International Journal for Housing Science and Its Applications
Publish August 6, 2025. Volume 46, Issue 3 Pages 5223-5230

https://doi.org/10.70517/ijhsa463449

Optimization Strategies for Reform Path of Civicizing Yoga Curriculum in Colleges and Universities Based on Data Mining

Kun Jiang[1,*], Congcong Ma[1] and Yong Li[1]
[1] Sports Department, Shaanxi Fashion Engineering University, Xianyang, Shaanxi, 712046, China
Corresponding authors: (e-mail: jiangkun1268@163.com).

Abstract For the association problems in the path of reforming the Civic Politics of yoga courses in colleges and universities, this paper takes the relevant elements between yoga and Civic Politics as the entry point. The text analysis method is used to mine the core elements of yoga courses and Civic and political courses. Aiming at the limitation of TF-IDF model in the accuracy of keyword extraction, the G1 assignment method is used to integrate the multidimensional features of keywords. Based on the different attributes of the words, the corresponding weights are assigned to rank the importance of the words in a more scientific way, and the comprehensive weights of the words are calculated accordingly. As a result, the keyword extraction algorithm SLPL-TF-IDF is proposed based on multiple features of words and semantic information extraction based on the Longformer pre-trained language model, and then the improved Apriori algorithm is introduced to mine and output the association rules between yoga courses and the knowledge of Civics and Politics, which provides effective technical support for the construction and optimization of the path of Civics and Politics reform of yoga courses in colleges and universities. In the experimental reform of yoga courses' ideology and politics in K colleges and universities, it is found that the subject word “yoga philosophy” has the highest correlation with other high-frequency subject words, which are all up to 0.90 and above. The theme of “Yoga Philosophy” should be used as an important bridge for the construction and optimization of the path of political reform of yoga courses in colleges and universities.

Index Terms text analysis, SLPL-TF-IDF, Yoga curriculum, political reform, Apriori algorithm, G1 empowerment method

I. Introduction

Yoga, as an ancient and comprehensive system of practice, integrates elements from many fields such as philosophy, psychology, physiology, and sociology [1]. It is not only regarded as a traditional way of exercising the body and mind, but also as a comprehensive discipline that can help people better understand themselves and society [2], [3]. The new era emphasizes that the socialist thought with Chinese characteristics leads the innovation of college curriculum ideology and politics, and it is necessary to realize the innovation of curriculum content and teaching methods, and integrate the great achievements of the new era into the curriculum content. And the yoga course in physical education can cultivate the spirit of patriotism and socialist core values of students by learning the eight precepts, eight systems and other knowledge, so that students consciously abide by the social legal system, cultivate traditional Chinese virtues and other civic will [4]-[6]. Therefore, colleges and universities should grasp the flow of ideology and politics in the curriculum, mobilize all means, actively seek the ideological and political elements of yoga classes, strengthen the effect of the curriculum, optimize the learning process of college students, and stimulate the growth of contemporary college students [7]-[9]. In the context of the construction of new liberal arts, in order to accelerate the modernization of education and the construction of a strong educational country, it is possible to promote the reorganization of traditional liberal arts in colleges and universities in order to construct new liberal arts courses, achieve a breakthrough in professional barriers and disciplinary limitations, and promote multidisciplinary interaction [10], [11]. By integrating modern new information technology into the professional learning and training of liberal arts, we can change the talent cultivation mode, provide students with a comprehensive interdisciplinary learning platform, and effectively cultivate students' ability to integrate cross-disciplinary knowledge and practical ability [12], [13].

