$sim\left(d_i, d_j\right) = \sum_t f\left(t, d_i\right) \times f\left(t, d_j\right) \tag{5}$$

In the formula, $f\left(t, d_i\right)$ is the weight value of feature item t in document d_i. Then the contribution of a certain feature item to the document collection is calculated by the following formula:

$$TC = \sum_{i,j \cap i \neq j} f\left(t, d_i\right) \times f\left(t, d_j\right) \tag{6}$$

All the features in the text set are arranged in descending order according to their contribution degree, and their contribution degree is used to represent the feature weight, and the comprehensive feature extraction results of civil engineering construction curriculum resources are obtained.

2.5 Classification of Civil Engineering Construction Course Resources

According to the CELTS-31 standard, the material types of this course are classified, as shown in Table 1.

In addition to the resource format of the civil engineering construction course, it can also be classified according to the content of the resource. The selected classification algorithm is the naive Bayes classification algorithm, which is a statistical classification algorithm based on Bayes' theorem [9]. The core idea is to calculate the posterior probability of each category of the text to be classified according to the known conditional probability and prior probability, and divide it into the category with the largest posterior probability. At the same time, we assume that the occurrence probability of all texts and the occurrence probability of each feature item are independent of each other. The calculation formula is as follows.

$$P\left(D \middle| C_j\right) = \prod_{11} P\left(W_k \middle| C_j\right) \tag{7}$$

Suppose that there are N category of texts in the corpus as C_n, and assume that D is arbitrary text, which is expressed as vector W_k and recorded as comprehensive

Table 1. Types of course resources and materials

Primary classification	Secondary classification	Classification description
Media resources	Text resources	Used to store basic information, including text, numbers and symbols
	Image resources	The image is stored in the form of vector graphics file; the image is stored in bitmap format after being digitized by scanning, digital camera, camera and other input devices
	Video resources	Store in video format
	Animation resources	Make animation of civil engineering construction and store it in the form of animation or video
Network courseware		Network version of teaching software, which is a detailed explanation of one or more knowledge points in the course, can run in the network operating platform through the browser, and can realize resource sharing in the network environment. The courseware can also be a stand-alone version, which can be downloaded and used through the network
Test paper materials		It includes the collection of test questions, analysis and analysis
Question bank		The teaching measurement of mathematical model is established to realize the collection of test questions in each learning stage of the course in the teaching resource platform
Case		The representative content of one or more civil engineering projects

feature vector. According to Bayes theorem, the posterior probability of class C_i can be expressed as follows:

$$P(C_j|D) = \frac{P(C_j)P(\mathrm{D}|C_j)}{P(\mathrm{D})} \tag{8}$$

According to the calculated posterior probability result, the category of text D can be judged, and the text D can be classified into the category with the highest posterior probability.

2.6 Realize the Integration of Civil Engineering Construction Course Resources

Curriculum integration does not have a fixed mode, it is a diversified curriculum design mode, which emphasizes the teaching process. We believe that the integration of engineering basic courses should mainly meet the following integration principles: the overall

principle of achieving the curriculum goal, the principle of multi-dimensional curriculum content optimization and integration, the principle of basing on the characteristics of the profession and the principle of effective integration of resources [10–12]. The classification results and feature extraction results of comprehensive curriculum resources are integrated, and all the civil engineering construction curriculum resources excavated are integrated, and the text curriculum resources and video curriculum resources integration results are obtained respectively.

The storage methods of civil engineering construction course resources can be divided into directory storage mode and database storage mode. The file directory storage mode mainly stores the resources in different directories of the server according to the classification results of civil engineering construction course resources, and operates and manages the resources through the operating system of computer and network platform.

3 Comparative Experiment Analysis

In order to test the integration effect and application performance of the civil engineering construction curriculum resource integration method based on big data mining, a comparative experiment was designed and the quantitative comparison results were obtained.

3.1 Experimental Environment Construction

This experiment takes the civil engineering construction teaching platform as the experimental environment, and configures the experimental environment. The specific configuration is shown in Table 2.

