Analysis of Needs of Higher Vocational College Students for Textbooks Based on Text Mining Techniques

Lijuan Xu[1,2(✉)], Yu Zhao[1,2], Zhenkui Xi[1,2], and Ping Wang[1]

[1] School of Business Administration, Shandong Institute of Commerce and Technology, Jinan 250100, China

[2] Innovation College, North-Chiang Mai University, Chiang Mai 50230, Thailand

g636501016@northcm.ac.th

Abstract. Text data mining techniques are used to process and analyze the questionnaire data collected. The needs of higher vocational college students for textbooks are explored in the aspects of textbook selection & management, content, form, and integration with teaching. Suggestions are given for textbook management departments, preparers, and teachers using them.

Keywords: Text mining technique · higher vocational college student · textbook need

1 Introduction

Textbooks are "the center of practical teaching activities for teachers and students" (Zeng Tianshan, 1997). Textbooks are the main carrier of education, the foundation for teachers' teaching, the main channel for students to acquire knowledge and develop abilities, and the key to achieving satisfactory teaching results and learning quality. Vocational colleges are an important player in cultivating high-quality technical and skilled personnel in China. Textbooks are an important carrier for students of higher vocational colleges, as one of the end users, to form the basic knowledge structure, and acquire and develop technologies and skills. There are certain requirements for the textbook in both form and content. High-quality textbooks will help students in acquiring knowledge and skills and thus promote vocational education institutions to cultivate highly-skilled talents. Understanding the needs of students for the use of textbooks in the process of learning may guide the textbook management departments, inspire the authors to optimize their content and form, guide teachers to put them to better use, and push forward the reform of curriculum and teaching.

In traditional surveys, the need for textbooks is usually obtained using questionnaires. However, the current learning needs of students are highly personalized, and their learning environment is information-based and intelligent. Although traditional surveys can offer data such as satisfaction with textbooks, it is difficult to explore the

V. Balakrishnan et al. (Eds.): ICMEIM 2022, AHSSEH 6, pp. 693–705, 2022.
https://doi.org/10.2991/978-94-6463-044-2_88

real needs of students for the use of textbooks. With text mining techniques, the sentiment tendency of text can be calculated through text mining, data processing, and data analysis. Through the sentiment tendency of students for texts, we can understand their real needs for textbooks. In this way, we can extract needed or unknown knowledge from the original text, fully understand user needs, upgrade products and optimize services to cater to such needs.

From the perspective of text analysis, this paper explores a series of needs of higher vocational college students for the use of textbooks, provides suggestions for the management departments of higher vocational colleges to select and manage textbooks and provides guidance for college teachers to prepare and use textbooks.

2 Research Design

2.1 Research Summary

The needs of teachers and students for textbooks have always been the focus of attention of scholars. They have analyzed and researched the need for the use of textbooks from different perspectives in theory and practice. Theoretically, Guo Huizhen (2009) proposed that the development of textbooks for vocational education should be aimed at meeting the needs of professional posts [1]. Based on the analysis framework of "two-way need", Xu Ye (2021) analyzed the adaptive development of textbooks for vocational colleges in China [2]; Dai Sijun et al. (2015) analyzed the characteristics of the market demand for undergraduate textbooks and put forward suggestions, which provided a reference for the paper [3]; In practice, Chen Zhiqi (2012) discussed the design ideas of higher vocational English textbooks based on the analysis of vocational needs [4]; because of the development practice of new-type textbooks, Su Chonglai et al. (2021) researched the development model of new loose-leaf textbooks on business in vocational colleges, which integrate theory, virtual and operation [5]. However, from current literature, scholars mostly conduct qualitative analysis to research and develop the needs for textbooks, in which there are certain limitations. Based on the above research, this paper employs the text mining technique to analyze the needs of higher vocational college students for the use of textbooks and put forward effective suggestions for the preparation and management of textbooks.

2.2 Data Collection and Processing

The data comes from the online questionnaire survey, and the respondents are freshmen and sophomores in higher vocational colleges. The questionnaire questions include questions about personal information and an open-ended question: Please describe your evaluation of textbooks used in the current course. A total of 6,023 questionnaires were gathered. The information gathered about open-ended questions in the questionnaire is shown in Table 1.