This paper firstly elaborates the extraction process of text analyzing method. On this theoretical basis, it analyzes the architectural composition of the TF-IDF model, and discusses the computational methods as well as the ideas of TF and IDF respectively. Secondly, SLPL-TF-IDF keyword extraction algorithm is used to mine and extract the topics and keywords of yoga course contemplation and politicization, and the sample size of each topic and keyword is counted. After obtaining the keywords and topics of yoga course politicization, the principles and

steps of the improved Apriori algorithm to output the association rules of yoga course and politicization knowledge are explained, so as to form the optimization strategy and method of the reform path of yoga course politicization in colleges and universities. Again, the keywords and sample size of yoga course and civic and political knowledge are extracted and counted. At the same time, the degree centrality, near centrality, and middle centrality of the high-frequency word keywords are calculated to determine the hot topics of the reform of the yoga curriculum's civic and political reform in colleges and universities. Finally, K colleges and universities are used as experimental objects to conduct experiments on the reform of Civics and Politics of Civics in Yoga Courses. The improved Apriori algorithm is utilized to convert the average interest value of students on hot topics, mine the association rules, and calculate the correlation degree between hot topics.

II. Data Mining for the Civicization of Yoga Courses

II. A. Text analysis methods

Natural language processing technology is used to process the depth of the course's ideological and political materials. Taking the machine learning course as an example, text materials such as textbooks, lesson plans, classroom transcripts, etc. of the course are first collected. Then, the text is transformed into a computer-processable form through preprocessing steps such as word splitting and lexical labeling.

Text analysis is used to extract keywords related to the elements of Civics and Politics from the machine learning course, such as business ethics, social responsibility, and ethics. These keywords not only reflect the core content of the course, but also reveal potential Civic and Political elements. In addition, thematic modeling of the course text is performed to identify themes such as professional ethics and business ethics, which are closely related to the elements of Civics and provide directions for the design of the course's Civics.

II. B. TF-IDF model

TF-IDF is composed of two parts, word frequency (TF) and inverse document frequency (IDF), by which it can be used to evaluate the importance of a feature word in a document in a document library. The word frequency is calculated as in equation (1):

$$TF_{i,j} = \frac{n_{i,j}}{\sum_{k=1}^{k} n_{k,j}} \tag{1}$$

$TF_{i,j}$ represents the frequency of occurrence of the word w_i in the document d_j, $n_{i,j}$ is the number of occurrences of w_i in the document d_j, and the denominator is a summation of the frequency of occurrence of all the feature words in the document d_j, and k is the number of different words in the document d_j. The other part of the model, IDF, is computed as in equation (2):

$$IDF_i = \log\left(\frac{n_d}{df(d, w_i)+1}\right) \tag{2}$$

where IDF_i represents the inverse document frequency of the feature word w_i in the document d_j, n_d is the number of all documents in the text library, and $df(d, w_i)$ is the number of documents in the document library that contain the feature word w_i, with 1 added in order to prevent the number of occurrences to be zero.

The idea of TF indicates that if a certain feature word occurs more times in a certain document, then this feature word can be used as the keyword of the document, and its correlation with the text containing this keyword is very large, and it can be used as the feature vocabulary for classification.The idea of IDF indicates that if only a small amount of text contains a certain feature word, then this keyword can be used as a categorization feature word, which is more conducive to distinguish different categories of text. Generally, TF and IDF are used together to calculate the TF-IDF value, so as to achieve the purpose of text vectorization.

II. C. TI-IDF Keyword Extraction Algorithm Fusing Multiple Features

The core of SLPL-TF-IDF keyword extraction algorithm lies in the following: generating the set of importance weights of semantic similarity, word position, lexicality, and word length through the G1 assignment method, and optimizing the calculation results of the weight values in the TF-IDF algorithm, so as to realize the precise extraction of keywords. The algorithm divides words according to different attributes, making keyword extraction more targeted and precise. Compared with the original algorithm, this improved method significantly improves the

precision and recall of keyword extraction, thus optimizing the effect of text processing.The SLPL-TF-IDF calculation formula is shown in equation (3).

$$S(v_i) = (1-d) \times TF - IDF + d \times W(V_i) \tag{3}$$

where d denotes the weight relationship between TF-IDF and G1 assignment method in the comprehensive weight calculation, which is obtained through experimental iterations.The overall flow of SLPL-TF-IDF algorithm implementation is shown in Fig. 1.