In this experimental environment, the integration method of civil engineering construction course resources based on big data mining is transformed into program code that can be read directly by computer and put into the experimental environment. When the plug-in installation and operation successful interface pops up in the civil engineering construction teaching platform, it proves that the integrated design method runs successfully in the experimental environment.

3.2 Experimental Data Set

Collect civil engineering construction data and build a corpus as an experimental data set. The specific resource data are shown in Table 3.

3.3 Experimental Process

In order to form an experimental comparison, the traditional resource integration method and the cloud computing-based integration method were set as the two comparison methods of the experiment, and they were imported into the experimental environment in the same way. Because the designed integration method uses data mining technology, it is necessary to set up data mining association rules in the experimental environment. Through the integration of three kinds of civil engineering construction course resources, the final integration result is obtained.

Table 2. Platform configuration environment table

Hardware environment	Application server	Database server	Client
Hardware configuration	CPU: Intel (R) Celeron (R) CPU 2.40 GHZ stepping01 Memory 1048256k ID: ST380817AS 80G SATA	CPU: Intel (R) Celeron (R) CPU 2.40 GHZ stepping01 Memory 1048256k ID: ST380817AS 80G SATA	CPU: Intel (R) Celeron (R) CPU 2.40 GHZ stepping01 Memory 1048256k ID: ST380817AS 80G SATA
Software configuration	OS: Microsoft Windows Server 2003. Standard JDK 1.5.0_06 Tomcat	OS: Microsoft Windows Server 2003. Standard JDK 1.5.0_06 Oracle10g	Window 2000 Professional (SP2) IE6. 0.2900 2180. xpsp_sp2
Web environment	10MLAN Huawei routers and switches	10MLAN Huawei routers and switches	10MLAN Huawei routers and switches

Table 3. Experimental data set

Numbering	Name	Source	Platform language database
1	Education resource library in e-era	APBABI	UNIX/JSP/ORACLE
2	Educational Digital Library	development by mandate	WIN2000/ASP/MSSQL
3	Teaching synchronization resource package	Self development	WIN2000/ASP/MSSQL
4	City civil engineering construction data management system	CNKI	WIN2000/ASP.net/MSSQL
5	Educational electronic journals	K12	WIN2000/ASP.net/
6	Intelligent question bank system	Tongfang	WIN2000/JSP/MYSQL
7	District a education resource library	Entrusted development, central audio visual education center	WIN2000/ASP/MSSQL

(*continued*)

Table 3. (*continued*)

Numbering	Name	Source	Platform language database
8	District B education resource library	Self development	WIN2000/ASP.net/MSSQL
9	District C education resource library	Self development	WIN2000/ASP.net/MSSQL
10	District D education resource library	Self development	LINUX/JSP/ORACLE
11	E District Education Resource Library	Self development	LINUX/JSP/ORACLE

3.4 Analysis of Experimental Results

The experimental test indicators are set to integrate packet loss rate and retrieval speed. The integrated packet loss rate is mainly used to test the integration effect of the design method, and the retrieval speed is used to reflect the application performance of the integration method in the civil engineering construction teaching platform.

Comparison Results of Integrated Packet Loss Rate

Through the statistics of the integration result data of the three resources, the test results about the integrated packet loss rate are obtained, as shown in Table 4.

Table 4. Comparative data of integration effect of construction curriculum resources

Input course resource data/GB	Data loss of traditional teaching resources integration method/GB	Data loss of integration method based on cloud computing/GB	Data loss of civil engineering construction course resource integration method based on big data mining/GB
10	1.14	0.85	0.21
20	2.02	1.23	0.34
30	3.11	1.66	0.64
60	5.62	2.04	0.89
80	10.51	4.28	1.34

It can be seen from Table 4 that the packet loss rates of the three integration methods are 11.9%, 5.79%, and 1.82%, respectively. It can be seen that the designed integration method has a lower packet loss rate, that is, the integration effect is better.

