

## Retraction

# Retracted: Construction of Multimedia English Teaching Resource Base Based on *K*-Means Clustering Algorithm

### Security and Communication Networks

Received 17 October 2023; Accepted 17 October 2023; Published 18 October 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our own Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

### References

- [1] J. Yan, L. Wu, and Q. Wang, "Construction of Multimedia English Teaching Resource Base Based on *K*-Means Clustering Algorithm," *Security and Communication Networks*, vol. 2022, Article ID 4397617, 12 pages, 2022.

## Research Article

# Construction of Multimedia English Teaching Resource Base Based on K-Means Clustering Algorithm

Jinyan Yan <sup>1</sup>, Lianfang Wu,<sup>1</sup> and Qinghong Wang<sup>2</sup>

<sup>1</sup>Basic Teaching Department, Aba Vocational College, Maoxian, Sichuan 623200, China

<sup>2</sup>Economic Management Department, Aba Vocational College, Maoxian, Sichuan 623200, China

Correspondence should be addressed to Jinyan Yan; 18990432506@163.com

Received 27 May 2022; Revised 27 June 2022; Accepted 12 July 2022; Published 26 August 2022

Academic Editor: Tao Cui

Copyright © 2022 Jinyan Yan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Because of the explosive development of the Internet today, network education resources are becoming more and more valuable, and so is the rapid development of the information teaching environment. Let us pay more attention to the construction of teaching resources. In order to help schools improve the quality of teaching and promote the reform of basic education in China, the teaching resource bank must be enriched in connotation and perfect in function. However, the current teaching database is still facing some difficulties in the construction process, such as insufficient information, and database content structure relative disorder. In fact, the progress of computer technology is likely to help overcome the above difficulties, but there are still some problems in the construction process of the current teaching resource database, such as the connotation of teaching resources is not full, the organization structure of the content of the education resource database is relatively messy. The development of computer technology is likely to help solve the above problems. Therefore, the research on constructing the service system using computer technology has become a very key and hot topic today. Based on the current situation of resource library construction and the classification and summary of some technical problems, the author of this study puts forward a set of methods and procedures for the construction of a multimedia English teaching resource library. An improved hybrid clustering algorithm ISPO + K-means was proposed to integrate PSO and K-means. Compared with other algorithms, the algorithm clustering results are better. Based on this process, this study can automatically sort and classify the collected learning resources, reduce the consumption of manpower and material resources, and improve the service efficiency of the resource library. The purpose of this study is to find the methods that can use information technology to serve the construction of a multimedia English teaching resource bank, hoping to provide a valuable reference for the related research on the construction of teaching resource bank.

## 1. The Introduction

With the in-depth development of the Internet information technology era, social education has officially entered a new era of highly informationized network education. Through the connection of mobile Internet, teaching and resources have gone beyond the previous time and space scope, and the range of access to knowledge information materials has rapidly expanded from the existing classroom, laboratory, and school library to other places covered by the Internet. However, the network is only one of the information carriers in the dissemination of knowledge and information resources. How to efficiently absorb and share learning information resources is still the concern of the vast majority

of network users. Rich resources effective multimedia network and education technology will be independently personalized to guide students to the active learning process of cooperative learning and creative key, is also help teachers create interactive classroom interactive teaching, orientation training, and other interactive educational activities necessary foundation support and guarantee conditions for success [1]. In the field of educational theory and information computer science, many experts and scholars at home and abroad have invested a lot of energy, time and money, and material resources to study and discuss various new network education information technology schemes, education information platform construction and advanced application of interactive virtual learning technology to

create a new interactive teaching environment. However, it is worth noting that no matter how valuable the program, platform resources, and environment are, their value cannot be separated from the support of quality educational resources [2]. It can be seen that the government should pay more attention to the construction of high-quality resources for public education. Therefore, how to efficiently collect and manage information and optimize the campus information resource database has gradually become an increasingly important practical subject for schools, especially to explore how to make the school information technology directly serve the optimization construction of campus education resource database, which has been a long-term prominent issue of concern in the field of university education and technology [3].

At present, the rapid development of information technology innovation involves a variety of information resource data automatic collection, sorting and classification, and rapid retrieval, including automatic network information crawler, automatic classification, information batch extraction, automatic query questions, data mining, natural language analysis, and processing functions. At the same time, there are all kinds of intelligent data collection and automatic retrieval of relevant information data system tools, such as intelligent web crawler system tools and various general intelligent search engines, such as Google, Yahoo, Baidu search engine, and other application systems, are specially equipped with their specific needs [4]. However, because the network users in different information fields have many special requirements of different levels, the general database system can not fully meet the diversified query needs of its users. Especially at present, with the gradual enrichment of various types of network data modules and the continuous in-depth development of the next-generation network technology platform technology, a large number of existing network multimedia resources (image, audio, video, etc.) and access to database resources, Single query technology can not effectively and timely obtain and quickly discover semi-structured or even semi-non-fully structured data. All these need to fully combine the existing advanced technology, in order to fully meet the domestic and foreign different aspects of enterprises and industry users of the diversified service needs [5]. Currently, the connectivity between the various technologies needed to collect and classify resources is largely manual. You can see that in this case, it is necessary to establish a set of automatic resource collection and sorting templates for a specific domain of multimedia English education resource databases [6].

## 2. Research Status

At present, teachers at home and abroad are still gradually committed to the establishment of educational resource databases. Through an extensive collection of high-quality teaching materials, teaching plan information, and other teaching data resources related to teacher training, and further processing and sorting, the basic purpose of fully integrating resources and reasonably sharing the content of

the existing digital teaching database resources is achieved. In this regard, there are three foreign projects: GEM in the United States, Edna in Australia, and Edusource in Canada.

At present, both at home and abroad are devoted to the establishment of educational resource database [7]. Through collecting teachers' teaching materials, teaching plans and teaching resources, and further sorting, to achieve the purpose of integrating and sharing the content of these teaching resources. There are three foreign projects in this regard: GEM in the United States, Edna in Australia, and Edusource in Canada [8].

Gem is an American educational resource database development and application project. One of the main ideas of the initial system design proposed by him was to re-classify the massive educational information integrated by online communities and the rich and diverse resources of global high-quality online education and learning projects through systematic, effective, and rigorous organization and integration design. Further, help school teachers and ordinary young students to choose and use these high-quality network resources safely and efficiently. Its overall design and functions are mature and complete, and the content level is reasonable and rich. It is equipped with a simple and practical search engine system and an interactive online question-and-answer query system. Gem does not require deliberately collecting the most abundant and reliable network information resources, but it is still in essence to build a network information education resource directory system with larger-scale resources. Gemcat, GemHarvest, and Browser Builders are the other three main technology tools used by GEM to develop a platform for building educational resource portals. Using metadata tools from Gemcat standard, users can easily and quickly automatically classify, sort out, query, and describe information data of various education and training resources according to Gemcat standard. It also allows each resource directory creator to automatically generate metadata and control the vocabulary for each resource tree. Metadata is a web page inserted as a meta tag into a resource format. Gemsearch data collection and analysis tool is a featured search engine. Through the adoption of this kind of similar to the robot, the operation mode of the automatic search engines can fast accurate reference to the local contains a large number of gem format metadata resources page, directly from the source web directory can automatically extract relevant resources data, form for the index of local content, and at the same time to realize automatic assembly and add the information to the local portal directory list. The file migration tool effectively and conveniently helps customers to solve a series of technical problems when they do not register address information in the directory file list, or they cannot effectively migrate the registered address directory list and resource file directory.