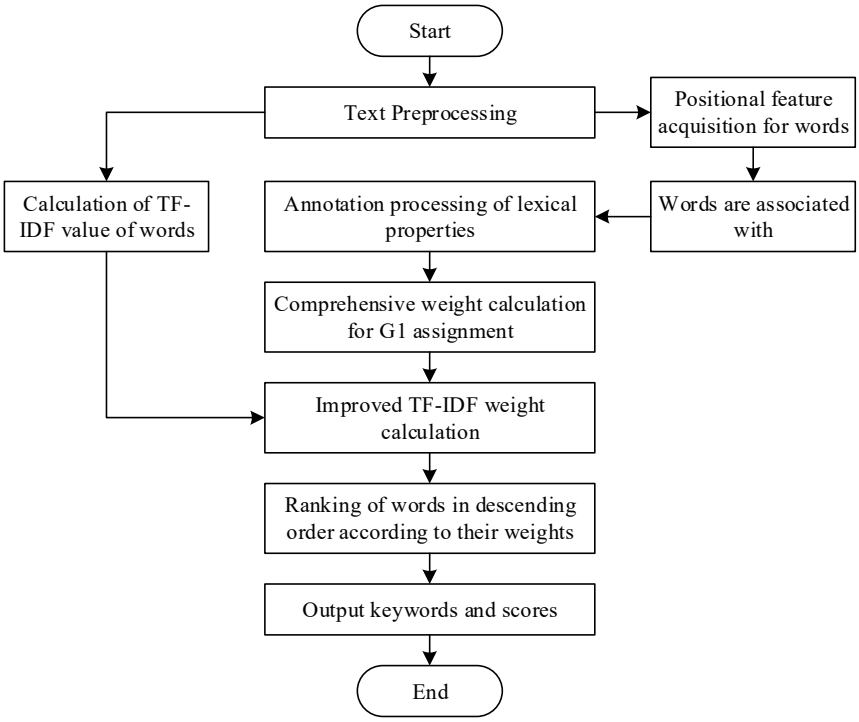

Figure 1: The SLPL-TF-IDF algorithm process

III. Association analysis based on the Apriori algorithm

III. A. Analysis of the correlation between course content and Civics themes

After constructing the library of Civics elements, it is necessary to further study the connection between course content and Civics topics. The goal of this link is to discover the potential connection between the knowledge points of the course and the elements of Civics, so as to provide a scientific basis for the design of Civics in the course.

III. B. Improved Apriori Algorithm

In this paper, an improved Apriori algorithm is proposed:

(1) For the problem of frequently scanning the whole dataset, this paper starts from reducing the approximate capacity of the scanned dataset. First of all, starting from $k=1$, each time after searching for $k-$ frequent item set, the transaction items in the data set that do not contain $k-$ frequent item set and its subsets are all eliminated from the original data set D, generating a new small data set Di, and searching for $k+1-$ frequent item set will be scanned in the new small data set Di, and so on each time to narrow the scanned data set about capacity for iteration, thus reducing the overhead caused by frequent scanning of data sets and improving efficiency.

(2) For the equality rule of items in Apriori algorithm, since this paper wants to mine the association rules between yoga courses and Civic and political knowledge, this paper will weight the category field inside yoga courses to ensure the results of the association rules related to the priority yoga courses and Civic and political knowledge field. The data are processed as follows: firstly, the yoga course field is subjected to data cleaning, word splitting, removal of deactivated words and other operations, secondly, the collection of processing results is used

as an item set input item, and finally, the yoga course is added as an item to jointly constitute a Civic Politics of a transactional item set, and the weight of the weighting of the skill requirement field is uniformly set to 1.1.

