TEXT MINING IN THE TREND OF EDUCATION 4.0:
A STUDY ON CLUSTERING MATHEMATICAL TERMS OF ALGEBRA TEXTBOOKS IN VIETNAMESE HIGH SCHOOLS

Tuan Anh Bui[1],
Diem-Chinh Thi Ho[2,+],
Chi Hong Nguyen[1],
Huynh Nhu Phan[1],
Linh Yen Nguyen[1]

[1]Department of Mathematics Education, Teachers College, Can Tho University, Vietnam;
[2]Department of Mathematics and Computer Science, University of Science, Vietnam National University Ho Chi Minh City, Vietnam
[+]Corresponding author ● Email: hothidiemchinh@gmail.com

ABSTRACT
In the context of Educational Revolution 4.0, text mining with digital tools plays an important role. Various techniques and softwares have been employed in text mining, among which the clustering technique using Atlas.ti, a German software, is widely used thanks to its versatility and open access. This article presents the results of clustering Mathematical terms in Algebra textbook in Vietnamese high schools with the support of Atlas.ti. Initial research results can yield the insight into the relationship among Mathematical terms in the curriculum, thereby, aiming for a better teaching process.

1. INTRODUCTION

In the era of 4th industrial revolution, the growth of digital data accompanies the demand for data transformation into useful and meaningful information and knowledge. This has attracted great attention to knowledge discovery and data mining (Gaikwad et al., 2014). One of the strikingly used techniques is text mining – a popular technique for text analysis. An outstanding study on this method, Tuan et al. (2019) employed text mining to analyze the poem Divan of Hafiz. Also, Don Swanson determined the cause of rare diseases by looking for indirect links in different subsets of biological literature (Hearst, 2003). Besides, Korhonen et al. (2009) used text mining assistive technology to identify and organize the scientific evidence needed to evaluate cancer risks. The results of the tests prove that the classification generated by experts is accurate, well-defined, and may be useful in practice. The emergence of online documents, especially big data, somehow troubles traditional data analysis methods (Jalali et al., 2020); yet has encouraged the development of tools to support the analysis of documents such as HyperResearch, Atlas.ti, CI-SAID which allow researchers to encode images, digitized speeches, and videos. Besides, NVivo and Atlas.ti allow video segments to be hyperlinked in a limited way (Gibbs et al., 2002).

In education, textbooks are an essential tool for both students and teachers, so the quality of textbooks is of particular interest. Textbooks are published according to the needs of educational institutions to meet the general and specific educational goals. According to Ivić et al (2013), improving the quality of textbooks is a crucial and meaningful factor to improve the quality of education in general; "particularly in developing countries, there is no single factor in improvement in the quality of education which is comparable with textbooks in terms of its impacts". Okeeffe (2013) pointed out that a good textbook must encourage the thirst for knowledge, consist of valuable contents and creativity-inspired aspects, motivational factors, accessibility, illustrations, learning guides. Textbook analysis is practical and meaningful in assessing the quality of textbooks as well as gaining a more general view of the role, the primary content of the textbook, the meaning and value that textbooks offer. Leburn (2002) presented textbook analysis in elementary teaching in Quebec over 40 years ago. The main goal is to examine the interaction between existing materials and how teachers exploit them. Chall's textbooks analysis (1997) showed the relationship between the difficulty level of elementary and secondary textbooks, teacher workbooks, and guides with learners' scores achieved on the SAT. Another study analyzed different versions of three textbooks for teacher education, explore the relationship between professional standards and the educational program for teachers in the field (Tummons, 2014).

Given the research literature, this paper focuses on analyzing the textbooks of Algebra and Analysis of grades 10, 11, and 12 (bilingual version) in Vietnam by text mining technique with the support of Atlas.ti software. The outline of the paper is offered as follows: Section 2 presents the theoretical basis of Text mining technique, the data

clustering method, and Atlas.ti software; and Vietnamese bilingual Algebra and Analysis textbooks grade 10, 11, 12 are also introduced afterwards. Section 3 describes the research methodology and textbook analysis results. Conclusions and implications for further research are offered in the last section.

2. LITERATURE REVIEW

2.1. Text Mining and Data Clustering Method

Text mining/Text data mining is a process of processing and extracting information contained in a document. This process involves a series of activities: text summary; document retrieval; clustering of documents; text classification; language identification; copyright; phrases identification. Besides, there are structured phrases and main phrases, extracted entities such as names, dates, and abbreviations; positioning of acronyms and their definitions; filling out predefined patterns with extract information, and so on (Witten et al., 2004). Previously, text mining and qualitative methods were thought to be different because using word count algorithms used to be considered as a quantitative method. Nevertheless, Krippendorff (2004) pointed out that text analysis is qualitative, reading text and counting words performed by humans or computers do not eliminate the qualitative nature of text. With the same view, Janasik et al. (2009) stated that the qualitative attribute of a study lies not in the data collection method, but in the data type and data analysis methods.

