[1] Tianyu Gao

Realistic Problems and Practical Paths of "Three-Wide Education" in Higher Education Based on Text Analysis and Mining

Journal of Electrical Systems

Abstract: - "Three complete education" is a powerful attempt to meet the requirements of the times. Students must have high quality, and they are compound talents who meet the requirements of the times "Universal education" is to provide students with comprehensive training to improve their cultural level and spiritual and moral level to meet the requirements of future social competition. To study the comprehensive reform and development of the moral education in universities, this paper reviews and standardizes the formation and development of the comprehensive reform in universities, which is related to the development of the comprehensive reform from three aspects: strengthening the system interconnection in the process of all staff training, strengthening the effective interaction between cultures in the whole process, and ensuring the complete organic integration of cultures. It puts forward a solution to the comprehensive reform of "three complete education" for universities. Provide some reasonable suggestions for the intensive development of moral education in universities.

Keywords: Higher education, Triple education, Text mining.

I. INTRODUCTION

The impact of multiple ideologies in the international society on the moral education of universities needs to be addressed urgently[1-3]. The continuous development of China's economic globalization, cultural pluralism, network informatization and big data application has made college students suffer from the impact of various western cultural trends in the context of globalization, and in the stormy international environment, if we want to guide college students to continue to carry forward positive ideology and discard negative and backward corrupted ideas, we must establish a full, full and all-round work mechanism of moral education [4].

The state places high hopes on students. The younger generation is the future of the motherland, the central force of society, and an important force for the future progress of the country. The vast number of young people have proved by their actions that in the new century, Chinese youth are excellent and should shoulder heavy responsibilities, and extend my cordial greetings to you and to the vast number of young people fighting on all fronts of the anti epidemic war[5]. If the young generation has dreams, skills and responsibilities, the country has a future, and the people have a future. She is to make the party and the state attach importance to young students in real life, and all teachers should attach importance to youth work and work together Good moral education [6].

As the basic concept of higher education, "three comprehensives" is a reliable guarantee to strengthen and improve the moral work of higher education[7]. With the rapid dissemination of information on the Internet, some selfish hedonism thoughts have penetrated into students' consciousness, changed their values and behaviors, made moral education difficult, and made the traditional moral education model difficult to adapt to the changing environment [8]. The "Trinity" education mode determines the position of teachers in the process of education, enables them to actively participate.

II. RELATED WORK

The three whole education includes, the whole staff education. The "four-in-one" nurturing mechanism is composed of college staff, family members, social forces and students themselves [11]. First of all, as the front position of teaching and educating students, teachers and staff should cooperate with each other to implement teaching, management and service education [12]. Secondly, families have an irreplaceable role in cultivating and shaping students' correct three views and developing good habits and character [13]. Finally, all sectors of society, including the government, enterprises and institutions, and outstanding individuals, also need to take responsibility for the cultivation of students [14] [15]. All-round education. All-round education means that the education work should

[1] Quality Control Office, Tourism College of Zhejiang, Hangzhou 311231, Zhejiang, China;

*Corresponding author: 99494934@qq.com;

stand at a global and strategic level and be expanded and extended continuously [16-17]. In addition, nurturing work should penetrate into all aspects of college students' life, study and work, and cultivate students from multiple perspectives and levels, so as to finally make them become talents, thus realizing the goals pursued by higher education [18-19].

There are many researches on the three holistic education. In [20], it is pointed out that under the concept of "adult-achieving" education, the "Three Associations and Two Guides" education mode and "Adult-achieving Integration" education mode should be explored. In the article, it was pointed out that under the concept of "adult-achieving" education, the " Three-Whole Education" system was constructed on the basis of the "integrated adult-achieving" education system. In [21],they propose "Building a "three-wide education" system with the foundation of moral education", it is proposed that the construction of the three-wide education model "first of all, we should implement the spirit of the National Education Conference", take the foundation of moral education as the foundation, integrate moral theory courses with general education courses through the means of information technology.

III. METHODOLOGY

Text analysis and mining is a process of obtaining information from text data, which is usually not structured. Researchers obtain information according to needs or values. It is generally believed that text mining is a process of extracting unknown and available knowledge from a large number of texts, which can interpret information and provide help and reference for research.