The modified formula for calculating the degree of support is shown in Equation (4) and Equation (5):

When $A \in G$ or $B \in G$:

$$New-Sup(A \to B) = \frac{\omega(A,B)Q(A,B)}{Q(R)} \tag{4}$$

When $A \notin G$ and $B \notin G$:

$$New-Sup(A \to B) = \frac{Q(A,B)}{Q(R)} \tag{5}$$

The modified confidence level is calculated in the following equations (6) and (7):

When $A \in G$ or $B \in G$:

$$C(A \to B) = \frac{\omega(A,B)Sup(A \cap B)}{\omega(A)Sup(A)} \tag{6}$$

When $A \notin G$ and $B \notin G$:

$$C(A \to B) = \frac{Sup(A \cap B)}{Sup(A)} \tag{7}$$

$\omega(A)$ denotes the weighting of the itemset A, $\omega(A,B)$ denotes the one with the highest weighting of $\omega(A)$ and $\omega(B)$, $Q(A,B)$ denotes the transaction set that contains both itemset A and itemset B, and G denotes the yoga course field. The classic algorithm for mining association rules between data is Apriori algorithm, in this paper, based on the traditional Apriori algorithm, we modify the capacity of the data set about each scan, and change the traditional algorithm to scan all the data set about each scan to scan the frequent itemset and its subsets each time, and at the same time, we increase the weighting rules, and use the modified support and confidence to filter the association rules, the detailed steps are as follows:

(1) Set the values of min support and confidence, which will be used to produce the subsequent frequent itemsets as well as the associated rules.

(2) Scan the complete dataset D. The first scan will produce a candidate 1-item set, denoted as $H1$, which is based on the weighted rule formula to compute the new support and compare it with the min support set in the first step, thus producing a frequent 1-item set, denoted as $P1$.

(3) Take the itemsets containing $P1$ and its subsets together to form a new dataset $D1$. From the frequent 1-item set $P1$, a candidate 2-item set $H2$ is generated directly, and the new support is calculated according to the weighting rule formula and compared with the min support set in the first step to generate the frequent 2-item set $P2$.

(4) Continuously repeat step 3 to generate candidate $k-$itemsets Hk and frequent $k-$itemsets Pk until no new candidate $k+1-$itemsets $Hk+1$ and frequent $k+1-$itemsets $Pk+1$ can be generated, and at this time, all the frequent itemsets have been obtained from the whole dataset.

(5) According to the values of min support and confidence set in the first step, the final association rule results are obtained.

IV. Optimization of the path of reform of the yoga curriculum in higher education for the purpose of political thinking

IV. A. Keywords and Themes Mining for Civicizing Yoga Curriculum

IV. A. 1) A clear categorization of research topics in curriculum philosophy and politics

The algorithm proposed above was used to obtain 3789 subject words about yoga programs. According to the number of occurrences of each subject term, 35 subject terms with the highest number of occurrences were selected for statistical analysis, the average TF - IDF value of each subject term was calculated and the research type of each subject term was labeled, and the subject terms were labeled as the object of study (O), the content of the study (T), and the method of the study (M), respectively. Five theme words with the same meaning or

ambiguous research type were excluded when labeling the theme words, and the remaining 30 theme words were statistically analyzed. The results of theme word analysis are shown in Table 1.

Table 1: Analysis results of subject terms

Number	Subject term	Frequency	Type	TF-IDF
1	Undergraduate	10160	T	0.43
2	Education	9007	T	0.39
3	Thought	7660	T	0.4
4	Ideological and political education	5119	O	0.44
5	Colleges and universities	2733	T	0.75
6	Yoga	1923	T	0.62
7	Ideological	1469	T	0.64
8	Course	1309	T	0.73
9	Ideological and political courses	1244	T	0.45
10	Yoga course	1203	M	0.42
11	Yoga teaching	1138	T	0.46
12	Posture	1104	M	0.37
13	Yoga equipment	1095	T	0.46
14	Equipment	1078	T	0.33
15	Breathe	1075	M	0.52
16	Respiratory control	902	T	0.38
17	Emotion	796	T	0.37
18	Transformation of education	776	M	0.65
19	Yoga teacher infiltration of culture	722	T	0.49
20	Theory course	571	T	0.64
21	Yoga Philosophy	496	O	0.45
22	Meditation	458	M	0.61
23	Practice	450	T	0.57
24	Ideological and political theory	446	T	0.46
25	Ideological and political work	437	M	0.39
26	The history of yoga	433	M	0.67
27	Value	419	T	0.35
28	Socialism	10160	T	0.43
29	Reform	9007	T	0.39
30	Pose techniques	7660	T	0.4