Furthermore, numerous researchers have applied text mining to qualitative research projects and see text exploitation as a feasible qualitative research method (Rajman & Besançon, 1998; Camillo et al., 2005; Hong, 2009; Janasik et al., 2009; Yu et al., 2011; Mahmoudi & Abbasalizadeh, 2019; Tian et al., 2019; Wang et al., 2020). According to Rose & Lennerholt (2017), advances in text mining accompany with the widespread adoption of information technology has granted wider access to large amounts of text documents as well as automated analysis techniques. Thus, it has opened up new possibilities for qualitative researchers in information systems and business and management fields. Gaikwad et al. (2014) proposed that the process of text mining should be conducted via 5 steps as follows (see Figure 1).
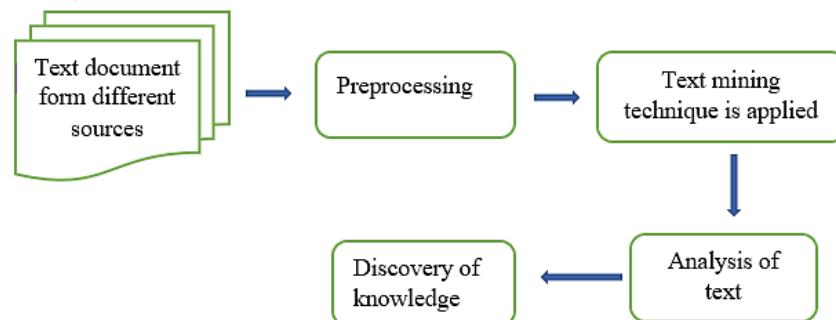

Figure 1. Text mining process

An effective text mining technique is based on cluster analysis with the task of partitioning the data set into groups called clusters, which involve using descriptor, a set of words that describe the content in cluster, and extract the descriptor. The goal of the data clustering method is to group data points in a database into clusters so that data points in the same cluster have great similarities and points that are not in the same cluster have little or no similarity.

Crawford et al. (2015) presented a study using cluster analysis to group fluorescent particles for discriminating and quantifying primary biological aerosol particles (PBAPs). The results showed that an improved method by combining cluster analysis increases the accuracy in quantifying PBAP classes. Decherchi et al. (2009) suggested that in digital forensic activities, seized digital devices can provide valuable information about events or individuals in the investigation if a digital text analysis strategy, using text mining based on clustering, was introduced for investigation purposes. This method is applied experimentally with the email data set Enron - the real-world dataset suitable for the context of forensic analysis. The results show that the clusters can be exploited to get useful survey information. In 2018, to investigate the appropriateness of existing resources provided by textbooks for e-learning purposes, Lau et al. (2018) analyzed 100 textbooks that are commonly used. Cluster analysis was performed to identify clusters of main learning resources in two aspects - content complexity and ease of use. This study shows that most of the learning resources in the sampled textbook are only suitable for low-to-intermediate learning according to the revised Bloom's taxonomy. Also, cluster analysis is employed in exploiting exam results to standardize the quality of multiple-choice questions, exploiting data of customers who used hotel services.

2.2. Atlas.ti Software

Atlas.ti is a computer-assisted qualitative data analysis (CAQDAS) provides some positioning, coding, and annotation tools (Barry, 1998). It aims to help researchers systematically explore and analyze complex phenomena hidden in unstructured data (text, multimedia, geospatial), thereby consider and evaluate the importance and relationship between them (Silver, 2014). ATLAS.ti helps us easily and quickly navigate through large data sets utilized in several fields such as anthropology, art, architecture, media, educational science, management research, market research, psychology, sociology, etc. The essence of Atlas.ti is the discovery approach to formulating theory by segmentation, coding (a piece of text), and traceability with the construction of concepts and text structures. Researchers can draw realistic maps, connecting lines, describe the interconnection of concepts as a network. Also, features like compiling text units and cutting and pasting operations between the various available text windows are used to form a coherent text. Moreover, Alas.ti also supports data analysis in many formats such as text, audio, graphics, etc. It will be of great help to analyze unstructured and non-numerical data, identify topics, patterns and meanings. Atlas.ti's strengths relate to directness, image quality and space, its creativity and interconnection, visualization, and integration. Nevertheless, users also have a lot of anxiety during using Atlas.ti due to its loose, free structure.