The second tool, Edna, mainly contains metadata collected through the GemHarvest program to create a simple interactive HTML teaching page, Australia's most famous online education portal for a wide range of teachers. After searching the required resource catalog on the website, the visitor will connect to the website storing the information of these resources through the name index of these resource catalogs.

The focus of the Canadian Education Project and its development is to create a network of teaching and learning resources based entirely on metadata standards as the main framework by updating teaching and learning resources. Another Careo project in Canada is an educational reference site for multilingual subjects that can be accessed free of charge at any time via the Internet. CAREO pays special attention to modular use of learning resources, reuse of resource content, and standardization of learning resource metadata.

### 3. Method Research

**3.1. Collecting and Sorting Models.** In view of the problems existing in the construction of the existing teaching resources database and the specific needs of the existing electronic teaching system, this chapter proposes a structured teaching resources collection and collation method based on the focused crawler technology. It can realize automatic online collection and collation of teaching resources in specific fields on the Web.

The main idea is as follows: first, the teaching resources are trained by using the method of learning from the topic classification structure table, and the characteristic items of the teaching resources of each topic are extracted. At the same time, the common sense and domain knowledge related to the theme are summarized. Then, using this knowledge, the focused crawler is guided to search and collect relevant thematic Web pages on the Web. After the web page information containing structured teaching resource information is successfully captured by the search engine, the technology of structured data extraction is applied to further extract and screen out all the structured web page teaching subject resources actually containing information content, and form a structured file. Finally, all of these files are stored in subject and architecture categories. The index function such as the index table should be established to facilitate query retrieval and retrieval. The model mainly consists of the following parts: focused crawler based on ant colony algorithm, automatic clustering, data extraction, and structured repository engine. A good automatic resource collection and organization model is the basis for improving the quality of the teaching resource library, and each part of the method will be introduced in detail below in Figure 1.

**3.2. Focused Crawler Based on Ant Colony Algorithm.** The overall distribution of information resources in the Web search space is unknown to focus on crawlers, so it is impossible to predict their creeping direction.

Ant colony algorithm technology, not only can simultaneously support a variety of intelligent algorithm search, global performance and optimization, but also has its own sets, positive feedback, distributed computing, and many other features, easy to quickly combine with other advanced algorithm technology. The use of the positive feedback principle can help further accelerate the overall speed of the evolution of human-computer intelligence development;

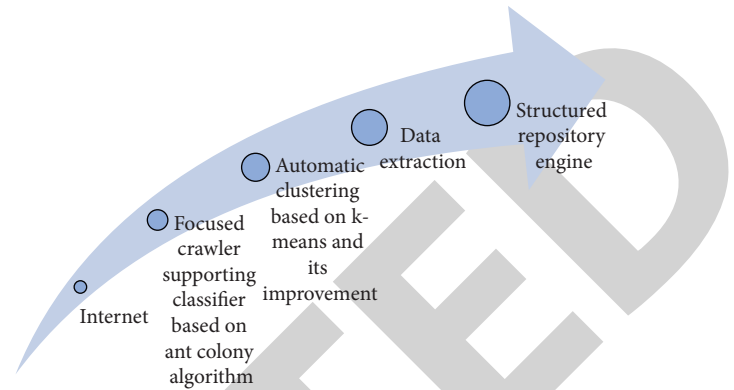


FIGURE 1: The process of building a multimedia English teaching resource library.

distributed parallel computing system facilitates high-speed parallel analysis and rapid execution between various complex system algorithms, and it is easier to carry out high-speed and continuous real-time exchange and sharing of data between the members of the personal computer system, which facilitates the user to quickly find and design the best solution with higher overall performance. It is also easier to fall into the misconception of optimizing the local performance of the system; an easy combination of heuristic algorithms can improve its performance; therefore, variations based on the ant colony algorithm model can be used to solve other problems.

The literature points out that it is much more effective to use topic classifiers to guide a website's commitment to crawl sliders.

In view of the above, we propose a focused crawler module based on the ant colony algorithm using classifier architecture by introducing the ant colony algorithm into the focused crawler search strategy and using the "inspiration" of the ant colony algorithm to guide the focused crawler on the Web.

The core of this module is the design of the focused crawler and the classifier (subject feature item library and knowledge base). The classifier in the focused crawler is mainly responsible for training the behavioral characteristics of the crawl target, calculating the relevance of the target web pages, and classifying the web pages. This module actually also enables the downloading of resources, the automatic filtering of URL content (using heuristics to filter out unwanted URLs), and the automatic filtering of web page content (it also enables the horizontal comparison of the relevance value of the result obtained from the calculation with the threshold value of the given automatic filtering, and if the value is greater than the threshold value, the links are retained and extracted, and the new links are continued to be tracked, otherwise they are discarded). Otherwise, they are discarded).

The web crawling sub-module undertakes the task of connecting the URL evaluation sub-module with the topic relevance analysis sub-module. This module first selects the URLs with the highest hyperlink scores in the URL link seed

set discerned from the URL evaluation sub-module and fetches the corresponding web pages. The web pages are divided into authoritative Hub pages and Content pages. Hub pages, i.e., directory pages, provide users with topic-related websites, and only extract URLs and anchor texts, without entering the topic relevance analysis sub-module, because it must contain many classes. Content pages are entered into the analysis sub-module of topic relevance for processing.

The website subject relevance analysis subtype (sorting module) crawl subproject is loaded, and the relevance to the subject is determined by focusing on the crawler through the subject relevance analysis subtype, thus guiding the subsequent crawler access process. After loading the page, if the value exceeds the threshold of one or more classes, it will be saved to the corresponding class, otherwise, it will be discarded. Fast and accurate access to structured data and metadata information on relevant topic websites is the basis for assessing and classifying the meaning of a topic. The so-called structured data and metadata are key information fields extracted from information sources. The specific approach can be based on keywords (feature elements) or can be traced back to the semantic and conceptual level [9]. The main idea of keyword-based thematic relevance analysis is, first, to determine the structure of the thematic classification system, with the participation of experts in the field; second, the classification system focuses on the taxonomy in reptiles, i.e., the taxonomy of reptiles. E.Education prepared the sample with the participation of experts in this field, identifying a set of features with weight, reflecting the characteristics of a specific field, identifying specific topics, and forming a feature and knowledge base that provides the basis for a targeted collection of information on reptiles. This is followed by linking to URL strings, text on anchors (the advantage of links and web links is that links describe the text more accurately than the page itself, and links describe the text linked to the page, an operation that helps to search for non-textual information), or extracting the page text by keyword [10]. Finally, the page topic is calculated to derive the relevance of the page content to the website topic and determine the threshold value, which determines the page trade-off [11]. The basic principle of the vector space model algorithm is that by converting a text space of a given form (article, query or paragraph in an article, etc.) into a text vector of high spatial dimensionality, and then using their corresponding operator formulas, respectively, the degree of similarity in size between any two of their text vector spaces is calculated, i.e., the correlation in size between the text vector and the two text vectors to which the corresponding operation is to be applied.