IV. A. 2) Sample Size of Civicizing Elements in Yoga Courses

In this part of the work the main task is to complete the text categorization task using EasyDL platform. According to the computer-based curriculum Civics framework to design the type of Civics course, when collecting materials, it is necessary to screen them according to certain criteria, such as the content should be appropriate to the type of Civics and the source should be reliable, and so on. The distribution of the number of various samples is shown in Figure 2, and the statistics show that the number of samples of the Yoga Civics element example is 1990.

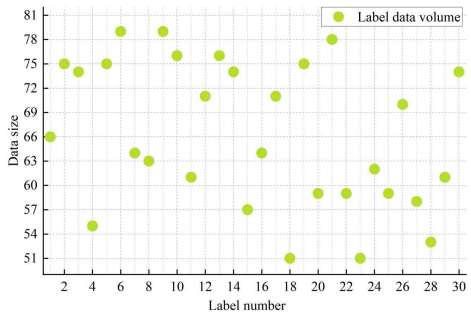

Figure 2: The sample size of yoga ideological and political examples

IV. A. 3) Initial establishment of hot topics

In social network analysis, the "power" of a node in the social network is usually explored by studying the "centrality" of the node. Commonly used indices include degree centrality (D), proximity centrality (C), and intermediate centrality (B). Degree centrality measures the ability of a node to establish a direct association with other nodes. Proximity centrality determines a node's ability to transmit information by describing its "distance" from other nodes. Intermediate centrality depicts the ability of a node to act as an "intermediary" between other "point pairs". The degree centrality in this article reflects whether a keyword appears with other words in a document. Proximity to centrality measures the probability that a keyword in a network will co-appear with other keywords. Intermediate centrality is an indicator that measures the ability of a keyword in the network to influence other keywords to appear together in a journal paper.

Calculate the centrality values of 30 keywords. Among them, the average values of degree centrality, proximity centrality, and intermediate centrality are 33.232, 60.418, and 1.302, respectively. There are 10 high-frequency keywords whose values of the three kinds of centrality are higher than the average values are shown in Table 2.

Table 2: Measurement results of the centrality of high-frequency keywords

High-frequency key words	D	C	B
6	83.641	85.943	9.562
9	74.55	79.715	6.764
21	58.187	70.518	3.47
10	63.641	73.338	3.437
13	63.641	73.338	5.181
2	60.005	71.434	3.236
5	60.005	71.434	4.693
19	50.914	67.078	2.454
4	56.369	69.625	2.836
15	45.46	64.711	2.216

Among the high-frequency keywords in Table 2, the three node centrality of the keywords "yoga", "ideological and political course" and "yoga philosophy" are much higher than those of other keywords, and they are in the absolute core position of the network. Among them, the degree centrality and near-centrality of the keyword "yoga" exceeded 80.00, which were 83.641 and 85.943, respectively. The median centrality exceeded 9.000 and was 9.562. Although the centrality of the three nodes of the keywords "yoga course" and "thought" is not as good as the first three keywords, it also ranks in the top 1/2 and is in the sub-core position. Among them, the degree centrality and proximity centrality of "yoga course" are high, but the middle centrality is slightly lower, indicating that the keyword has a large "power" in the network and often appears in a paper with other keywords, but has insufficient influence on the co-occurrence of other keywords. Therefore, it can be preliminarily determined that "yoga", "ideological and political courses", "yoga philosophy", "yoga courses" and "thoughts" refer to the hot topics of ideological and political reform of yoga courses in colleges and universities.

IV. B. Reform Path of Yoga Curriculum Civicization

IV. B. 1) Data conversion

In the Apriori algorithm analysis, Boolean data can simplify the data set, that is, let each data item only expressed as 1 or 0, can more intuitively indicate the existence or otherwise of the data item, can also reduce the complexity of the calculation, so as to improve the algorithm's operation speed, and more efficiently discover the frequent item set. Therefore, after obtaining the above pre-processed raw data, in order to facilitate the subsequent processing and analysis, it is also necessary to carry out data conversion to convert the data object into Boolean data.