It should be noted that Atlas.ti does not replace thinking as it cannot do intellectual work because only humans can think; the analyst is the one who has to do the real analysis. According to Konopásek (2007), Atlas.ti is a support tool "stored, recollected, classified, linked, filtered out in great numbers … and made meaningful sum". Thus it will allow researchers to think tangibly. Besides, it also supports several other utilities such as notes, comments, adjust the size, color of the text and the code as described by the analyst. From there, they have an overview for everything they think about their data. In particular, multiple documents can be added to the same project so that they can be compared and contrasted, and in a teamwork environment. Hence many developers conduct their coding and then merge and discuss, saving time and increasing work efficiency. (Hwang, 2008). The first version of Atlas.ti was developed by Thomas Muhr at the Technical University in Berlin in the context of the ATLAS project 1989-1992 (Muhr, 1991). Atlas.ti is available for Windows, Mac, Android, and iPad operating systems.

2.3. Vietnamese Algebra and Analysis Textbooks of Grades 10, 11, and 12

The project "Teaching foreign languages in the national education system between 2008 and 2020" was proposed by the Ministry of Education and Training and approved by the Government. The primary objective is that "by the year 2020, most of the young Vietnamese graduates from high schools, colleges and universities will be able to use foreign languages independently with confidence in communication, learn and work in an integrated, multilingual and multicultural environment; turning foreign languages into an advantage of the Vietnamese people, serving the cause of industrialization and modernization of the country" (Prime Minister, 2019). Grasping that spirit, textbooks published by Vietnam Education Publishing House, used nationwide, have two versions: Vietnamese and bilingual Vietnamese - English. This study uses textbooks Algebra 10, Algebra and Analysis 11, Analysis 12 Vietnamese-English bilingual editions published at Binh Dinh People Printing Co., Ltd. in 2015 (see Figure 2). We hereinafter collectively call Algebra and Analysis 10, 11, and 12. The textbooks are all uniformly structured, including the following main topics: theory, examples, exercises at the end of each lesson, exercises at the end of each chapter. Besides, at the end of the book, there are instructions for solving exercises and answers, look up the terminology table and the table of contents at the end of the book. At the end of some lessons, there are two sections: 'Further Reading' and 'Do you know?'. The contents of Vietnam general education curriculum are required to ensure comprehensive education, forming and developing necessary qualities and capabilities to meet the requirements of the country's industrialization and modernization. It is also required to focus on practice, associated with real-life, appropriate age psychology; create favourable conditions for the implementation of educational methods to promote students' activeness, self-awareness, initiative, and creativity, fostering self-study capacity.

In that view, the Mathematics program in high school is developed with specific goals of basic knowledge and skills for numbers and calculations on a real set of numbers, complex numbers, clause and set, algebraic and trigonometric expressions. The book also presents the equation (first order, quadratic, quadratic, converted to first and second, trigonometric, exponential, logarithmic), a system of equations (first order, quadratic); inequalities (first order, quadratic, converted to quadratic, exponential, logarithmic). It also includes level 1 inequalities (one hidden, two hidden); functions, limits, derivatives, primitives, integrals, and their applications. Besides, it also introduces geometric relations and some common shapes (points, lines, planes, triangles, circles, ellipses, polyhedra, rotated

circles). There are also references to displacement and homomorphism in the plane; vectors and coordinates; statistics, combinations, probabilities (Ministry of Education and Training, 2017).

The content of each textbook is shown in Table 1 and an illustrative lesson for a textbook is shown in Figure 3.

Table 1. Contents of Algebra and Analysis textbooks 10, 11, 12

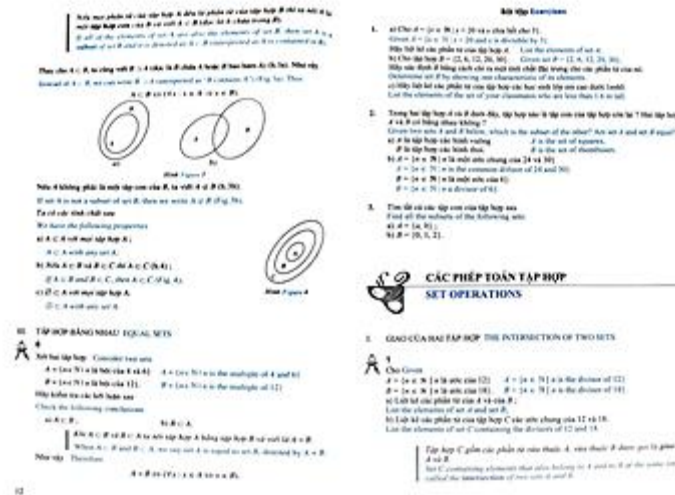| Chapter		Exercises	Number of pages
		End-of-chapter	
Algebra 10	Propositions and sets	22	17	30
	Linear and quadratic functions	12	15	21
	Equations and systems of equations	19	17	20
	Inequalities and inequations	21	17	34
	Statistics	15	11	20
	Trigonometric Arcs and angles. Trigonometric formulas	20	14	26
	End-of-term Review	8 theoretical questions	12	2
Total	**6 chapters**			
Algebra and Analysis 11	Trigonometric functions and equations	21	10	41
	Combinatorics and Probability	31	15	37
	Sequences: Arithmetic sequences and geometric sequences	21	19	31
	Limits	21	15	35
	Derivatives	24	13	37
	End-of-term Review	18 theoretical questions	20	4
Total	**5 chapters**			**191**
Analysis 12	Applications of differentiation in assessing and graphing functions	27	17	47
	Power functions, exponential functions, and logarithm functions		15	44
	Antiderivatives-Integrals and their applications		13	37
	Complex numbers			16
	End-of-term Review	10 theoretical questions	16	4
Total	**4 chapters**	**27**	**17**	**47**