The principle of text mining is similar to that of digital mining, but it is different from traditional data mining in some aspects. The object of data mining is mainly composed of structured data in the database, and some data mining methods are not applicable in text mining. The characteristic of text mining is that the text it studies is semi-structured or unstructured, and there is no certain form, so it is difficult for machine tools to understand the content contained therein. In short, we need to extract features from text, analyze meaningful information, and establish useful models. Chinese text search includes the following steps, as shown in Figure 1.

Figure 1 Flow chart of text analysis and mining

This study adopts a standardized qualitative research methodology to develop qualitative coding in the context of China's education informatization 2.0 era, conducts cluster analysis and multidimensional scale analysis based on the research method of text co-word analysis, and follows the logic of thinking from theory to practice and from the particular to the general.

How to represent text is an important part of the task. Computer encoding of natural language solves problems such as transferring and storing natural language, but the encoding itself does not contain semantic information. If the natural language encoding is used directly for processing tasks, it will result in a lack of semantic information. This is where a language model is needed to transform the text into a specific input containing semantic feature information. One of the common language models is the statistical language model, which is essentially a probabilistic model for computing sentence probabilities, and the construction of such a model also needs to rely on a corpus. Suppose $W(w_1, w_2, ..., w_T)$ denotes a corpus of T words $w_1, w_2, ..., w_T$ a sentence text composed in order. Then $w_1, w_2, ..., w_T$ have the joint probability of

$$p(W) = p(w_1^T) = p(w_1, w_2, ..., w_T) \quad (1)$$

This is the probability of the text W. Using the Bayesian formula, the above equation can be decomposed as

$$p\left(w_1^T\right)=p\left(w_1\right)\cdot p\left(w_2|\ w_1\right)\cdot p\left(w_3|\ w_1^2\right)\cdots p\left(w_3|\ w_1^{T-1}\right) \quad (2)$$

The implementation idea of the above measurement method is simple, but there are two such problems in practical application.

(1) The number of model parameters tends to increase exponentially with the growth of the text length. For example, Equation (2) calculates the probability of a text of length T, which requires T parameters. Now suppose there is a corpus corresponding to a dictionary D,and consider any text of length T. Theoretically, there are N^T possibilities, each of which requires the calculation of T parameters. In total, TN^T parameters need to be computed without considering duplicate parameters. This results in huge computational and storage overheads when dealing with larger text sizes.

(2) The data sparsity is more serious. This will lead to a probability result of 0 for most of the sentences we finally compute. The N-gram model is a typical statistical language model, currently more commonly used are the binary Bi-gram model and the ternary Tri-grammodel, which provide feasible methods to calculate the model parameters and solve both problems. First, considering the approximate calculation of $p\left(w_k|\ w_1^{k-1}\right)(k>1)$, using the Bayesian formula yields.

$$p\left(w_k|\ w_1^{k-1}\right)=\frac{p\left(w_1^k\right)}{p\left(w_1^{k-1}\right)} \quad (3)$$

According to the large number theorem, when the corpus size is large enough, $p\left(w_k|\ w_1^{k-1}\right)$ can be approximated as

$$p\left(w_k|\ w_1^{k-1}\right)\approx\frac{c\left(w_1^k\right)}{c\left(w_1^{k-1}\right)} \quad (4)$$

Where $c\left(w_1^k\right), c\left(w_1^{k-1}\right)$ indicate the number of occurrences of word strings w_1^k, w_1^{k-1} in the corpus, respectively. When k is large, the time overhead of counting the number of these two strings is also large.

$$p\left(w_k|\ w_1^{k-1}\right)\approx p\left(w_k|\ w_{k-n+1}^{k-1}\right)\approx\frac{c\left(w_{k-n+1}^k\right)}{c\left(w_{k-n+1}^{k-1}\right)} \quad (5)$$

When $n=2$, equation (5) becomes

$$p\left(w_k|\ w_1^{k-1}\right)\approx p\left(w_k|\ w_{k-1}\right)\approx\frac{c\left(w_{k-1}^k\right)}{c\left(w_{k-1}\right)}=\frac{c\left(w_{k-1},w_k\right)}{c\left(w_{k-1}\right)} \quad (6)$$

Such simplification makes the matched word strings shorter for statistics and shortens the statistical time. It also makes the total number of model parameters smaller, which solves the first problem mentioned above. The second problem is solved by smoothing the data.