The data comes from the questionnaire survey. The text information about the response by students is in disorder. Thus, it is necessary to preprocess the gathered text information, so that valuable and meaningful research results can be obtained in

Table 1. Information from questionnaires gathered (partial)

No.	Response text information	No.	Response text information
1.	There are pictures, words, and stories.	11.	None
2.	Brief	12.	Detailed content
3.	Make full use	13.	Get more knowledge from the new edition
4.	Talk more about key points. Too many words in textbooks are annoying	14.	Very satisfied
5.	Color pictures are needed	15.	Take current professional certificate examinations into consideration
6.	Clear and simple	16.	Not bad
7.	Online and offline combination	17.	None
8.	Hope for key points, and they are a bit more other things	18.	Closely related to life
9.	Good	19.	Simple and easy to understand
10.	Very good	20.	Novel and many key contents

the analysis. The data is processed mainly by removing meaningless numbers or many meaningless numbers reflecting emotions (such as "111" and "none") and some pause words (such as "yes", "but", "maybe" and "although").

2.3 Analysis Methodology

2.3.1 Introduction to TF-IDF Concept and Algorithm

To research the evaluation of textbooks by higher vocational college students, explore their needs for the use of textbooks, and provide strategies for the preparation of textbooks, this paper uses the TF-IDF method to extract keywords and judge their importance.

TF-IDF is an effective keyword extraction algorithm, which judges the importance of a word according to the word frequency and inverse text frequency. The word frequency often reflects the importance of the evaluation content. Therefore, compared with TextRank and LDA topic algorithm, it can better reflect its superiority. The algorithm is divided into two parts: one is the TF algorithm, and the other is the IDF algorithm. In actual use, the calculation formula of TF is as follows:

$$\mathrm{tf}_i = \frac{n_{ij}}{\sum_k n_{kj}} \tag{1}$$

where n_{kj} refers to the appearance frequency of a word in the document. After the word frequency is normalized, the denominator is the total appearances of each word in the statistical document.

IDF is calculated as:

$$idf_i = \log\left(\frac{|D|}{1 + |D_i|}\right) \tag{2}$$

where D refers to the total number of documents in the document set, and D_i refers to the number of documents in which word i appears in the document set. Laplacian smoothing is used to add 1 to the denominator, which avoids the situation that the disappearance of some new word segments in the corpus causes the denominator to be zero, and enhances the robustness of the algorithm. Then, the TF-IDF value of word i is as follows:

$$tfidf_i = tf_i \times idf_i \tag{3}$$

2.3.2 Semantic Network Analysis

For the data about the evaluation of the use of textbooks by higher vocational college students, it is suitable to conduct semantic network analysis, which is divided into two steps: Step 1: keyword extraction, word frequency, and sentiment analysis. Step 2: co-occurrence analysis, which obtains the co-occurrence matrix by the importance of keywords, and then further understands the relationship and intimacy between words according to the co-occurrence matrix.

2.4 Research Tools

Python and ROSTCM6 tools are mainly used for analysis. First, Python is used for data and preprocessing. Based on the TF-IDF algorithm, the ROSTCM6 tool is used to calculate the TF-IDF value of data about textbook evaluation by students to extract keywords. Second, Sentiment analysis is performed using SnowNLP, a third-party library for Python. SnowNLP is a class library written in python. Most current natural language processing libraries are for English. While SnowNLP is intended for Chinese, it can easily process Chinese texts. Finally, ROSTCM6 is used for semantic network analysis, and the co-word matrix obtained from the semantic network analysis is visualized.

3 Results of Data Analysis

3.1 Calculation and Analysis of TF-IDF Value and Frequency of Each Keyword

3.1.1 Calculation and Analysis of TF-IDF Values of Keywords

The gathered text data is collated into a CSV file, and the TF-IDF values of data about student evaluation of textbooks are calculated by using the ROSTCM6 tool to extract keywords. The TF-IDF values of the top 30 keywords are in descending order, and the results are shown in Table 2.