Figure 2. Lesson "Sets, Textbook Algebra 10 English, Vietnamese version"

In the next section, the method of analyzing Algebra and Analysis textbooks for grades 10, 11, 12 bilingual version with text mining technique based on cluster analysis would be presented. This study is supported by Atlas.ti software to assess the suitability of the educational program as well as the second language response to the subject.

3. RESEARCH METHODS AND RESULTS

3.1. Research methods

Using the tools provided in Atlas.ti 8.4.21, we propose the process of analyzing the qualitative data based on cluster analysis of textbooks as presented in Figure 3.
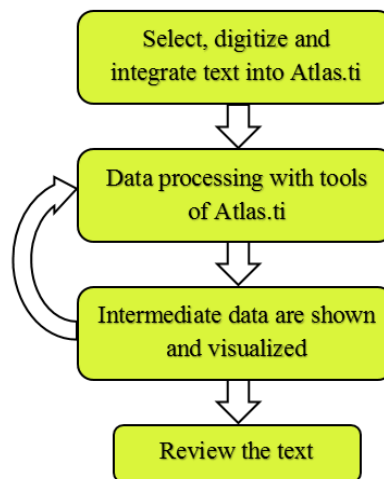

Figure 3. Process of qualitative data analysis using Atlas.ti software

In Step 1, all contents of Algebra and Analysis textbooks for grades 10, 11, and 12 of the English version are digitalised into words, including 18 files corresponding to the chapters and end-of-term reviews to facilitate analysis and tracking as well as to synthesize and correlate the topics. To optimize the time, we use ABBYY FineReader software - converting images, PDFs, etc. into words and allowing them to be modified according to the needs. Then, create a project named "SGK" in Atlas.ti, each file is included in the project using the Add Document function on the toolbar. Finally, the files were divided into three groups using the "Create Document Group" function, named Algebra 10, Algebra & Analysis 11, Analysis 12, respectively.

In Step 2, word processing using various text analysis techniques is proposed by Salloum et al. (2018) including Clustering, Association rule, Visualization, and Terms frequency using Atlas.ti's available tools 8.4.21. An intermediate data obtained is shown and visually retrieved in Step 3. According to Gupta & Lehal (2009), the

techniques need to be repeated until the information is extracted. Finally, the intermediate data in Step 3 becomes the input of Step 2 for the next round. This process ends when the extracted data satisfies the analyst's requirements.

The techniques of text analysis in step 2 are as follows. Firstly, use the Word List and Word Cloud tools for each file to get an idea of the frequency of all words used in each chapter. The main purpose is to make it easy to grasp the topic of the chapter as well as to assess the suitability for the goal of the textbook program by chapter. On the toolbar, click and select Word List/Word Cloud in Documents (see Figure 4). Also, Atlas.ti supports exporting words in tabular form, on the toolbar, select Export (Excel) in Documents to convert the frequency table received to an excel file with the frequency of occurrence of words from low to high so that it is easy to see the topics and issues around.
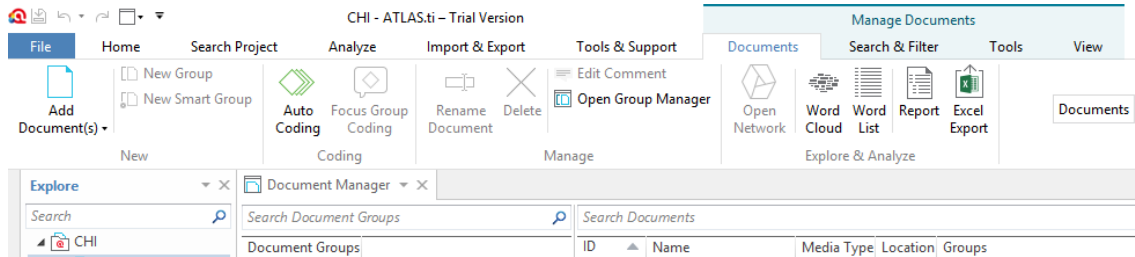

Figure 4. Toolbar of Atlas.ti

Next, read each text file, cluster the texts using the K - mean algorithm (Irfan et al., 2015) and encode specialized terms around the topic being analyzed (note, customize colors, dimensions as needed) as in Figure 6. Based on encrypted data, create networks in each topic to visually examine the relationship of words as well as the frequency of these words in the text. Go to Home / New Entities / New Network to create a diagram showing the links of Mathematics vocabularies.