The Neural Network Language Model (NNLM), which solves the problems of traditional statistical language models, but has difficulties in obtaining context-dependent long time and lacks sufficient generalization ability. The model captures contextual information and obtains similarities between words by training distributed word vectors. Figure 2 shows the network structure of the model.

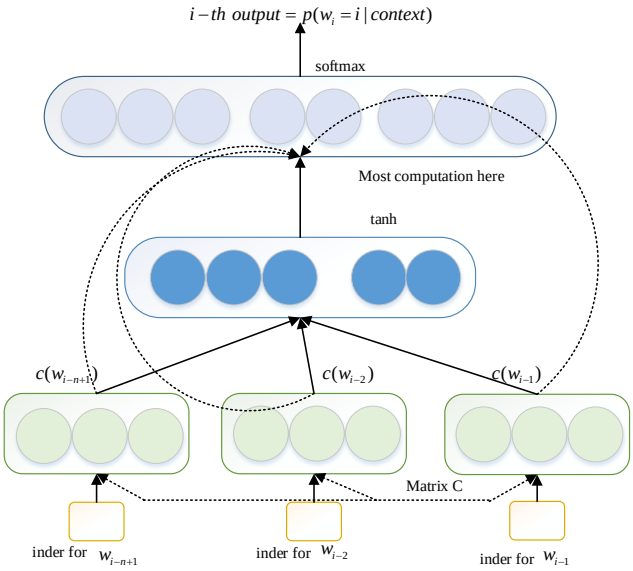

Figure 2 Text language model

The above parts constitute the overall word vector training model. The word vector parameters during training are shared by these three parts, and the overall training process is completed by the joint training of these three parts. In the specific implementation, the stochastic gradient ascent method is used for training, and the negative sampling method is used to optimize the three parts. Therefore, the core algorithms of the three parts have the same idea. Taking the first part as an example, the word vector update algorithm incorporating contextual information is shown in algorithm 1.

Algorithm 1 Word vector training algorithm incorporating contextual information

Algorithm: word vector update
Input: the ith sentence in the dataset s_i and the central word w containing m characters Output: word vector matrix Begin 1: obtain the vector representation e_w of the word w in the word vector matrix and the vector representation of each character that makes up w. The vector representation of the k th character is c_k. 2: calculate the vector representation of word w at training time using equation (12) x_w. 3: select the context word $\{v\}$ of w according to the set window size, and obtain its negative sampling result NEG(v), and then merge the two sets to form a new set $T = \{v\} \cup NEG(v)$. 4: initially define the update intermediate quantity p such that p =0; 5: sequentially traverse the set T and select word $u \in T$. 6: combining u auxiliary vector θ^u, indicator function $L^v_{(u)}$, learning rate η, and activation function σ, compute the intermediate quantities $q = \sigma(x_w^T * \theta^u), g = \eta(L^v_{(u)} - q)$. 7: Compute $p = p + g\theta^u$ and then $\theta^u = \theta^u + gx_w$ based on the intermediate quantity g. 8: If the set T has not been traversed, then turn 5 9: update word vector e_w of w, make $e_w = p + e_w$, then update all character vectors of word w, update character vector c_k of the k th character to $c_k = p + c_k$, end of one update of word vector. End

In summary, the overall training steps of the exacting word vector pre-training model can be described as.

(1) Preprocessing the training set with the text preprocessing method described in 3.1.2 and constructing the seed sentiment lexicon.

(2) Initialize the training parameters of the model, such as the window size of the interception context, the dimensionality of the training vector, the auxiliary vectors of each part, and the learning rate.

(3) Input the training set data into the model for training.

(4) After the training of the model is completed, the completed word vectors are obtained to complete the exact word vector pre-training process.

The co-word analysis method can visualize the network formed by the keywords and the sparsity characteristics, which can grasp the structure and characteristics of the research object more completely and accurately in the study. The preliminary keyword word frequencies were obtained by counting the class keywords derived from the rooting theory, and the keyword word frequencies were imported into Excel for data processing to obtain the keyword co-word matrix (the number of occurrences of every two keywords in the same policy). Then SPSS is used to further process it into a similarity matrix, and the high-frequency keywords are clustered and analyzed through the similarity matrix, and the keywords are grouped into large and small clusters by the clustering algorithm, so as to obtain a clustering dendrogram, and the structural model of the policy is constructed again according to its development, and the optimal one is selected in comparison with the model obtained from the zapping theory.