The URLs evaluation submodule is used to evaluate the relevancy of URLs parsed from topic-related pages and guide the crawling process of focused crawlers using the ant colony algorithm.

**3.3. Automatic Clustering.** Clustering is an important analysis method in data mining, which has been widely used in business intelligence, Web mining, and other fields. The clustering module can automatically sort the teaching

resources of specific topics collected by web crawlers, reduce manual labor and improve sorting efficiency [12].

The clustering algorithm is a classification process that divides each population into several completely different levels according to the differences in data characteristics between different population levels by using computer technology. The main algorithm principle of clustering analysis of individual data distribution is to make the average data distance between two categories of the same individual population as small as possible, and to make the average data distance between two categories as large as possible. The k-means algorithm and the recent neighbor algorithm are the most flexible, stable, and reliable algorithms in the research of clustering algorithm, and the most flexible and widely applicable in the field [13]. In the K-means algorithm,  $n$  target points are considered to be divided into  $K$  target clusters. First, a representative target point with at least  $K$  targets is selected, and according to the minimum relative distance between it and other representative points, all target clusters of other categories can be considered to be re-divided into another target. Therefore, each class is represented by its center of gravity, and the other objects are re-segmented. This process is performed iteratively until convergence occurs. K-means algorithm has a simple structure and is easy to implement quickly, with higher efficiency and linear time complexity. In addition, it can always converge to the local optimal from the search to the initial position of the object, but the number of cluster centers at the specified position must be calculated in advance [14]. The initial location of cluster analysis center has a great influence on the result of cluster analysis. This is actually a deterministic local search algorithm.

PSO algorithm is a global optimization algorithm based entirely on the characteristics of the population. Its inspiration partly comes from the simulation of the foraging behavior of the bird population. Although the algorithm has a fast convergence speed, it can easily fall into the approximate local optimal solution [15]. When it converges completely to a nearly global optimal solution, the overall accuracy of the solution may be difficult to improve further [16]. The intelligent single particle algorithm proposed in the literature can not only ensure that a single particle can randomly search each region in the search space but also can effectively decompose the single search space into at least several small areas, small particle size search space, so as to better solve the problem of PSO algorithm. Combining k-means algorithm with an intelligent single particle algorithm, an improved ISPO + K-means hybrid clustering algorithm is proposed [17]. The basic processing flow of the automatic clustering module is shown in Figure 2. See Chapter 4 for details.

**3.4. Data Extraction.** After the access to the user-related structured information content, according to customer's demand, also need to automatically extract calculate the related structured web page information in the data, the user information stored on that particular page specific content of the page data, by automatically extracting method,

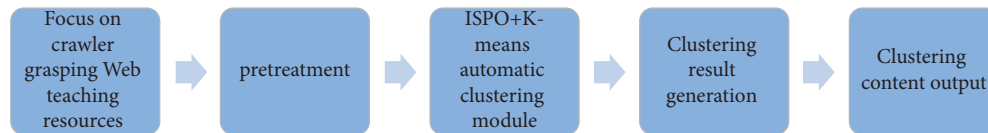


FIGURE 2: Processing flow of automatic clustering module.

structured information data is converted to a particular web page [18]. The structured database is regarded as a relatively small storage unit in the system, and these structured data are stored and processed separately according to the data classification principle of the information subject system and the design of the information architecture. You also need to build indexes and other functions for queries and recovery. Therefore, the data extraction module came into being [19]. The data extraction module is to extract and sort the theme-related web pages collected by the crawler according to format and category. In order to realize the fast and accurate extraction of web content rules and the fast and effective integration of various web data from various web resource environment systems under different levels of use, we need a specially constructed web wrapper library. The main content of the wrapper is to extract the extraction rule code base from each page of website content and a computer program code base that can effectively apply the extraction rule code embedded in all websites. Wrappers should be able to retrieve relevant information from specific information sources and integrate information from different information sources into a database. Web page data extraction system generally adopts the form of plug-in [20]. Through plug-in management services, different plug-ins are used to process web pages with different structures. The advantage of this approach is that it is extensible: in the future, whenever a new type appears, its processing methods can be transformed into plug-ins and added to plug-in management services [21].

The classification module automatically classifies subject-related web pages, stores and maintains them according to the classification structure, and serves as the input of the data extraction module. The data extraction module formulates the corresponding extraction rules and modes according to the characteristics of each discipline field and the learning of the sample web pages marked by experts and generates the corresponding wrapper library. The data extracted by the wrapper is stored as an entity file (Structuralfile) according to a pre-set format, using XML to represent it. It is then processed by the structured repository engine module and stored in the structured repository. As the new structured files are populated, the new feature items are periodically or quantitatively retrained and added to the feature item library of the classification sub-module. This makes the classifiers built in the classification sub-module adaptive [22]. The method of obtaining extraction rules and patterns in the wrapper also has a self-learning function, and the extraction rules and patterns are continuously improved as new samples appear.

There are two kinds of extraction rules and patterns used in the wrapper system design procedure. One main

approach is to use manual methods to prepare extraction rules and procedures directly so that the wrapper system can effectively deal with several specific types of knowledge information extraction-related problems at the same time, which is called the extraction knowledge engineering method [23]. The other one is to learn the extraction rules of the marked corpus and then process the new text through training or learning, this method is called the automatic recognition pattern method, the advantage is fast, but the disadvantage is the high requirement for training data. In this study, we propose to use the two methods, knowledge engineering and automatic recognition mode, together.

**3.5. Structured Teaching Resource Library.** The collection and arrangement of network teaching resources is a continuous and evolving process [24]. The continuous updating of teaching resource standards and requirements requires the teaching resource library to be constantly supplemented and updated timely and automatically; without unified teaching material standards, the downloaded resources cannot be effectively stored in a structured manner and cannot form an effective utilization scheme, thus making it difficult to realize the ease of use, sharing and It is difficult to realize the ease of use, sharing and expandability of online teaching resources. To address these problems, we propose the concept of creating a structured teaching resource library [25].

The structured teaching resource library is defined as a teaching resource library that is built based on structured metadata information related to the topic and pays more attention to specialized application and structural analysis, and these structured metadata have well-defined syntax and clearly defined semantics, which can reflect the status and characteristics of teaching resources in an integrated manner [26].

