

INTERNATIONAL JOURNAL
ON INFORMATICS VISUALIZATIONjournal homepage : www.joiv.org/index.php/joiv

Relationship between Korean Informatics Curriculum and Textbook Learning Element Considering Compound Word

Jaehong Kim^a, Hosung Woo^b, Jamee Kim^c, WonGyu Lee^{a,*}^a Dept. of Computer Science and Engineering, Graduate School, Korea University, 145 Anam-ro Seungbuk-gu, Seoul, South Korea^b Dept. of E-learning, Graduateschool, Korea National Open University, 86, Daehak-ro, Jongno-gu, Seoul, South Korea^c Major of Computer Science Education, Graduate School of Education, Korea University, 145 Anam-ro Seungbuk-gu, Seoul, South Korea

Corresponding author: *lee@inc.korea.ac.kr

Abstract— With the development of information and communication technology, countries around the world have strengthened their computer science curriculums. Korea also revised the informatics curriculum (The name of a subject related to computer science in Korea is informatics.) in 2015 with a focus on computer science. The purpose of this study was to automatically extract and analyze whether textbooks reflected the learning elements of the informatics curriculum in South Korea. Considering the forms of terms of the learning elements mainly comprised of compound words and the characteristics of Korean language, which makes natural language processing difficult due to various transformations, this study pre-processed textbook texts and the learning elements and derived their reflection status and frequencies. The terms used in the textbooks were automatically extracted by using the indexes in the textbooks and the part-of-speech compositions of the indexes. Moreover, this study analyzed the relevance between the terms by deriving confidence of other terms for each learning element used in the textbooks. As a result of the analysis, this study revealed that the textbooks did not reflect some learning elements in the forms presented in the curriculum, suggesting that the textbooks need to explain the concepts of the learning elements by using the forms presented in the curriculum at least once. This study is meaningful in that terms were automatically extracted and analyzed in Korean textbooks based on the words suggested by the curriculum. Also, the method can be applied equally to textbooks of other subjects.

Keywords— K-12 computer education; textbook analysis; Korean natural language processing; term extraction.

Manuscript received 8 Feb. 2021; revised 12 Apr. 2021; accepted 20 Apr. 2021. Date of publication 31 Dec. 2021. International Journal on Informatics Visualization is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

An educational curriculum is a systematized guideline with all plans for education. it serves as a comprehensive educational framework that guides teachers to plan courses, such as the subject contents and items, according to educational purpose. The curriculum provides what and how to teach and evaluate. Among them, learning elements of the curriculum indicate the specific knowledge to be taught.

With the development of information and communication technology, countries around the world have revised their curriculums to incorporate the learning elements of computer science to reflect international trends and social needs. The UK introduced a subject called Computing in elementary and middle schools in 2014 and included it in its regular

curriculum. Israel introduced computer science as a¹ regular subject for high schools in 1994, and added a subject called Computer Science Literacy in elementary school curriculum in 2016. Furthermore, a variety of countries, such as the United States, India, Japan, and China, have reinforced computer science curriculum, and its cycle of change has also shortened with the rapid development of computer-related technologies. The Korean government also reinforced computer science education in the 2015 revised curriculum and emphasized problem-solving in real life by using basic concepts, principles, and technologies of computer science in the informatics subject [1]. The Informatics subject, which only few students took selectively, became mandatory in elementary and middle schools in the new curriculum. The Korean government applies one nationally standardized

¹ This thesis was written based on Jaehong Kim's master's thesis.

curriculum to all schools. In other words, the revision of the informatics curriculum made it essential for all Korean students to learn computer science [2]. In addition, the Korean government announced a new curriculum revision in 2022 and is developing AI-related subjects [3],[4]. Therefore, it is expected that the informatics curriculum including artificial intelligence is emphasized once again.

Although the Korea government operates the nationally standardized curriculum, the textbook accreditation system allows multiple textbooks. Private companies can author, develop, and publish textbooks, and schools can use textbooks that have gone through the accreditation system. Textbook developers can publish textbooks using their own interpretations as long as the textbooks include the contents presented in the curriculum. Therefore, there are currently 17 informatics textbooks for middle schools and 9 for high schools in Korea, and each school can freely select one of these textbooks for education.

Although the core contents of different textbooks are the same, the expression, terminology, and volume of the contents may vary. In this regard, the current study aimed to examine whether Korean high school textbooks reflect the learning elements of the curriculum, to extract which terms other than the learning elements are used in the textbooks, and to analyze their relevance to the learning elements.

Textbook content analysis studies have mainly used the content analysis method in which researcher read and analyze the contents of the textbooks firsthand. It has the advantage of being able to analyze both the manifest and latent contents of the research subjects in the textbook contents. For example, a study analyzed codes in computer science textbooks and teaching characteristics of teachers, finding that the code style of the textbooks influenced the teachers [5][6].