The process required constant sampling, repeated reading of the data and revision and comparison of the codes until no new concepts or classes were found. In the process of coding the 73 policy texts, starting from the 57th text, the 58th to the 73rd text, the coding process did not extract any new concepts and genera, and the codes were all able to be grouped into the previously coded genera, indicating that the current codes were saturated.

After the axial coding is completed, the word frequency of keywords can be counted according to the coding. 27 keywords of policy texts are obtained here (e.g., the word frequency of "education equity" is 12, which means that the keyword appears in 10 policy texts, and only one time is counted when it appears in the same policy several times), and the keywords are counted and organized. Table 1 shows the key words.

Table1 Frequency of keywords in the texts of the three comprehensive education policies

Keywords	frequency	Keywords	frequency	Keywords	frequency
Regulatory coordination	42	Information Technology Education Applications	22	Cross-border integration	12
Resource Development and Sharing	38	Informatization Support Services	21	Educational Equity	12
Three Communication Project	37	Public Service System	21	National Network Construction	12
Teachers Training	33	Investment and financing support	15	Information capacity	6
Education Management and Government	29	Information Security	16	Talent Development	5
Policy Environment	29	Rural Education	16	Cultural Development	4
Campus Network Construction	24	Pilot Demonstration	13	Distance Education	4

Teacher Education Informatization	22	Education Reform	13	Smart Education	4
Platform Construction	23	Dedicated Funds	11	Teaching materials construction	4

Based on the word frequency of the policy source material and keywords, the number of times each two keywords appeared in the same text was counted, and the pivot table function of Excel was used to draw the co-word matrix (e.g., the data in row 2 and column 3 of the matrix represent the number of times two keywords with serial number 2 and serial number 3 appear together in the same policy). The more times a pair of words appear together in the same policy, the more closely connected the topics represented by the pair are. The keywords obtained from Table 1 were organized and the co-word matrix generated is shown in Table 2.

Table 2 Keyword co-word matrix of the texts of the three whole education policies (partial)

	PSS	NNC	TC	TEI	TT	ER
Public Service System(PSS)	1	6	0	11	10	7
National Network Construction(NNC)	6	1	0	7	4	4
Textbook Construction(TC)	0	0	1	1	2	0
Teacher Education Informatization(TEI)	11	5	2	0	14	4
Teacher Training(TT)	10	6	0	12	1	6
Education Reform(ER)	5	4	0	6	5	1

The prerequisite for performing clustering analysis is the need to convert the co-word matrix into a similarity matrix that represents the correlation between word pairs. In this paper, we use cosine similarity, and in text mining each keyword is also given dimensions, and each dimension can be represented by a vector, and the frequency of the keyword in each dimension is also its frequency in the document. Therefore, the similarity between keywords can also be derived using cosine similarity. Some of the results of the similarity matrix obtained by cosine similarity in this paper are shown in Table 3.

Table 3 Similarity matrix of keywords in the text of Sankyo Education Policy (partial)

	Public Service System	National Network Construction	Textbook Construction	Teacher Education Informatization	Teacher Training	Education Reform
Public Service System	1.000	0.861	0.786	0.853	0.872	0.818
National Network Construction	0.863	1.000	0.618	0.905	0.871	0.826
Textbook Construction	0.788	0.615	1.000	0.598	0.692	0.495
Teacher Education Informatization	0.853	0.904	0.596	1.000	0.780	0.737
Teacher Training	0.870	0.872	0.692	1.782	1.000	0.781

Education Reform	0.817	0.828	0.495	0.737	0.782	1.000

IV. CASE STUDY

The investigation is carried out from three perspectives: the all-round development of personality, the whole process development and the all-round development on which moral education work is based. The content of the questionnaire is also designed around these three aspects. By asking a series of questions, we can obtain real information about the overall composition of the University, the whole process and the whole scope of education from students and staff.