Retrieval Speed Comparison Results

The retrieval speed is mainly obtained by obtaining the background operation data of the civil engineering construction teaching platform. Collecting the background data before and after the application of the integration method can obtain the resource retrieval time. The statistical results of retrieval time-consuming are obtained through multiple experiments, as shown in Table 5.

Table 5. Statistical comparison results of retrieval time

Experiment number	Search content input time	Output time of retrieval results without application of integration method	Output time of search results using integration method
1	08:00:00	08:00:39	08:00:06
2	08:05:00	08:05:44	08:05:05
3	08:10:00	08:10:23	08:10:07
4	08:15:00	08:15:38	08:15:04
5	08:20:00	08:20:36	08:20:06

Through the calculation of the data in Table 5, it can be found that the average retrieval time of civil engineering construction curriculum resources before application is longer than that after application, which shows that the retrieval speed of curriculum resources has been improved after the application of integration method.

4 Concluding Remarks

The network teaching platform of civil engineering course design is the result of teaching reform in major colleges and universities combining the needs of engineering education development and learning from advanced educational concepts. Through the application of big data mining technology, a resource integration method for civil engineering construction courses is designed and implemented, which provides effective auxiliary tools for the teaching of civil engineering. However, in the process of research, the data of civil engineering construction curriculum resources are not compressed, so the integration time of civil engineering construction curriculum resources is still long. In order to improve the integration efficiency, the following research will further preprocess the data of civil engineering construction curriculum resources, so as to shorten the integration time of civil engineering construction curriculum resources.

References

1. Roure, B., Anand, C., Bisaillon, V., et al.: Systematic curriculum integration of sustainable development using life cycle approaches: the case of the civil engineering department at the Université de Sherbrooke. Int. J. Sustain. High. Educ. **19**(3), 589–607 (2018)

2. Kuzucuoglu, D., Ayhan, B.U., Tokdemir, O.B.: Integrating lean construction into the civil engineering curriculum*. Int. J. Eng. Educ. **35**(6), 1787–1802 (2019)
3. Kordi, N.E., Zainuddin, N.I., Taruddin, N.F., et al.: A study on integration of building information modelling (BIM) in civil engineering curricular. In: IOP Conference Series: Materials Science and Engineering, vol. 849, no. 1, p. 012018 (8pp) (2020)
4. Al-Harthy, A., Waris, M., Sana, A.: Assessments of the civil engineering program using the new abet accreditation criteria. Luis Gómez Chova, pp. 861–867 (2019)
5. Pan, C., Chen, K., Shen, X.: Research on teaching reform of civil engineering construction course under the background of new engineering. In: IOP Conference Series: Earth and Environmental Science, vol. 510, no. 6, p. 062011 (6pp) (2020)
6. Bae, Y., Woo, S.: Study of satisfaction levels on construction engineering education in cyber university and its improvements. KSCE J. Civ. Eng. **23**(2), 499–506 (2018)
7. Arashpour, M., Wakefield, R., Abbasi, B., et al.: Optimal process integration architectures in off-site construction: theorizing the use of multi-skilled resources. Archit. Eng. Des. Manag. **14**(1–2), 46–59 (2018)
8. Liu, S., Lu, M., Li, H., et al.: Prediction of gene expression patterns with generalized linear regression model. Front. Genet. **10**, 120 (2019)
9. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mob. Netw. Appl. **24**(1), 1–4 (2019)
10. Liu, S., Glowatz, M., Zappatore, M., et al. (eds.): e-Learning, e-Education, and Online Training, pp. 1–374. Springer, Heidelberg (2018)
11. Sethi, A., Khan, R.A.: Curriculum integration: from Ladder to Ludo. Med. Teach. **42**(12), 1–3 (2019)
12. Waterman, K.P., Goldsmith, L., Pasquale, M.: Integrating computational thinking into elementary science curriculum: an examination of activities that support students' computational thinking in the service of disciplinary learning. J. Sci. Educ. Technol. **29**(1), 53–64 (2020)