The structured teaching resource library is based on a certain classification system structure. The classification system structure includes concepts, ontologies, knowledge points, etc. At present, most structured teaching resource libraries are built on the common classification system structure of specific disciplines, which is usually a tree structure. The first-level directory is the main content within the subject (usually corresponding to the chapter of the textbook); the second-level directory is the main knowledge points under this content; the third-level directory is the common types of resources used under this knowledge point (e.g., courseware, materials, exercises, extensions, lesson preparation reference, etc.); the fourth-level directory is the form of resource media presentation (text, pictures, sound, images, etc.); and the fifth-level directory is the specific resources [27]. The nodes at each level of the directory

provide corresponding functions, for example, the functions of adding, deleting, modifying, copying, moving up and down, and exporting of nodes; the functions of limiting the total amount of resources; and the functions of limiting the types of applications of resources.

## 4. Algorithm Experiment

**4.1. Particle Swarm Algorithm.** The particle swarm algorithm is a swarm intelligence optimization algorithm proposed by Kennedy and Eberhart in 1995, inspired by the foraging behavior of bird and fish flocks. In the particle swarm algorithm, it is first initialized as a swarm of random particles, the position of each particle represents a candidate solution, and the degree of superiority of the solution is determined by the fitness function. In each iteration, the particle updates the velocity and position of that particle by dynamically tracking the individual extremum pbest and the global extremum gbest. The pbest represents the optimal solution found by the particle itself, while the gbest represents the optimal solution found so far by the whole population. The iterative search is terminated by reaching a specified number of iterations, or by the positions searched so far satisfying a specified error criterion.

In the PSO algorithm, the particle population is searched in a  $D$ -dimensional space, and if a population is formed by  $n$  particles, the position vector of the  $i$ th particle ( $i = 1, 2, \dots, n$ ) is noted  $z_i = (z_{i1}, z_{i2}, \dots, z_{id}, \dots, z_{iD})$  as. The superiority of the particle position is calculated and measured according to certain criteria [28]. The velocity vector of particle  $i$  is the distance moved by the particle in each iteration,  $v_i = (v_{i1}, v_{i2}, \dots, v_{id}, \dots, v_{iD})$  denoted as, the optimal position searched so far by the  $i$ th particle is  $pbest_i$ , the optimal position searched so far by the whole particle population as,  $gbest_i$ , then the update velocity of the  $d$ th dimensional component of any particle  $i$  in the PSO population at the  $k + 1$  th iteration is determined by equation (1), [29].

$$v_{id}^{k+1} = wv_{id}^k + c_1r_1^k(p_{id}^k - z_{id}^k) + c_2r_2^k(p_{gd}^k - z_{id}^k). \quad (1)$$

The update speed of the particle is composed of three components, the first term is the inertia weight term, which remembers the effect of the particle's previous movement speed; the second term is the cognitive term, which represents the effect of the historical optimal position of the particle population; and the third term is the social term, which represents the effect of the optimal position once possessed by the particle population.

The position of the particle after moving is determined by equation (2), i.e., the sum of the position and movement at the previous moment is

$$z_{id}^{k+1} = z_{id}^k + v_{id}^{k+1}, \quad (2)$$

where  $k$  is the number of iterations and  $r1$  and  $r2$  are random numbers uniformly distributed between  $[0, 1]$ .  $c1$  and  $c2$  are the learning factors. The particles keep following the guidance of local optimization and global optimization to

search in the search space until the number of iterations is reached or the error threshold is satisfied [30].

**4.2. Intelligent Single-Particle Algorithm.** In the traditional PSO algorithm, the overall quality of the solution vector is often judged by changing the values of each dimension in the whole solution vector simultaneously and getting an adaptation value based on the updated solution vector. However, this does not tell whether the partial dimension moves in the optimal direction. To address this problem, the literature proposes the solution of dividing the particle into several small spaces of lower dimensions for searching while ensuring that it can search the whole space. The basic idea is as follows.

The algorithm uses a particle to search in the solution space, and the position vector of the particle is divided into several subvectors, and then the particle is updated by the subvectors. In the process of subvector updating, a new learning strategy is introduced by analyzing the previous velocity updates, so that the particle can dynamically adjust its velocity and position in the search space, thus approaching the global optimum.

The algorithm principle is described in detail as follows.

The first is the subvector. Unlike the traditional particle swarm algorithm, in the intelligent single-particle algorithm, each particle represents the entire position vector, and when updated, the entire  $D$ -dimensional space is divided into  $m$  parts, i.e., the entire position vector is divided into  $m$  position subvectors, and each position subvector and its corresponding velocity subvector are denoted as and,  $j = 1, \dots, j, m$ ,  $v_j$  respectively. For simplicity, assuming that  $D$  is exactly divisible by  $m$ , each position subvector  $l$  ( $l = D/m$ ) includes a dimension.

The second is the update process. The update process of the intelligent single-particle algorithm is that each position subvector is updated cyclically from  $\bar{z}_1$  in  $\bar{z}_m$  sequential order, and the iterations are executed  $N$  times, and the speed and position adjustment equations at update are as follows:

$$\begin{aligned} \bar{v}_j^k &= \left( \frac{a}{k^\rho} \right) \times r + b \times L_j^{k-1}, \\ \bar{z}_j^k &= \begin{cases} \bar{z}_j^{k-1} + \bar{v}_j^k, & f(x_1^k) > f(x_2^k), \\ \bar{z}_j^{k-1}, & f(x_1^k) \leq f(x_2^k), \end{cases} \\ L_j^k &= \begin{cases} \frac{L_j^{k-1}}{s}, & f(x_1^k) \leq f(x_2^k), \\ \bar{v}_j^k, & f(x_1^k) > f(x_2^k), \end{cases} \end{aligned} \quad (3)$$

where the parameter  $L$  denotes the learning variable,  $r$  denotes a random vector uniformly distributed on the interval  $[-0.5, 0.5]$ ; the constants  $a, \rho, b \geq 1s$ , and  $(b)$  denote the diversity factor, descent factor, contraction factor, and acceleration factor, respectively; and  $f()$  is the adaptation value function.

During the subvector update process, the velocity subvector determines the position subvector, and each velocity subvector consists of two parts: the learning part  $b \times L_j^k$  and the diversity  $(a/k^p) \times r$  part. The diversity  $(a/k^p) \times r$  part decreases with the increase of iterations, which will make the particle dynamics gradually switch from global search to local search.

Third is the learning strategy. The intelligence of the ISPO algorithm is mainly reflected in its learning part. It follows a new learning strategy that uses the learning part of the velocity  $b \times L_j^k$  subvector to intelligently adjust the velocity subvector according to the previous velocity updates of the particles so that the velocities of the particles have greater diversity and thus avoid falling into local optima.

The final intelligent single-particle algorithm pseudo-code can be obtained.

**4.3. K-Means Algorithm.** So far, researchers have proposed or designed a variety of classical document clustering algorithms for this purpose, among which the  $K$ -means algorithm is a classical algorithm that can really solve the problems of clustering and analysis of complex documents quickly. The  $K$ -means clustering algorithm is based on the distribution of the probability predictions of the category categories of the most similar documents.