Finally, select one or more citations as goals, double click on another target, drag and drop to the target citation to create the link. When a list of contacts is displayed, select a relationship to link the two citations (Figure 7) or we can also create custom relationships by selecting "Open Relation Editor". According to the results that assess the alignment and level of use of specialized terms.
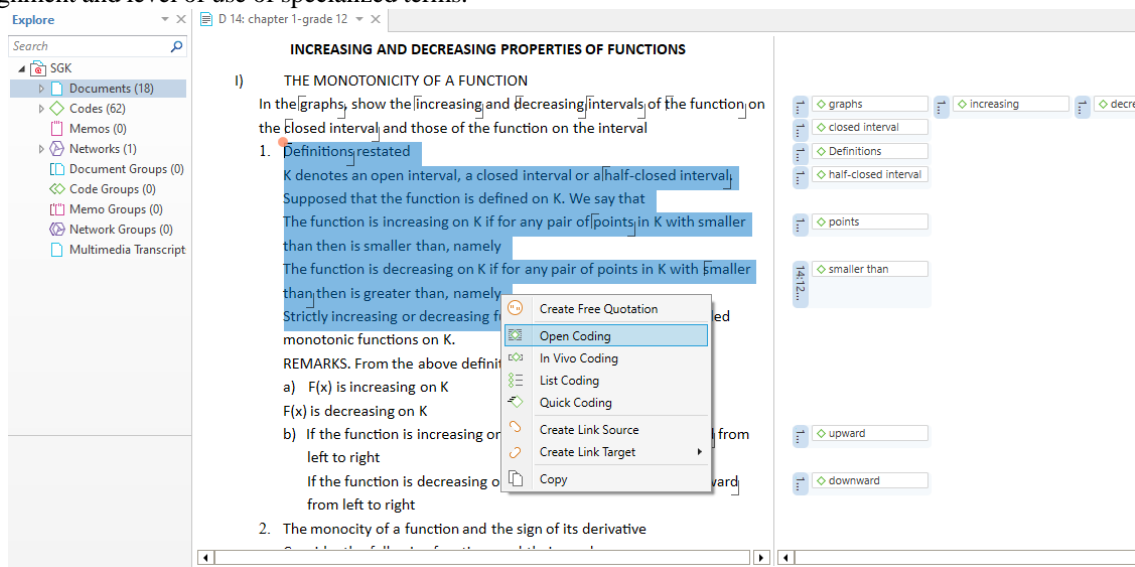

Figure 5. Coding vocabulary

In Step 4, we summarize the performance results for each class and the entire Mathematics program in high school, consider the relationships between chapters in a class and the relationships between classes, bring focused topics around the content of the whole high school algebra program; in particular, consider the effectiveness and challenges of English math in developing English. The above results will be presented in Section 3.2.
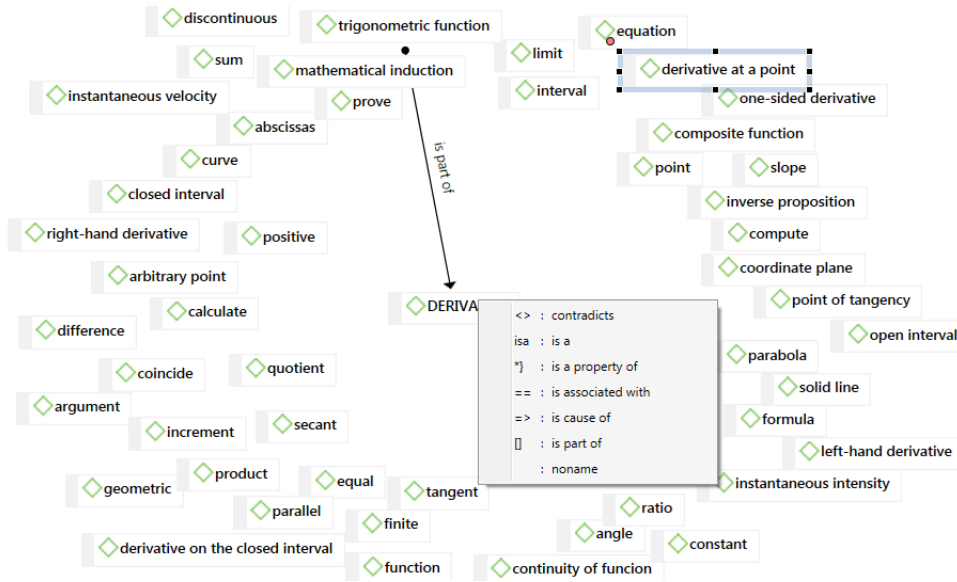
Figure 6. Creating Network

3.2. Research results

After conducting an analysis using Atlas.ti 8.4.21 to analyze grades 10, 11, and 12 bilingual math textbooks, some initial results are shown below.