As seen in Table 2, “content” is the keyword with the highest TF-IDF value, indicating that students using textbooks are most concerned about the content of textbooks. Secondly, several words such as “practical”, “combine”, “novel”, “rich” and “interesting” appear most frequently in the content of the texts evaluated, showing that most students pay more attention to the novelty of knowledge points in the content of the textbook, whether they are rich and interesting, whether they combine theory with practice, and hope to use textbooks with both rich and interesting content and novel design.

Table 2. TF-IDF values for top 30 keywords (The table is organized by the author based on the data analysis results.)

No.	Keyword	TF-IDF value	No.	Keyword	TF-IDF value	No.	Keyword	TF-IDF value
1	Content	25300	11	Comprehensive	7600	21	Practical	4700
2	Pragmatic	18200	12	Theory	7500	22	Knowledge points	4600
3	Combine	13900	13	Cheap	7400	23	Popular	4500
4	Novel	12600	14	Clear	6700	24	Course	4400
5	Teacher	8500	15	Learning	5800	25	Practice	4100
6	Based on	8400	16	Price	5700	26	Vivid	4100
7	Knowledge	8000	17	Detailed	5600	27	Simple	3700
8	Easy to understand	7800	18	Hope	5600	28	Professional	3100
9	Rich	7700	19	Key point	5200	29	Student	2800
10	Interesting	7700	20	Nowadays	5000	30	Easy	2800

3.1.2 Frequency Calculation Result and Cloud of Each Keyword

Collate the gathered text data into the CSV file, and use Python for word segmentation and word frequency statistics. The following 39 words (in total) appear more than 20 times, as shown in Table 3. At the same time, the cloud of 39 words is sorted out, and the visualization results are shown in Fig. 4:

It can be seen from Table 3 and Fig. 2 that in the top 30 keywords, in addition to the keywords (such as "content" and "reality") related to the content of textbooks, there are also "cheap" and "price" words, showing that students hope that the textbooks will be practical, easy to use and economical. The word "teacher" also appears frequently, showing that students are equally concerned about the use of textbooks by teachers in the classroom (Fig. 1).

3.2 Sentiment Tendency Analysis

3.2.1 Overall Sentiment Proportion Analysis

As the sentiment dictionary is easier to obtain than the corpus, the sentiment dictionary is used for sentiment analysis. The sentiment dictionary is used to conduct analysis mainly as per the following steps (taking sentiment polarity analysis as an example):

1) Read the sentiment dictionary. Acquire a commendatory word list, a derogatory word list, and a neutral word list; and acquire a sentiment classification word list and sentiment intensity thereof.
2) Process the text to be analyzed. It is mainly to read the text, split it into sentences, and make word segmentation for each sentence.

Table 3. List of words and frequencies that appear more than 20 times (The table is organized by the author based on the data analysis results.)

No.	Word	Frequency	No.	Word	Frequency
1.	Content	218	21.	Course	42
2.	Novel	125	22.	Close to	42
3.	Teacher	82	23.	Practice	41
4.	Interesting	72	24.	Combine	39
5.	Theory	70	25.	Price	35
6.	Knowledge	65	26.	Vivid	34
7.	Take reality into consideration	64	27.	Student	28
8.	Cheap	63	28.	Picture	25
9.	Easy to understand	62	29.	Useful	25
10.	Learning	58	30.	Better	25
11.	Detailed	56	31.	Practicality	24
12.	Understand	55	32.	Based on actual conditions	24
13.	Simple	54	33.	Ok	24
14.	Clear	50	34.	New	23
15.	Teach	48	35.	Professional	22
16.	Practical	47	36.	Picture and text	22
17.	Key point	46	37.	High	21
18.	Knowledge point	46	38.	Moderate	21
19.	Consistent	44	39.	Full of pictures and texts	21
20.	Popular and easy to understand	43			

3) Calculate the sentiment score of sentences. Find the sentiment classification of each word in a sentence, read its emotional intensity, and subtract the negative sentiment score from the positive sentiment score to get the total sentiment score of the sentence. It should be noted that negative words and degree adverbs (if any) in the sentence will affect the sentiment direction and intensity of the sentence.

3.2.2 Data Analysis Results

Collate the gathered text data into a CSV file, use Python's third-party library SnowNLP for sentiment analysis, and make sentiment classification statistics, as shown in Table 3.