The  $K$ -means clustering algorithm works as follows: first,  $K$  clusters are selected as the initial center of mass, where  $K$  is a pre-specified cluster parameter, i.e., the number of cluster points expected by the user. Each point is assigned to the initial center of the nearest cluster, and each cluster assigned to the initial center is automatically clustered into the next cluster. Then, the center of mass of each cluster is updated according to the points assigned to the cluster. The two steps of assigning and updating are repeated until the clusters do not change anymore, or are almost identical, and the center of mass does not change. The specific steps are as follows.

Input:  $N$  documents to be clustered, number of clusters  $K$ .

Output:  $K$  clusters and convergence of the criterion function.

The method is shown in Figure 3.

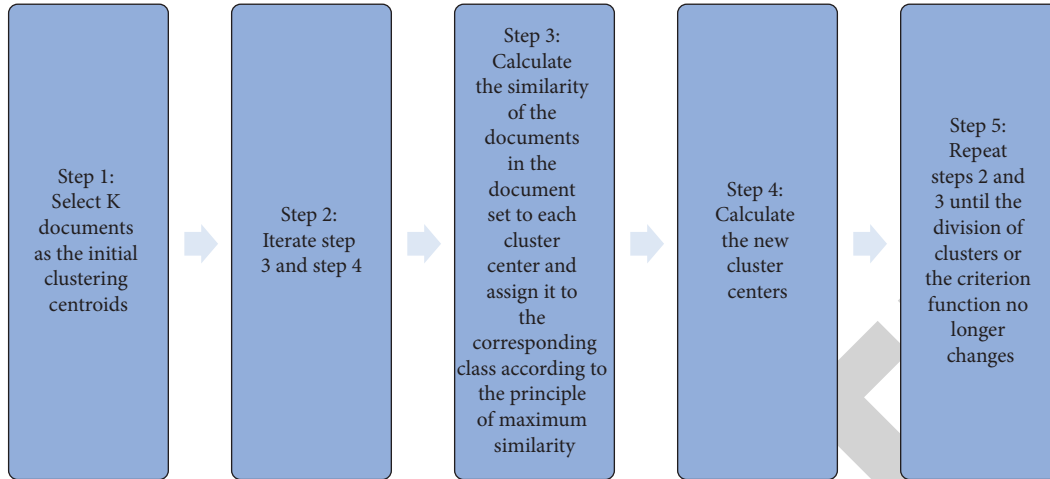
The  $K$ -means algorithm is simple, effective, and fast to process, especially for large document sets. However, the traditional  $K$ -means classification algorithm has two inherent drawbacks: (1) random initial value selection may lead to different clustering results, and even the existence of no solution and more sensitive to the initial clustering center; (2) the algorithm is based on gradient descent algorithm, so inevitably often fall into the local optimum. These two major defects greatly limit its application scope.

#### 4.4. Improved K-Means Algorithm Based on Intelligent Single-Particle

**4.4.1. ISPO + K-Means Algorithm Idea.** To overcome the above-mentioned shortcomings, many studies in recent years have combined the Particle Swarm Optimization (PSO) algorithm to improve the  $K$ -means algorithm. Omran et al. first applied the particle swarm algorithm to cluster analysis in 2002 and showed experimentally that the clustering effect based on the particle swarm algorithm is better than the simple  $K$ -means clustering algorithm. According to the literature,  $k$ -means should be combined with particle swarm for clustering, and it is proposed that in the initialization stage, the fast clustering results of  $K$ -means should be used to initialize and assign values, and then the clustering algorithm of particle swarm is used. In the literature, random functions were used to allocate the initial positions of particles in the initialization process, and then the  $k$ -means clustering algorithm was used for optimization in the subsequent new generation of particles. A method of spatial clustering analysis with obstacle constraints combining PSO algorithm and partition is proposed in this study. This method makes good use of the global solution searching ability of the particle swarm optimization algorithm and fully considers the effect of real obstacles on the clustering results so that the clustering results are more practical. DKPSO, a dynamic clustering algorithm based on particle swarm, is proposed in this study, which can automatically determine the optimal number of clustering centers when running. In the literature,  $k$ -means calculation is performed on all particles in each PSO iteration to seek the optimal solution. Although the clustering performance is improved, the algorithm still fails to get rid of the dilemma that the PSO algorithm falls into the local optimal solution, but greatly increases the computational amount of the algorithm, resulting in slow convergence. The Adaptive particle swarm optimization (APSO) algorithm proposed in the literature can estimate different evolutionary states of the algorithm, and design an effective parameter Adaptive control strategy according to the different evolutionary states, so as to speed up the solution of the algorithm optimization problem.

The clustering methods of combining the PSO algorithm with the  $K$ -means algorithm described in the above literature can be divided into three categories:  $k$ -means first followed by particle swarm optimization can be simply written as  $K$ -means + PSO; PSO first followed by  $k$ -means can be written as PSO +  $K$ -means; The recombination of the above two types is abbreviated as  $K$ -means + PSO +  $K$ -means. Literature has proved through experiments that the latter two combination methods have better effects, and the probability of obtaining the best solution is significantly higher than the  $k$ -means method, basic PSO method, and  $K$ -means + PSO method. The study analyzes the problems existing in the traditional PSO algorithm, proposes the



FIGURE 3:  $k$ -means algorithm flow.

intelligent single particle optimization algorithm, and proves through experiments that the algorithm can overcome the defects of the traditional PSO algorithm. In view of this, the intelligent single particle optimization algorithm is introduced into the clustering analysis, and an improved particle swarm and  $K$ -means hybrid clustering algorithm ISPO +  $K$ -means is proposed.

The ISPO +  $K$ -means clustering algorithm proposed in this study consists of two modules: the ISPO module and the  $K$ -means module. First, the ISPO algorithm is used to find the optimal  $m$  initial clustering centers, then the  $k$ -means algorithm is used to find the clustering results, and finally, the clustering results are output. The  $m$  initial clustering centers found by the ISPO algorithm means that  $M$  initial clustering centers are randomly given for the  $k$ -means algorithm. In the subsequent execution steps of the algorithm, the optimal one is selected from  $m$  kinds of clustering, so that the defect of the  $k$ -means algorithm sensitive to the initial clustering center can be significantly improved.

**4.4.2. ISPO +  $K$ -Means Algorithm Coding and Fitness Function.** ISPO +  $K$ -means clustering algorithm adopts the coding method based on clustering centers. Each particle is encoded with location, velocity, and adaptive value. The dimension of the sample vector is  $D$  and the number of clustering centers is  $M$ , so the location and velocity of particles are  $m \times D$ -dimensional variables and the adaptive value function is  $F()$ .

**4.4.3. ISPO +  $K$ -Means Algorithm Description in Detail**

(1) ISPO Module

Step 1: given data  $D = (x_1, x_2, \dots, x_n)$  set,  $z^0$  represented as position vector of data set  $D$ . The entire position vector is divided into  $m$  position subvectors, i.e.,  $z^0 = (\tilde{z}_1^0, \tilde{z}_2^0, \dots, \tilde{z}_m^0)$  and the fitness is  $f(z^0)$  calculated.

Step 2: initial number of sub-vectors  $j = 1$  and learn  $L_j^0 = 0$  variables.