The respondents of this survey include students and teachers from multiple institutions, with 585 students and 173 teachers from China, and their institutions cover finance and economics, politics and law, teacher training, science and technology, medical science, and vocational education. The subject majors of the interviewees include science and technology, literature and history, and art, and the academic levels include college students, undergraduates, master students and doctoral students.

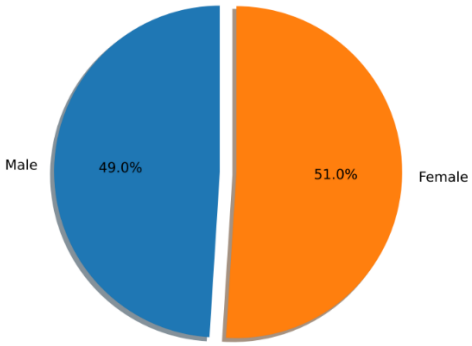

Figure 3 Gender of Survey Respondents (Students)

As shown in Figure 3, this chart reflects the gender situation of the survey respondents, from the data, the proportion of men and women is relatively balanced, with 287 male respondents and 298 female respondents, accounting for 49% and 51%, respectively. From Figure 4, we can see that among all the respondents, there are 150 specialists, accounting for 25.64%; 292 undergraduates, accounting for 49.91%, the largest number and the highest percentage; 137 master students and 6 doctoral students, accounting for 23.42% and 1.03% respectively.

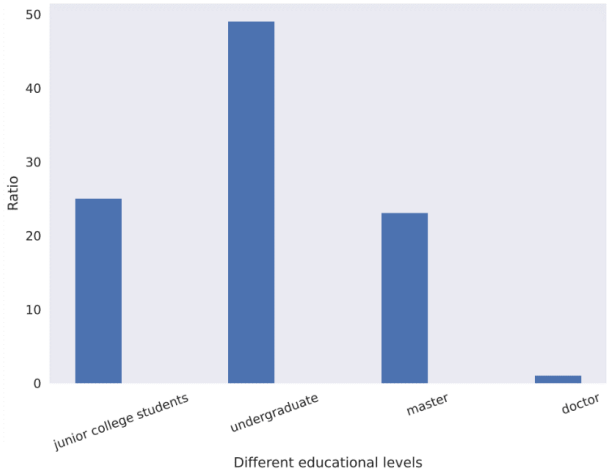

Figure 4 Survey respondents' education status (students)

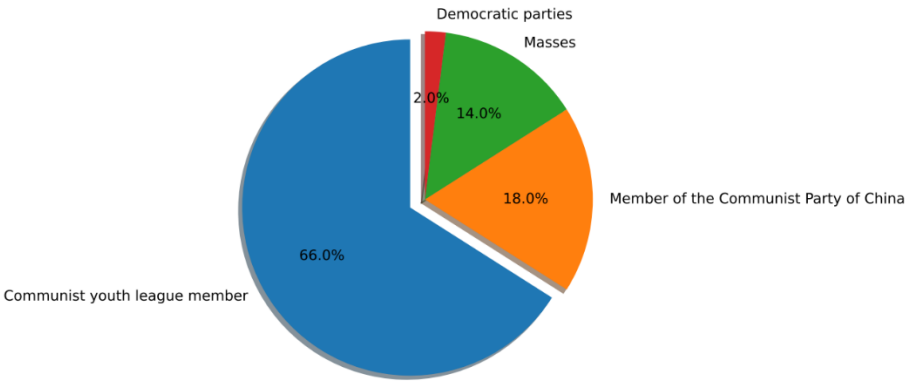

Figure 5 Political outlook of the respondents (students)

From Figure 5 that this study focuses on the work of moral education of college students, so the political outlook of the respondents is also crucial. Among the 585 students surveyed, 387 were members of the Communist Youth League, accounting for 66.32%, followed by 105 members of the Communist Party of China, accounting for 17.88%, and including 14.51% of the masses and 1.29% of the democratic parties, with 85 and 8 members respectively.

Among the surveyed teachers, the distribution of men and women is roughly balanced, with 92 male teachers and 81 female teachers, accounting for 53.18% and 46.82% respectively, and the gender distribution of the survey is reasonable, see Figure 6.