In the Algebra 10 textbook, Word List and Word Cloud functions show the frequency of prominent words including "frequency" (154 times), "solution" (163 times), "inequation" (153 times), "function" (141 times), (Figure 7-a). In Word Cloud, words from large to small sizes are arranged based on the frequency with different colors to create visualization. From this, it can be seen that the Mathematics program 10 focuses mainly on inequalities and inequations. As for Algebra and Analysis 11 textbooks (Figure 7-b), words appear with high frequency are "function" (383 times), "sequence" (230 times), "equation" (158 times), "solution" (132 times); In Analysis textbook 12 (Figure 7-c), "function" (348 times), "graph" (161 times), "interval" (131 times), etc. The huge number and great frequency of specialized vocabularies show that to study Mathematics in a bilingual program, it is necessary to study and master specialized vocabulary.

The network results of the vocabulary in Chapter 5, Algebra and Analysis 11 – Figure 8, show that each branch is a word that revolves around the main topic of the chapter and has a high frequency throughout the chapter. The central word in the network is "derivative", the problems revolve around "interval", "derivative at a point", "point of tangency", "derivative on the closed interval", "continuity function". The network shows the relationship between words and the main topic of the chapter with a diagram that provides a more general view of the relationship between them.

Through the summary of the frequency of mathematical terms of textbooks through excel (Figure 9), it can be seen that the frequency of the term "function" is the largest (598 times), followed by the term "Equation" (461 times). It can be said that this is a cross-cutting theme in the high school education program, with the aim of consolidating and improving knowledge through each school year.
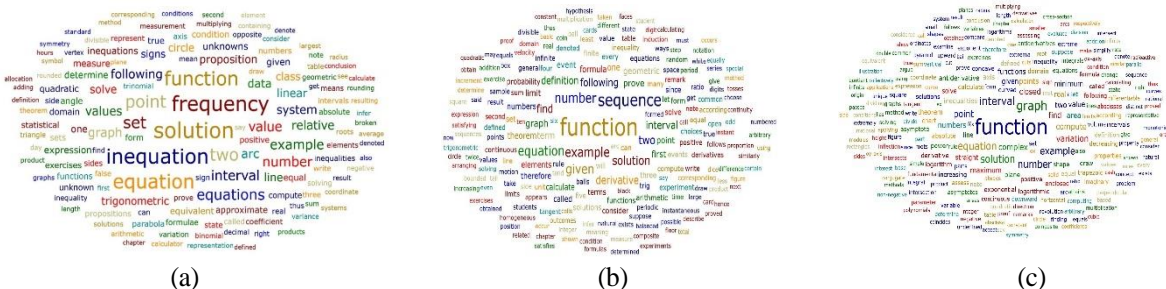

| (a) | (b) | (c) |

Figure 7. Word Cloud results for textbooks (a) Algebra 10, (b) Algebra and Analysis 11, and (c) Analysis 12

As for vocabulary in Chapter 2 Algebra 10 - Figure 9, the branches revolve around two main topics about "linear function" and "quadratic function". Some phrases revolve around a topic and have phrases related to both, such as "the highest point of a graph" and "vertex of parabola" appearing in the topic "quadratic function" and clusters "straight line" and a line "revolve around the topic of linear function", which appear on both topics such as "odd functions", "variation trend". The network shows the relationship between words and the main topic of the chapter by schematic giving a more general view of the relationship between them.
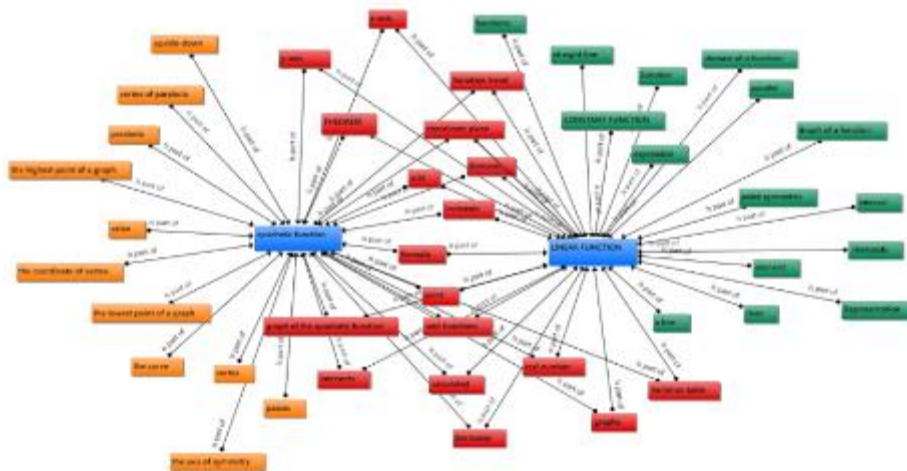