Step 3: initial sub-vector iterations  $k = 1$ . If the maximum number of iterations  $N$  is reached, skip to Step 4; otherwise, continue. Update the speed of the  $J$ TH subvector according to formulas (5-3) and (5-5). Update the position of the  $J$ TH subvector according to formula (5-4).

Step 4: perform the next subvector according to Step 2 and Step 3. If all the  $m$  subvectors are updated, skip to step 5. Otherwise, continue the execution.

Step 5: output the center point values of  $m$  clusters.

(2)  $K$  Means the Module

Step 1: take the values from  $I$   $x_i \in D = 1$  to  $n$  respectively, and then find the center point from the  $m$  cluster centers selected by ISPO module according to  $O_j$  the nearest  $x_j$  neighbor rule, and  $C_j$  add clustering.

Step 2: recalculate the center points of each new cluster,  $j = 1, 2, \dots, M$ ,  $O_j$  that is, calculate the average value of the members of the new cluster and use this average value as the center point of the new cluster.

Step 3: repeat Step 2 and Step 3 until there is no change in each clustering center, that is, the algorithm converges, and the final clustering result is output.

**4.5. ISPO +  $K$ -Means Algorithm Experimental Design.** All experiments were carried out on a PC with CPU 2.0 GHz and 1G memory, operating system Windows XP and Matlab 6.5. The experiment evaluates the clustering performance of low-dimensional data, high-dimensional data, and multidimensional sample high-dimensional data sets, respectively. We compared the experimental results of ISPO +  $k$ -means with those of PSO +  $K$ -means and  $K$ -means.

Three kinds of data sets commonly used in similar studies were used for experimental data: The low-dimensional feature data set was the Iris plant sample data

TABLE 1: Parameter Settings of is PO + K-means algorithm.

The data set	$a$	$b$	$p$	$l$	$s$	$\epsilon$
Iris	15	2	40	1	4	$1.00e+00$
Winc	300	2	3	1	4	$1.00e+00$
Cancer	2000	2	40	1	4	$1.00e+00$

TABLE 2: Comparison of stability of three  $k$ -means clustering results.

Data set	Iris		Winc		Cancer	
	Mean	Variance	Mean	Variance	Mean	Variance
$k$ -means	33.1022	9.5668	140.1022	38.5623	701.1236	331.1211
PSO + $k$ -means	12.9673	4.0006	71.1389	5.1363	271.8983	44.9673
IPSP + $k$ -means	11.7652	3.9362	62.4187	3.1364	206.3255	16.2147

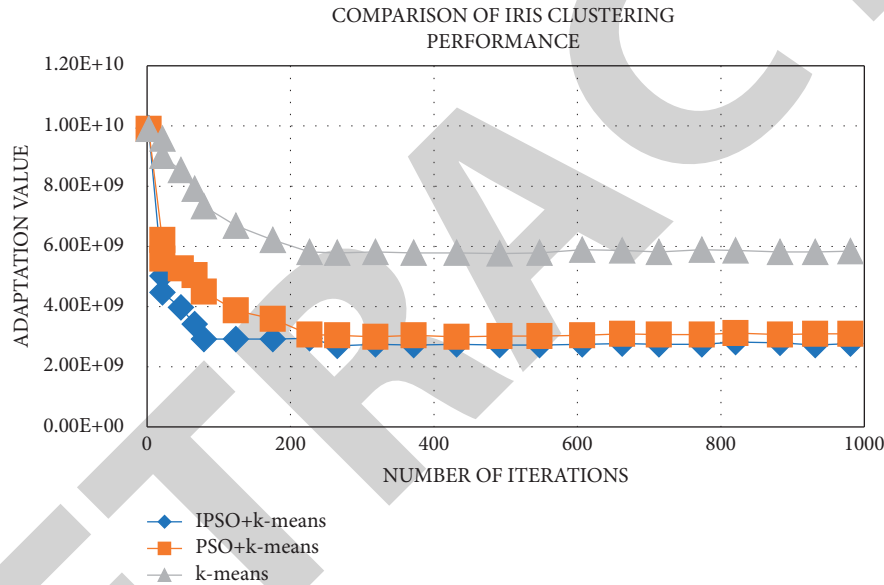


FIGURE 4: Comparison of IRIS clustering performance.

set, which was divided into three categories with four attributes for each data object and a total of 150 samples. The high-dimensional feature data set adopts the Wine data set, which is divided into three categories. Each data object has 13 attributes and 178 samples in total. The WisconsinBreastCancer dataset was used for the multiple high-dimensional data, which was divided into two classes with nine attributes for each data sample, a total of 663 samples.

In the experiment, the value of the adaptive function  $f()$  is the function value corresponding to the current position. Parameter Settings of the  $K$ -means algorithm and PSO +  $K$ -means algorithm are referenced. Table 1 describes the parameters of the ISPO +  $K$ -means algorithm.

**4.6. ISPO +  $K$ -Means Algorithm Experimental Results Analysis.** The clustering experiment results are shown in

Table 2 and Figures 4–6. It can be seen from Figures 4–6 that the  $k$ -means algorithm has the fastest convergence speed, but it is easy to fall into local optimal. By combining PSO +  $K$ -means clustering algorithm with the traditional particle swarm optimization algorithm, the global optimal solution searching ability of the algorithm is improved. However, as the number of iterations increases and the effect of  $k$ -mean decreases, the particle swarm focuses and converges gradually, and then falls into the local optimum. Different from the above two methods, the ISPO +  $K$ -means algorithm proposed in this study does not update the whole velocity vector or position vector at the same time as the traditional PSO at the ISPO stage. Instead, it divides the whole vector into several sub-vectors and updates the sub-vectors at the same time. By introducing a new learning strategy, the particle can make an intelligent analysis of its previous velocity update and determine the velocity of the next iteration, thus increasing the velocity diversity and making it

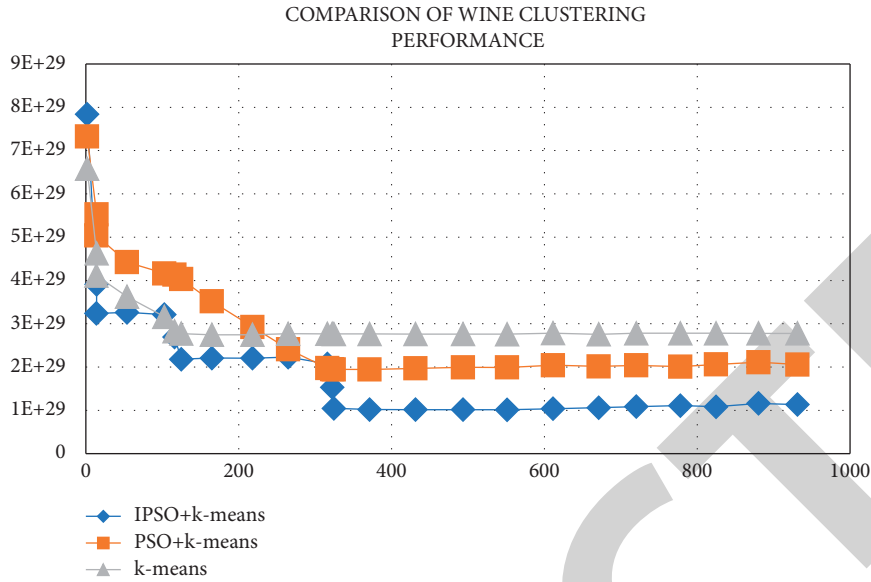


FIGURE 5: Comparison of clustering performance for WINE.