Taking K college as the experimental object, the experiment of reforming the yoga course of thinking and politics is launched. The average interest value of 30 students in the five hot topics above is randomly selected as the feature identification, and if the interest value of students in a hot topic is greater than or equal to the average interest value of the subject, the hot topic is considered to be a key topic, and is recorded as 1. Otherwise, it is considered to be a non-focused topic, and is recorded as 0. At the same time, in order to make the data table easy to view, the student number is expressed using the numbers ranging from 01 to 30, and the 5 hot topics are expressed by the codes from C1 to C5, and some of the final data obtained after data conversion are shown in Table 3.

Table 3: Partial data conversion results

Student number	Hot topics				
	C1	C2	C3	C4	C5
01	1	0	1	0	1
02	0	0	0	1	1
03	1	1	1	0	0
04	0	1	0	1	1
…	…	…	…	…	…

IV. B. 2) Data mining

After the analysis of the improved Apriori algorithm, a total of 10 association rules for generating support above 0.5 and confidence above 0.93 for the ideology and politicization of yoga courses in colleges and universities are shown in Table 4, and the frequency of appearance of each hot topic in the association postitems is the same, but the frequency of appearance of C3 (Yoga Philosophy) in the association postitems is the highest at 0.7.

Table 4: Degree of association

Number	Associated consequent	Relating to the preceding item	Support degree	Degree of confidence	Upgrade degrees
1	C1	C5	0.61833	1.000	1.601
2	C1	C5, C3	0.53833	1.000	1.601
3	C2	C3	0.57833	1.000	1.601
4	C2	C3, C1	0.63167	0.98372	1.601
5	C3	C1, C2, C4	0.59167	0.98227	1.598
6	C3	C3	0.55167	0.98061	1.596
7	C4	C1, C3, C5	0.51167	0.97868	1.593
8	C4	C3, C2	0.51167	0.97868	1.593
9	C5	C1, C5	0.55167	0.93183	1.484
10	C5	C2, C4, C3	0.63167	0.94117	1.409

IV. B. 3) Visualization of mining results

This section further visualizes the correlations within the five hot topics mined in Figure 3, where C3 (Yoga Philosophy) has strong correlations with other topics with correlations of 0.9 and above.

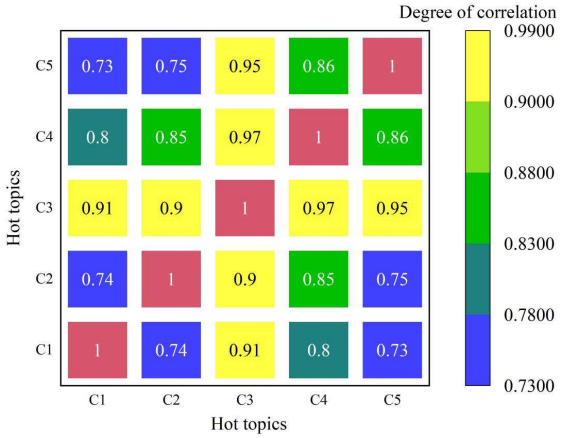

Figure 3: The degree of correlation among hot topics

V. Conclusion

The main research conclusions of this paper are as follows:

(1) Aiming at the problem of low accuracy of keyword extraction caused by the TF-IDF model algorithm due to the shortcomings of restricted corpus, low degree of text utilization, and inability to capture the deep semantic information of the text, we construct the set of importance weights based on the G1 assignment method. By adding multi-dimensional feature attributes of words on the basis of TF-IDF model algorithm, optimizing the calculation

results of TF-IDF algorithm on keyword weight value, so as to realize high-precision keyword extraction and mining, the design of TI-IDF keyword extraction algorithm integrating multiple features is completed.