Figure 8. Network results of Chapter 2, Algebra 10

Through the frequency summary of mathematical terms of the textbooks through excel (Figure 10), it can be seen that the frequency of the term "function" is the largest (598 times), followed by the term "equation" (461 times). It can be said that this is a cross-cutting topic in the high school education program, developed in the direction of consolidating and improving knowledge through each school year.

	Word	Total	
1	Word	Total	
2	Variation	30	
3	Statistical	36	
4	expression	56	
5	Arithmetic	59	
6	Proposition	62	
7	Limit	66	
8	complex	68	
9	Inequality	79	
10	line	87	
11	geometric	89	
12	graph	158	
13	frequency	161	
14	set	171	
15	interval	189	
16	inequation	194	
17	sequence	230	
18	example	258	
19	Solution	376	
20	Equation	461	
21	Function	598	

Figure 9. Summary of the frequency of specialized vocabulary for Algebra and Analysis 10, 11, 12

Through the analysis results of the grade 10, 11, and 12 bilingual math textbooks, the focused contents are clearly shown with the relationships illustrated visually, providing a view general overview as well as the understanding of the core factors for the entire high school algebra program.

4. DISCUSSION AND CONCLUSION

This study highlights the importance of analyzing textbooks and conducting analysis using Text Mining techniques with the support of Atlas.ti software 8.4.21, thereby presenting the analysing steps using textbook analysis as well as introducing features and how to use Atlas.ti. Regarding English - Vietnamese bilingual textbooks, besides the role of a regular textbook, it also provides a large range of vocabulary and sentence structures to facilitate learners' access to English as well as English for special purposes.

Based on the relationship of Mathematics vocabulary around the topic of a chapter shown in the network, it can be concluded that learning English vocabulary for Mathematics is essential and meaningful for Mathematics college students and high school students to have access to the bilingual program of the Ministry of Education and Training. The results of this research can be considered as a reference for effectively building the curriculum as well as learning for students.

The text mining technique has been widely utilized in scientific research with numerous qualitative analysis softwares that have features similar to Atlas.ti such as HyperResearch, CI-SAID, Nvivo, etc. Therefore, it is recommended to further examine the use of these softwares as well as conduct deeper subject analysis for entire textbooks.

Moreover, Atlas.ti's features of segmentation, annotation, commenting, network creation, archiving, etc. offer a user-friendly technological support. Hence, another research direction is to use this software to support the teaching and learning process. To look further beyond, due to the importance of specialized English, research and development of an application for specialized English learning is also a necessary and significant research direction.

REFERENCES

Barry, C. A. (1998). Choosing qualitative data analysis software: Atlas.ti and Nudist compared. *Sociological research online*, *3*(3), 16-28.

Camillo, F., Tosi, M., & Traldi, T. (2005). *Semiometric approach, qualitative research and text mining techniques for modeling the material culture of happiness*. Berlin, Germany: Springer.

Chall, J. S. (1997). *An analysis of textbooks in relation to declining SAT scores*.

Crawford, I., Ruske, S., Topping, D. O., & Gallagher, M. W. (2015). Evaluation of hierarchical agglomerative cluster analysis methods for discrimination of primary biological aerosol. *Atmospheric Measurement Techniques*, *8*(11), 4979.

Decherchi, S., Tacconi, S., Redi, J., Leoncini, A., Sangiacomo, F., & Zunino, R. (2009). Text clustering for digital forensics analysis. In *Computational Intelligence in Security for Information Systems* (pp. 29-36). Springer, Berlin, Heidelberg.

Gaikwad, S. V., Chaugule, A., & Patil, P. (2014). Text mining methods and techniques. *International Journal of Computer Applications*, *85*(17).

Gibbs, G. R., Friese, S., & Mangabeira, W. C. (2002). *The use of new technology in qualitative research*. Introduction to Issue 3(2) of FQS. In Forum Qualitative Sozialforschung/Forum: Qualitative Social Research (2002).

Gupta, V., & Lehal, G. S. (2009). A survey of text mining techniques and applications. *Journal of emerging technologies in web intelligence*, *1*(1), 60-76.

Hearst, M. (2003). *What is text mining*. SIMS, UC Berkeley.

Hong, C. F. (2009). Qualitative chance discovery - Extracting competitive advantages. *Information Sciences*, *179*, 1570-1583.