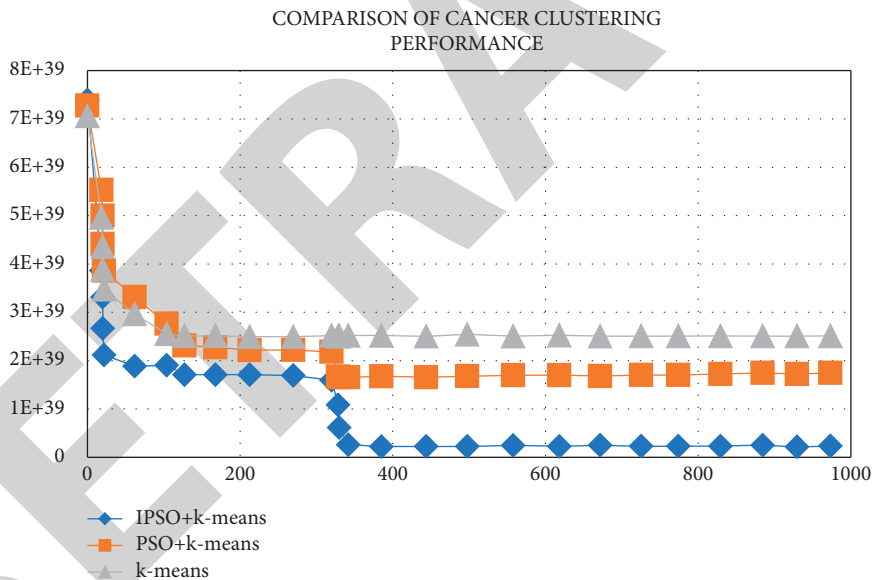


FIGURE 6: Comparison of clustering performance for CANCER.

easy to jump out of the local optimal and approach to the global optimal.

The comparison results of data clustering stability of the three algorithms are shown in Table 2. It can be clearly seen from the sub-variance value and mean difference of the above algorithms that the stability of the k-means algorithm is often extremely poor due to the large influence range of the initial solution, and it is very easy to cause the calculation results to fall into local optimal. The PSO + K-means hybrid clustering algorithm system is a k-means operation introduced at least once in each iteration of the algorithm, which makes the system algorithm in the initial stage of the entire iteration can always have a good and stable processing ability

of distribution search results of global initial solutions. Therefore, the algorithm architecture itself greatly reduces the dependence of the algorithm on the global initial solution distribution search results, and the stability of the overall structure of the algorithm is relatively good. The proposed algorithm ISPO + K-means has no obvious advantage in processing low-dimensional feature data. However, in fact, it is difficult for the traditional PSO algorithm to take into account all the optimization directions of low-dimensional space when dealing with the problem of high-dimensional and diverse particle feature data. The ISPO + K-means algorithm proposed in this study decomposes the large particle search space into several low-dimensional

small Spaces in an ISPO module. And updated throughout the particles in the process of iteration, the introduction of some new learning optimization strategy, through the analysis of the real-time particle velocity and the update of the before, intelligent quickly achieve the real-time dynamic optimization to solve the speed adjustment, enhance to improve the global solutions for computing performance in the field of space exploration, making global algorithm significantly less dynamic dependence on initial solution speed. The dynamic stability of the algorithm is greatly improved. As can be seen from the mean and variance values shown in Table 2, the algorithm proposed in this study has obvious advantages in processing high-dimensional feature data, especially for complex and diverse high-dimensional feature data, and its performance is greatly improved compared with other algorithms.

Compared with the other two algorithms, the ISPO + K-means algorithm has the highest efficiency in terms of time, because K-means does not perform PSO optimization and clustering directly. Compared with PSO + K-means clustering algorithm, ISPO + K-means algorithm does not update all spatial dimensions at the same time, which greatly reduces the iteration time of the algorithm, especially in processing high-dimensional data.

In this chapter, the automatic clustering based on the previously proposed method is studied. Based on a full understanding of the clustering technology, the analysis of the classical clustering algorithm *K*-means algorithm and particle swarm algorithm, the hybrid clustering algorithm combining intelligent single particle and *K*-means algorithm is proposed for the problem that the particle swarm algorithm is prone to fall into the local optimum and the two inherent defects of *K*-means algorithm--ISPO + K-means clustering algorithm, including two modules: ISPO module and *K*-means module. The ISPO algorithm is used to find the optimal *m* initial clustering centroids; the *K*-Means algorithm is used to find the clustering results, and finally, the clustering results are output. The experimental results show that the proposed algorithm overcomes the problems of the above algorithms better, especially for the clustering of high-dimensional data.

## 5. Summary and Prospect

This thesis summarizes the current situation and problems of multimedia English teaching resource library construction, and the main content is to focus on the use of existing information technology as a means to carry out research related to the construction of the resource library. This chapter concludes the thesis with a specific summary of the main work and innovation points of the previous study, and also makes the corresponding construction and outlook for the next work and research direction.

*5.1. Summary.* The research content and main innovations of this study are mainly as follows: in view of the defects and shortcomings of some current teaching resources construction, a model is proposed to collect and automatically

organize the rich teaching resources on the network by using existing information technology. The overall architecture of the model and the design ideas of each module are also described. In addition, this study proposes a hybrid automatic clustering algorithm ISPO + K-means and K-means clustering algorithm and compares it with the traditional particle swarm *K*-means-based hybrid algorithm. The results show that the proposed algorithm can effectively overcome the problem that other algorithms tend to fall into local optimality, and verify that the algorithm has obvious advantages in processing multi-dimensional data sets.

*5.2. Prospect.* In the process of resource construction and management in the teaching field, we need tools for collecting, organizing, and retrieving various subject-specific resources, how to collect, organize and store the resources by using existing IT technologies? This requires us to integrate various technologies to form an automatic structured teaching resource collection (collection) and organization (process) method. This study mainly starts from the perspective of the current situation of teaching resources, and puts forward the corresponding collecting and organizing methods for the shortcomings and deficiencies in the current process of constructing multimedia English teaching resources library, due to the limitation of time, objective conditions, and our own ability, the study only does relevant research on part of the method, and the follow-up work can be carried out from the following aspects.

First, the other parts of the proposed automatic collection and collation method will continue to be studied. This includes data extraction, formation of structured resources, and development of a structured repository engine.

Second, optimize the proposed ISPO + K-means automatic clustering algorithm and try to further successfully improve the execution efficiency of the algorithm as well as the accuracy of the final clustering results.