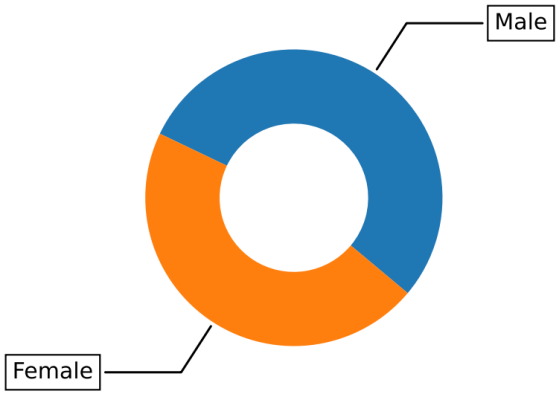

Figure 6 Gender of survey respondents (teachers)

Among these 173 faculty members, they come from three categories of posts, teaching and research posts, administrative posts and logistic service posts, including 55 teachers of general and specialized courses in terms of categories, accounting for 31.87%. In terms of categories of teachers interviewed the most, including full-time teachers of various majors and teachers of general courses, the number of teachers of moral theory courses interviewed 40 people, accounting for 23.15%. The number of interviewed teachers of moral theory courses is 40, accounting for 23.15%, and also includes 28 party and administrative managers, accounting for 16.32%, as well as 15 logisticians, accounting for 8.62%. In terms of political orientation, all teachers in universities are basically CPC members, with the proportion of party members reaching 82.78%, and some democratic parties and masses, which account for less than 20% in total, see Figure 7.

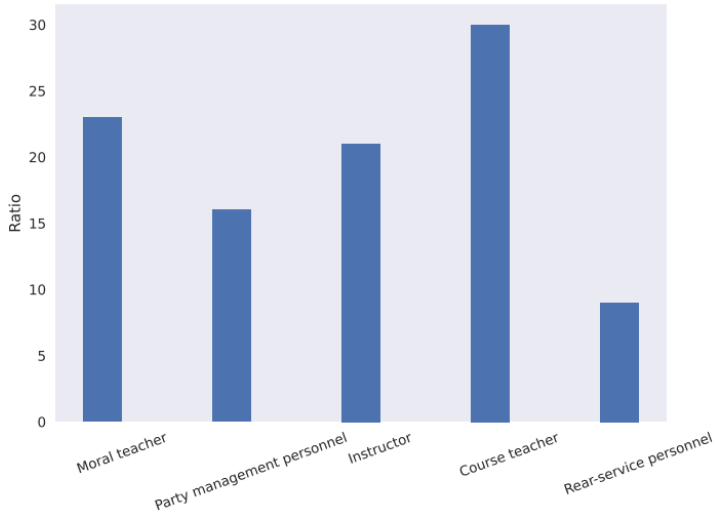

Figure 7 Job categories of interviewed teachers

V. CONCLUSION

The comprehensive reform of "triple education" is an increasingly profound and gradual process. In practice, universities should, according to their own conditions and students' characteristics, strengthen the cooperation of the education system for all, strengthen the effective interaction of the whole process of education, ensure the organic connection of high-quality education, and establish a long-term education mechanism. It has its own characteristics. Establish an effective mechanism for long-term education characterized by self-development.

ACKNOWLEDGMENT

This work no funding supported.

REFERENCES

[1] Yan, R. (2021). Psychological Motivations of Students for Book Borrowing in Libraries of Higher Vocational Colleges from the Perspective of “Three Comprehensive Education”. *International Journal of Emerging Technologies in Learning (iJET)*, *16*(8), 77-92.

[2] Liu, Y., Zhang, Y., & Yu, S. (2021). Research and Practice on the Training Mode of Computer Graduate Students from the Perspective of" Three Comprehensive Education". International Journal of Social Science and Education Research, 4(11), 242-245.

[3] Wang, C. (2021). Exploration of the Educational Path Based on the Integration of Innovation and Professionalism Based on the Tutorial System under the Concept of" Three-Dimensional Education". International Journal of Social Science and Education Research, 4(5), 329-332.

[4] Salmela-Aro, K., & Tuominen-Soini, H. (2010). Adolescents’ life satisfaction during the transition to post-comprehensive education: Antecedents and consequences. Journal of Happiness Studies, 11(6), 683-701.

[5] Wenliang, Z. H. A. N. G., & Hainan, L. I. (2020). Research on the Strategy of National Defense Education in Universities to Promote the Comprehensive Quality of College Students——Take Zhuhai College of Jilin University as an Example. The Theory and Practice of Innovation and Enntrepreneurship, 3(13), 68.