(2) The improved Apriori algorithm, which has the advantages of shorter running time, less memory occupation and higher efficiency, is selected to mine and output the association rules for the politicization of yoga courses in colleges and universities under the guidance of the theory of association analysis between the course content and the topic of politicization.

(3) Using the proposed methods to excavate and sort out the key words of the ideological and political education of yoga courses, five hot topics were preliminarily obtained: "yoga", "ideological and political courses", "yoga philosophy", "yoga courses" and "thoughts". Taking K university as the experimental object, the performance data of five hot topics in the ideological and political experiment of yoga courses were further transformed and mined, and a total of 10 rules related to the ideological and political education of yoga courses in colleges and universities with a support degree of more than 0.5 and a confidence degree of more than 0.93 were obtained. Through the visualization of the correlation degree of the five hot topics, it is found that the correlation degree of "Yoga Philosophy" with the other five hot topics is 0.90 or above, showing a strong correlation. The study points out that the ideological and political transformation of yoga courses in colleges and universities should take the theme of "yoga philosophy" as the starting point, and transition from "yoga philosophy" to ideology and politics with values as the theme, so as to effectively optimize the reform path.

Funding

This research was supported by the Education and Teaching Reform Research Project (Curriculum Ideological and political special project) of Shaanxi Fashion Engineering University (project No. 2024SZZX025).

References

[1] Weng, T. (2024). A Developmental Study of Yoga Courses and Yoga Teaching in General Colleges and Universities. Frontiers in Educational Research, 7(5).

[2] Berent, G. R., Zeck, J. M., Leischner, J. A., & Berent, E. A. (2014). Yoga as an alternative intervention for promoting a healthy lifestyle among college students. Journal of addictions nursing, 25(4), 167-171.

[3] Gaskins, R., Jennings, E., Thind, H., Becker, B., & Bock, B. (2014). Acute and cumulative effects of vinyasa yoga on affect and stress among college students participating in an eight-week yoga program: A pilot study. International journal of yoga therapy, 24(1), 63-70.

[4] Hu, M., & Wang, Z. (2023). The Ideological and Political Teaching of School Physical Education Courses under the Background of Big Ideological and Political. International Journal of Educational Curriculum Management and Research.

[5] Liu, W., & He, C. (2022). Curriculum-Based Ideological and Political Education: Research Focuses and Evolution. International Education Studies, 15(5), 28-35.

[6] Jiang, K., Ma, C., & Li, Y. (2025). Research on Teaching Reform Supported by Multi-Objective Optimization Algorithm for the Process of Civicization of Yoga Courses in Universities. J. COMBIN. MATH. COMBIN. COMPUT, 127, 5763-5783.

[7] Wang, C., & Gupta, P. (2023). The Path of Ideological and Political Construction in University Physical Education Curriculum Based on the Concept of Collaborative Education. Frontiers in Sport Research, 5(4).

[8] Guan, Y. (2023). Research on the Collaborative Path of Integrating Ideological and Political Content into College Physical Education Teaching. Curriculum and Teaching Methodology, 6(10), 130-134.

[9] Sun, Q. (2021, May). Organic Combination of Ideological and Political Education and Public Physical Education in Universities. In 2021 2nd International Conference on Computers, Information Processing and Advanced Education (pp. 579-582).

[10] Wang, L., & Liu, J. (2023). Reflection on Physical Education Teaching Practice in Colleges and Universities from the Perspective of Curriculum Ideology and Politics. International Journal of New Developments in Education, 5(5).

[11] Chen, Q. (2025). The Profound Impact of Informatization Reform of College Physical Education Courses on Network Ideological and Political Education Based on Data Analysis. J. COMBIN. MATH. COMBIN. COMPUT, 127, 7517-7530.

[12] Høydal, Ø. S., & Haldar, M. (2022). A tale of the digital future: analyzing the digitalization of the norwegian education system. Critical Policy Studies, 16(4), 460-477.

[13] Maltsev, D., Lazareva, A., Rudenko, O., Dudyk, R., & Zvarychuk, Z. (2022). Philosophy of education and science in the context of digitalization of society. Wisdom, (3S (4)), 75-82.