It can be found that, among the gathered data, the positive evaluation takes a big proportion (89.2%), followed by the neutral evaluation (8.09%), and the smallest proportion is the negative evaluation (2.71%); That is to say, the current sentiment towards the use of textbooks is relatively positive on the whole (Table 4).

Fig. 1. Keyword cloud of textbook evaluation text (The word cloud is drawn by the author.)

Table 4. Sentiment classification statistics (The table is organized by the author based on the data analysis results)

Negative	84	2.71%
Positive	2767	89.20%
Neutral	251	8.09%

3.2.3 Word Frequency and Visualization Analysis of Positive and Negative Sentiment Texts

Analyze and visualize positive and negative/neutral sentiment text contents respectively, with the results shown in Fig. 3 and Fig. 4.

In the word cloud generated by positive sentiment texts, It can be found that “novel”, “teacher”, “theory”, “knowledge” and “practice” receive high attention, indicating that students have high requirements for the novelty of textbook forms and contents, the reasonableness of the use of textbooks by teachers and the adequacy of integrating theory with practice in the process of using textbooks. In the word cloud generated by positive emotional text, words such as “content”, “actual conditions”, “utilization rate” and “update” draw attention, showing that students pay more attention to the content quality of textbooks, whether the knowledge and skill knowledge points of textbooks and the industry frontier can be updated in time when using textbooks.

3.3 Semantic Network Analysis

Through the ROSTCM6 platform, the information text is processed to extract high-frequency words, meaningless words are filtered, and line characteristics are extracted to build a semantic network and matrix. It can be found through the semantic network

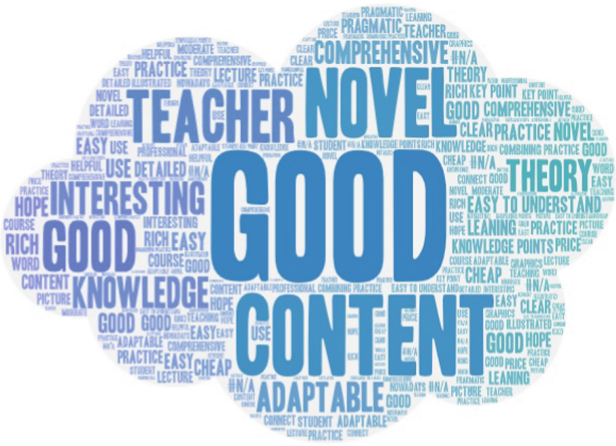

Fig. 2. Word cloud of positive sentiment text (The word cloud is drawn by the author.)

Fig. 3. Word cloud of negative/neutral sentiment text (The word cloud is drawn by the author.)

that, students' needs for textbooks focus on the knowledge and skills points of textbooks, and they are closely related to the paths of "conforming to the development of the times", "combining practice", "accommodating the use of courses", "interesting and vivid" and "easy to understand and clear knowledge points". This also reflects the students' attention to and need for the content quality of textbooks when using them.

4 Conclusions and Suggestions

4.1 Conclusions

Under the background of the elevated status of vocational education, the training of skilled talents requires high-quality textbooks that can better meet the needs of students. It is found through the above text analysis that, the current needs of higher vocational college students for textbooks mainly focus on the selection & management, content, form, and use of textbooks.

4.1.1 Students Hope that the Selection and Management of Textbooks Will Be Further Standardized

The textbooks that students want to use should be up to date and in line with the frontier of professional development. This requires attention to the novelty of the content of textbooks when selecting textbooks, which is reflected in the publication time of textbooks and whether they contain the frontier contents of the industry. At the same time, students hope to use textbooks with rich content and a high utilization rate, which requires that the use and evaluation of textbooks should also be assessed in the selection and management of textbooks. At the same time, attention should be paid to the acceptability of textbook prices to students, and the cost of textbooks should be reduced on the premise of meeting teaching requirements.

4.1.2 Students Hope that the Content of Textbooks Will Be More Novel, Interesting and Practical

Students' requirements for the content of textbooks are not only limited to the aspect of "novel", but also practicality, interest, and practicality. Practicality refers to the connection between the content of the textbook and the skills for practical operation on the post. Interesting means that textbook cases are attractive to students. Pragmatically means that textbooks can be used as a reference by students. In addition, highlighting the key points, combining pictures and texts, and new-type textbooks are also the requirements of students for textbooks.