Third, the model can be continuously improved and enhanced for this model, and a prototype system can be developed to make it more widely used in the field of resource search and organization.

## Data Availability

The labeled dataset used to support the findings of this study are available from the corresponding author upon request.

## Conflicts of Interest

The authors declare no conflicts of interest.

## Acknowledgments

This study was supported by the following projects: A Probe into the Mode Teaching of How to Make the English Study and Moral Education Integrate Based on the Internet in the Ethnic Tibet-Related Areas (No.SCJG21D021) and To Construct the Integration of Ideological, Political Course into College English in Ethnic Tibet-Related Areas (No.WYJZW-2021-2145).

## References

- [1] B. Kang, S. Kang, and X. Ning, "Chinese character learning information platform based on multimedia information technology[J]," *Wireless Communications and Mobile Computing*, vol. 2022, Article ID 1758129, 2022.
- [2] J. Wang, "Multimedia technology and application Course based on application talent training construction and practice of three-dimensional teaching resource base[J]," *Advances in Educational Technology and Psychology*, vol. 5, no. 4, 2021.
- [3] R. Liu, "Design of ideological and political multimedia network teaching resources integration system based on wireless network[J]," *Scientific Programming*, vol. 2021, Article ID 4293771, 2021.
- [4] D. Lv and L. Song, "Research on teaching method of graphic design software[J]," *Frontiers in Educational Research*, vol. 0, no. 5, p. 4, 0, 2021.
- [5] Q. Mamurova Mehrinoz, "How significant are multimedia resources in teaching and learning English language[J]," *Academica: An International Multidisciplinary Research Journal*, vol. 11, no. 5, 2021.
- [6] X. Luo, "Cloud classroom design for English education based on internet of things and data mining[J]," *Mobile Information Systems*, vol. 2021, Article ID 5555006, 2021.
- [7] I. Hilola and A. Dorr, "Smelser at al. The effect of applying the advantages of the infographic format in the process of media literacy education[J]," *Academica: An International Multidisciplinary Research Journal*, vol. 11, no. 4, 2021.
- [8] A. P. F. Vivas, "Didier at al. The teaching of neuroanatomy through the combined usage of traditional resources and multimedia devices [J]," *Revista Cubana de Educación Médica Superior*, vol. 35, no. 1, 2021.
- [9] S. H. Nopita, "The development of innovative learning resources with multimedia to support online learning in teaching industrial management[J]," in *Proceedings of the 6th Annual International Seminar on Transformative Education and Educational Leadership (AISTEEL 2021)*, p. 591, NY City, November 2021.
- [10] X. Qiao, "Integration model for multimedia education resource based on internet of things," *International Journal of Continuing Engineering Education and Life Long Learning*, vol. 31, no. 1, p. 17, 2021.
- [11] A. R. Bolinger and T. C. Burch, "Identifying and evaluating schéin's three layers of culture: the Texas A&M culture exercise," *Management Teaching Review*, vol. 5, no. 4, pp. 351–362, 2020.
- [12] S. D. Puspita, "Implementation of project-based learning resources with multimedia to improve student learning outcomes in teaching cation analysis[J]," in *Proceedings of the The 5th Annual International Seminar on Transformative Education and Educational Leadership (AISTEEL 2020)*, p. 488, Amsterdam, November 2020.
- [13] A. Predoi-Cross, "Teaching introductory undergraduate physics courses using multimedia resources[J]," *AIP Conference Proceedings*, vol. 2109, no. 1, 2019.
- [14] M. Heitzler, H. R. Bär, R. Schenkel, and L. Hurni, "The light source metaphor revisited—bringing an old concept for teaching map projections to the modern web," *ISPRS International Journal of Geo-Information*, vol. 8, no. 4, p. 162, 2019.
- [15] W. O. Anyim, "Multimedia instructional resources for effective library user education programme in universities in North-Central, Nigeria[J]," *Library Philosophy and Practice*, vol. 2019, 2019.
- [16] V. Victoria, "MERLOT (multimedia educational resource for learning and online teaching)[J]," *Charleston Advisor*, vol. 20, no. 2, 2018.
- [17] J. Wang, "Research on the effect and teaching mode of three-dimensional teaching resources construction -- the experience of the curriculum reform of "multimedia technology and application"" in *Proceedings of the 2017 7th International Conference on Education and Management (ICEM 2017)*, Amsterdam, July 2018.
- [18] W. O. Anyim, "Multimedia instructional resources for effective library user education programme in universities in north-central, Nigeria[J]," *Library Philosophy and Practice*, vol. 12, 2018.
- [19] W. Bian, "Construction of multimedia teaching resource management evaluation model in the big data environment," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 13, no. 03, p. 4, 2018.
- [20] B. Nataliia, "Kopniak. The use of interactive multimedia worksheets at higher education INSTITUTIONS[J]," *Informacijni Tehnologii i Zasobi Navčannâ*, vol. 63, no. 1, 2018.
- [21] Li. Pan and C. Li, "Construction of multimedia teaching platform for community nursing based on teaching resource library technology[J]," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 12, no. 07, 2017.
- [22] B. AA and V. V. Konnov, "Bizyaeva at al. Multimedia presentation as a form of E-learning resources in the educational process[J]," *Saratovskij Naučno-medicinskij Žurnal*, vol. 13, no. 2, 2017.
- [23] P. Kumar and X. Fan, "Developing multimedia resources using camtasia studio 8 to teach pathology in the era of subspecialized residency education[J]," *Modern Pathology*, vol. 30, 2017.
- [24] L. Zhu, "An analysis of English network teaching mode in higher vocational colleges in the internet era[J]," *Proceedings of the 2016 4th International Conference on Management Science, Education Technology, Arts, Social Science and Economics (msetasse-16)*, vol. 85, 2016.
- [25] Da. Liu, J. Shu, and X. Fan, "Research of the English network teaching resources system based on data warehouse technology[J]," *2016 4th International Conference on Machinery, Materials and Computing Technology*, vol. 1, 2016.
- [26] R. Matheson and I. Mathieson, "Ceramics studio to podiatry clinic: the impact of multimedia resources in the teaching of practical skills across diverse disciplines," *Innovations in Education & Teaching International*, vol. 53, no. 3, pp. 349–359, 2016.
- [27] A. Mohamed and A. Yasser, "Collection of web multimedia teaching resources using ontology[J]," *International Journal of Information and Education Technology*, vol. 6, no. 4, 2016.
- [28] X. Xu, "Research on the application of multimedia computer interactive teaching in higher vocational colleges education [J]," *RISTI: Revista Ibérica de Sistemas e Tecnologias de Informação*, vol. 2016, no. E10, 2016.
- [29] B. Q. Pan, "Study on multimedia art teaching resources management system based on internet technology[J]," *MATEC Web of Conferences*, vol. 44, 2016.
- [30] A. Chantyclaire, "Tiba, Janet at al. Digital storytelling as a tool for teaching: perceptions of pre-service teachers[J]," *The Journal for Transdisciplinary Research in Southern Africa*, vol. 11, no. 3, 2015.