[6] G. Cai, Y. Fang, J. Wen, S. Mumtaz, Y. Song and V. Frascolla, "Multi-Carrier M-ary DCSK System With Code Index Modulation: An Efficient Solution for Chaotic Communications," in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 6, pp. 1375-1386, Oct. 2019

[7] Wenyu, L. I. U., Jiayan, L. U. N., & Jing, F. A. N. (2020). Evolution and Implementation of “Three Comprehensive Education” Policies from the Perspective of Frame Analysis. Modern Education Management, (9), 14.

[8] K. Chandra, A. S. Marcano, S. Mumtaz, R. V. Prasad and H. L. Christiansen, "Unveiling Capacity Gains in Ultradense Networks: Using mm-Wave NOMA," in IEEE Vehicular Technology Magazine, vol. 13, no. 2, pp. 75-83, June 2018

[9] Saghezchi, Firooz B., Ayman Radwan, Jonathan Rodriguez, and Tasos Dagiuklas. "Coalition formation game toward green mobile terminals in heterogeneous wireless networks." IEEE Wireless Communications 20, no. 5 (2013): 85-91.

[10] Palanisamy, S.; Thangaraju, B.; Khalaf, O.I.; Alotaibi, Y.; Alghamdi, S.; Alassery, F. A Novel Approach of Design and Analysis of a Hexagonal Fractal Antenna Array (HFAA) for Next-Generation Wireless Communication. Energies 2021, 14, 6204. https://doi.org/10.3390/en14196204

[11] Goldfarb, E. S., & Lieberman, L. D. (2021). Three decades of research: The case for comprehensive sex education. Journal of Adolescent Health, 68(1), 13-27.

[12] Alsubari, S. N., Deshmukh, S. N., Alqarni, A. A., Alsharif, N., H., T. (2022). Data Analytics for the Identification of Fake Reviews Using Supervised Learning. CMC-Computers, Materials & Continua, 70(2), 3189–3204.

[13] Liu Qingfeng, Liu Chenxuan, Wang Yanan, etc. Integrating external dictionary knowledge in conference scenarios The field of personalized machine translation method [J]. Journal of Chinese Informatics, 2019, 33(10): 31-37.

[14] Freudenberg, N., Goldrick-Rab, S., & Poppendieck, J. (2019). College students and SNAP: The new face of food insecurity in the United States. American Journal of Public Health, 109(12), 1652-1658.

[15] Bansode, S. A., V. R. More, S. P. Zambare, and M. Fahd. "Effect of constant temperature (20 0C, 25 0C, 30 0C, 35 0C, 40 0C) on the development of the Calliphorid fly of forensic importance, Chrysomya megacephala (Fabricus, 1794)." Journal of Entomology and Zoology Studies 4, no. 3 (2016): 193-197.

[16] Orphan, C. M., & Broom, S. (2021). Life at the “people’s universities”: Organizational identification and commitment among regional comprehensive university faculty members in the USA. Higher Education, 82(1), 181-201.

[17] Sumuer, E. (2018). Factors related to college students’ self-directed learning with technology. Australasian Journal of Educational Technology, 34(4).

[18] Ali, Rahman, Muhammad Hameed Siddiqi, and Sungyoung Lee. "Rough set-based approaches for discretization: a compact review." Artificial Intelligence Review 44, no. 2 (2015): 235-263.

[19] Rotas, E. E., & Cahapay, M. B. (2020). Difficulties in Remote Learning: Voices of Philippine University Students in the Wake of COVID-19 Crisis. Asian Journal of Distance Education, 15(2), 147-158.

[20] Zhu, Y., Zhang, J. H., Au, W., & Yates, G. (2020). University students’ online learning attitudes and continuous intention to undertake online courses: A self-regulated learning perspective. Educational technology research and development, 68(3), 1485-1519.

[21] Jingchun Zhou, Qian Liu, Qiuping Jiang, Wenqi Ren, Kin-Man Lam, Weishi Zhang*. Underwater image restoration via adaptive dark pixel prior and color correction. International Journal of Computer Vision, 2023. DOI :10.1007/s11263-023-01853-3.