4.1.3 Students Hope that the Form of Textbooks Will Be More Digitalized and Interactive

Because of the novel form, loose-leaf textbooks and textbooks gradually become popular among students. Most of the loose-leaf textbooks published in recent years are concise, easy to understand, rich, colorful, vivid, and interesting, which can help students to improve their learning effect. Teaching contents are presented in the form of a textbook. Only when novel and attractive textbooks are recognized by teachers and students and students are more willing to take the initiative to learn, can the teaching effect can be enhanced.

4.1.4 Integration with Teaching

Textbooks and teaching by teachers do not replace each other and need to play their respective roles. Vocational education textbooks should function to educate students ideologically, transmit culture and experience, and build professional abilities, which will be accomplished by teaching teachers to students in and out of the classroom. It can be said that the same textbooks taught in different methods will deliver different effects. It is found from the above analysis that, students need textbooks commensurate with teaching, hope that textbooks can help them clarify key and difficult points, form clear learning objectives and boost motivation for learning. At the same time, students also realize that textbooks are not the only teaching resources and they should be complementary to other teaching resources such as micro-lesson.

4.2 Suggestions

It can be found based on the above analysis that, the preparation, use, and management of textbooks involve three key subjects: textbook management department, preparer, and user. This paper gives suggestions from the perspectives of the textbook management department, preparer, and user.

4.2.1 The Textbook Management Departments of Higher Vocational Colleges Should Step Up the Planning and Organizational Management of Textbooks

First, ramp up efforts in planning the construction of textbooks for higher vocational colleges. The textbook management departments should plan the construction of their textbooks according to the requirements of national strategies and the needs of local economic and social development, the characteristics and orientation of college running, professional and curriculum clusters, and teachers, and give financial support. Second, strengthening the organizational management and construction of textbooks, and the management of textbook preparation teams. In addition to post-event assessment, higher vocational colleges should also conduct pre-event guidance and concurrent supervision based on considering major subdivisions, human resource allocation, and other issues. Third, Higher vocational colleges should strengthen the organizational construction of textbook selection and teaching committees. Therefore, it is suggested that higher vocational colleges should strengthen the organizational construction of textbook planning and selection committees. On the one hand, more peer experts, and business people interested in education and social personages should be introduced in this regard. On the other hand, the committees should be granted practical power to promote their continuous and stable operation and establish their authority.

4.2.1.1 Textbook Selection Management Procedures and Rules Need to Be Clarified

Textbook management departments should establish and improve multi-level procedures for reviewing textbook selection to ensure sound, standardized and reasonable, review and supervise the ideology, utilization rate, selection, and evaluation of textbooks, and maintain the integrated supervision mechanism for the selection, use, and evaluation of textbooks used by colleges. At the same time, it is also necessary to effectively supervise

and evaluate the use of textbooks by teachers in the teaching process through teaching supervision and inspection, and other forms [6].

4.2.2 Textbook Developers Should Set Up a Diversified Preparation Team, Design High-Value Textbook Content, and Enrich Digital Interaction Forms

4.2.2.1 Developers Should Be Diversified

It is necessary to change the current situation of textbook preparers working alone, and form a development team of vocational education textbooks. There is no lack of textbook preparation teams in vocational colleges, but they have not been combined to form a diversified textbook development team who understands policies, is proficient in teaching, and has strong abilities. Enterprise personnel and social personages in appropriate majors can be introduced to cultivate talents by moving teaching research from the traditional classroom to the site combining enterprise and education, to form a diversified textbook development and construction team consistent with the requirements of the new era.