Hwang, S. (2008). Utilizing qualitative data analysis software: A review of Atlas.ti. *Social Science Computer Review*, *26*(4), 519-527.

Irfan, R., King, C. K., Grages, D., Ewen, S., Khan, S. U., Madani, S. A., … & Tziritas, N. (2015). *A survey on text mining in social networks*. The Knowledge Engineering R.

Ivić, I., Pešikan, A., & Antić, S. (2013). *Textbook Quality. A Guide to Textbook Standards*. Göttingen: V&R Unipress.

Jalali, S. M. J., Park, H. W., Vanani, I. R., & Pho, K. H. (2020). *Research trends on big data domain using text mining algorithms*. Digital Scholarship in the Humanities.

Janasik, N., Honkela, T., & Bruun, H. (2009). Text mining in qualitative research: Application of an unsupervised learning method. *Organizational Research Methods*, *12*(3), 436-460.

Konopásek, Z. (2007). *Making thinking visible with Atlas. ti: Computer-assisted qualitative analysis as textual practices*. Historical Social Research/Historische Sozialforschung. Supplement, 276-298.

Korhonen, A., Silins, I., Sun, L., & Stenius, U. (2009). The first step in the development of text mining technology for cancer risk assessment: identifying and organizing scientific evidence in risk assessment literature. *BMC bioinformatics*, *10*(1), 303.

Krippendorff, K. (2004). *Content analysis: An introduction to its methodology*. Thousand Oaks, CA: Sage Publications.

Lau, K. H., Lam, T., Kam, B. H., Nkhoma, M., Richardson, J., & Thomas, S. (2018). The role of textbook learning resources in e-learning: A taxonomic study. *Computers & Education*, *118*, 10-24.

Lebrun, J., Lenoir, Y., Laforest, M., Larose, F., Roy, G. R., Spallanzani, C., & Pearson, M. (2002). Past and current trends in the analysis of textbooks in a Quebec context. *Curriculum Inquiry*, *32*(1), 51-83.

Mahmoudi, M. R., & Abbasalizadeh, A. (2019). How statistics and text mining can be applied to literary studies?. *Digital Scholarship in the Humanities*, *34*(3), 536-541.

Ministry of Education and Training (2017). *High School Education Program*. Retrieved from https://moet.gov.vn

Muhr, T. (1991). ATLAS/ti-A prototype for the support of text interpretation. *Qualitative Sociology*, *14*(4), 349-371.

Okeeffe, L. (2013). A framework for textbook analysis. *International Review of Contemporary Learning Research*, *2*(1), 1-13. http://dx.doi.org/10.12785/irclr/020101

Rajman, M., & Besançon, R. (1998). *Text mining: natural language techniques and text mining applications*. In Data mining and reverse engineering (pp. 50-64). Springer, Boston, MA.

Rose, J., & Lennerholt, C. (2017). Low-cost text mining as a strategy for qualitative researchers. *Electronic Journal of Business Research Methods*, *15*(1), 2-16.

Salloum, S. A., Al-Emran, M., Monem, A. A., & Shaalan, K. (2018). *Using text mining techniques for extracting information from research articles*. In Intelligent natural language processing: Trends and Applications (pp. 373-397). Springer, Cham.

Silver, C., & Lewins, A. (2014). *Using software in qualitative research: A step-by-step guide*. Sage.

The Prime Minister (2019). Decision on approving the project "Teaching and learning foreign languages in the national education system in the 2008-2020 period". Accessed from http://vanban.chinhphu.vn

Tian, Y., & Pho, K. H. (2019). *A statistical view to study the aphorisms in Nahj al-Balaghah*. Digital Scholarship in the Humanities.

Tuan, B. A., Pudikova, G. N., Mahmoudi, M. R., & Pho, K. H. (2019). *Statistical approaches in literature: Comparing and clustering the alternatives of love in Divan of Hafiz*. Digital Scholarship in the Humanities.

Tummons, J. (2014). Professional standards in teacher education: tracing discourses of professionalism through the analysis of textbooks. *Research in Post-Compulsory Education*, *19*(4), 417-432.

Wang, Y., Garjami, J., Tsvetkova, M., Huu Hau, N., & Pho, K. H. (2020). *Statistical approaches in literature: An application of principal component analysis and factor analysis to analyze the different arrangements about the Quran's Suras*. Digital Scholarship in the Humanities.

Witten, I. H., Don, K. J., Dewsnip, M., & Tablan, V. (2004). *Text mining in a digital library*.

Yu, C. H., Jannasch-Pennell, A., & DiGangi, S. (2011). Compatibility between text mining and qualitative research in the perspectives of grounded theory, content analysis, and reliability. *The Qualitative Report*, *16*(3), 730-744.