4.2.2.2 The Content Should Be Cutting-Edge, Valuable, Interesting, and Interactive

The National Textbook Construction Plan for Universities, Colleges and Primary and Secondary Schools (2019–2022) issued by the National Textbook Committee points out that: The key to vocational education textbooks is to embody "new" and "practical", reflect new knowledge, technology, new process, and method, make a timely revision, enhance the ability to serve the national industrial development, and solve the problem of " a few of many textbooks are excellent". As an important tool in teaching, vocational education textbooks should not only be based on the current conditions of students but also accommodate the needs of students for future career development, to promote the development of students' capabilities [7]. First, the content of textbooks should be designed based on the frontiers of industrial development and global and future trends to embody the extensibility and growth potential. Second, textbooks should have a clear value in knowledge acquisition and skill enhancement and should be highly operable. Third, textbooks should be able to stimulate students' learning motivation and interest, and interesting case stories should be introduced to make them more interesting. Fourth, textbooks should help enhance the interaction on teaching content and the effectiveness of textbook use and teaching by teachers should be enhanced.

4.2.2.3 The Form of Textbooks Should Be Interactive, Digitalized, and Visualized

In the era of the digital economy, all walks of life have more connections with the Internet and big data, and the development of teaching content and the need for teaching reform impose new requirements on textbooks. It is proper to develop digital textbooks, link with the Internet platform, and create a model of "Internet + " for educational informatization. Various digital resources and educational informatization projects tend to become mature. Full use of various informatization application platforms and tools can meet the requirements of students that teaching resources (such as textbooks) are scientific "at any time and everywhere". At the same time, it is necessary to promptly update textbooks with the upgrading of information technology and industry, combine

paper textbooks with Internet resources, expand the boundaries of textbooks, and provide systematic services for students' autonomous learning.

4.2.3 Textbook Users Should Pay Attention to the Mutual Complementarity and Integration of Textbooks and Courses

In the era of educational informationization, classroom teaching in higher vocational colleges has the characteristics of "inquiry-based science education", "digitalization", "flexibility" and "modularization". The development of higher vocational textbooks should adhere to the principles of attaching equal importance to thinking and skill, carrying out moral education, paying attention to professional attributes, and strengthening the informationization of textbooks. Therefore, when using textbooks, teachers should pay attention to the fact that textbooks are passive and static information, and classroom teaching is to deepen textbooks, which can help students to allocate learning time effectively. At the same time, the use of multimedia resources to supplement the content of textbooks can better meet the requirements of curriculum standards for teachers' teaching and promote students' mastery of knowledge and skills.

Acknowledgment. This study was supported by the Research Foundation of:

1. The research project of vocational education teaching reform in Shandong Province in 2021, "Development and practice of loose-leaf textbooks based on competency-based business" (No. 2021031);

2. The Quality Engineering Project of Shandong Commercial Vocational and Technical College "Digital and Media-integrated Textbooks Construction and Path Research" (No. A105).

References

1. Guo Huizhen. (2009). Development of Vocational Education Textbooks to Meet Post Needs. Chinese Editors Journal (04), 36-37.
2. Xu Ye. (2021).Research on the Adaptive Development of Textbooks in Vocational Colleges in China under the Framework of "Two-way Need" Analysis. Education and Occupation (22), 63-68. doi:https://doi.org/10.13615/j.cnki.1004-3985.2021.22.010.
3. Dai Sijun & Gong Jingping. (2015). Characteristics Analysis of and Suggestions on Undergraduate Textbook Market Demand. Science & Technology and Publication (06), 87-89. doi:https://doi.org/10.16510/j.cnki.kjycb.2015.06.028.
4. Chen Zhiqi. (2012).Discussion on the Design of Higher Vocational English Textbooks Based on the Analysis of Vocational Needs. Vocational and Technical Education (20), 34-37.
5. Su Chonglai & Liu Xia. (2021). Development of New Loose-leaf Textbooks on Business in Vocational Colleges, Which Integrate Theory, Virtual and Operation. Vocational and Technical Education in China (11), 92-96.
6. Yang Xiaoli. (2010). Mining Publishing Resources from the Selection, Construction, and Management of College Textbooks. Editor's Friend (08), 38-40. doi:https://doi.org/10.13786/j.cnki.cn14-1066/g2.2010.08.001.
7. Zhao Hui. (2007). Strengthening the Planning of Textbooks Based on the Market Demand-Some Thoughts on the Planning of Computer Textbooks. Science & Technology and Publication (09), 3-5. doi:https://doi.org/10.16510/j.cnki.kjycb.2007.09.001.