Romey analysis method analyzes the exploratory tendencies of materials in textbooks. It examines whether the

elements of the textbooks are described to inspire students' intellectual curiosity. While the content analysis method and Romey analysis method can analyze all contents of the textbooks, the two methods both require researchers to read and analyze data firsthand, which takes a lot of time. Also, the researchers have to use their own discretion to quantify and classify the data, which may decrease objectivity [7].

The data mining method has also been conducted for the analysis of educational materials. It has the strengths of being able to quantitatively and automatically process unstructured data such as textbooks, and to exclude the subjectivity of researchers from the research process. Related studies analyzed whether the US and UK curriculums were applied to its actual operation [8] or textbooks [9] by using the topic modeling technique. A study on the AI curriculum in India analyzed the progress of discussions on the AI education by using the topic modeling technique of the AI education symposium in the United States [10]. A study in Korea analyzed the relevance of Korean information textbooks by using the frequency analysis and word-to-word association analysis techniques.

Research on information science, information systems, and technology management is being actively conducted using text mining[11]. Text mining, which processes information in a quantified and automated way, can maintain objectivity and takes less time. However, it requires researchers' interpretations of the results. Also, the type of natural language has a great influence on its performance. Due to the complex structure of Korean language and its versatile word transformations, it is difficult to classify Korean language by using text mining and natural language processing.

II. MATERIAL AND METHOD

TABLE I
LEARNING ELEMENTS OF KOREAN HIGH SCHOOL INFORMATICS CURRICULUM

Area	Learning Elements		
Information Culture	<ul style="list-style-type: none"> information science information occupation group software copyright 	<ul style="list-style-type: none"> information sharing inforamtion security cyber ethics 	<ul style="list-style-type: none"> Information protection system and method (→ information protection system, information protection method),
Data and Information problem-solving and programming	<ul style="list-style-type: none"> encoding big data problem analysis current state target state core element extraction problem decomposition modiling sequential structure selection structure iterative structure 	<ul style="list-style-type: none"> data collection data analysis algorithm efficiency text-based programming environment variable data type arithmetic operation comparison operation logical operation standard input/output 	<ul style="list-style-type: none"> information visualization database file input/ouput input/ouput design overlapping control structure one-dimension and two-dimension arrays (→ one-dimension array and two-dimension arrays), function software development
Computing System	<ul style="list-style-type: none"> operation system role resource management 	<ul style="list-style-type: none"> Wired/wireless network (→wired network,wireless network) IP adress 	<ul style="list-style-type: none"> physical computing system design and implementation (→ physical computing system design, physical computing system implementation)

The current study used text mining in order to maintain objectivity and automatically analyze the contents of rapidly changing informatics textbooks.

The research subjects included the learning elements of high school information curriculum and 4 informatics textbooks. There were 41 learning elements in 4 domains in the curriculum, as shown in Table 1. For automatic extraction and analysis based on the learning elements, the following pre-processing was performed.

A. Pre-processing of the learning elements and textbook texts

For the automatic processing of the learning elements and textbooks, two types of pre-processing were needed. The first was to separate two learning elements if they were presented together in the curriculum. For example, 'one-dimensional and two-dimensional arrays' in the domain of problem-solving and programming were divided into 'one-dimensional array' and 'two-dimensional array'. Since a total of 4 learning elements were subject to this type of pre-processing, the number of the learning elements for the analysis of this study was 45. In Korean language, affixes are attached to stems of agglutinating words to form words, and their meaning and grammatical function can greatly vary and determined according to the context [12][13]. Furthermore, unlike English, Korean words divided into sub-words may not use spaces and be composed of several morphemes [13][14]. Such characteristics complicate the law of spacing and notation in Korean. In particular, computer science terms that frequently use foreign and technical words are often compound words including spaces. For example, the term 'big data' in the domain of data and information was presented with a space in the majority of the textbooks analyzed in this study, whereas

one textbook used the term 'bigdata' without a space. Since it is not that a textbook using the different law of spacing does not reflect the contents presented in the curriculum, this study deleted all the spacing of the learning elements and texts in the textbooks.

Among the 45 learning elements, 30 learning elements had spaces. Table 2 shows whether the learning elements used the form presented in the curriculum (excluding space notation). Among the 45 learning elements, 13 learning elements were never mentioned in more than one textbook. By domain, 3 were in the domain of information culture (38%), 2 were in the domain of data and information (33%), 5 were in the domain of problem-solving and programming (21%), and 4 were in the domain of computing systems (57%). The domain with the highest non-reflection rate was computing system, and all of the two learning elements related to physical computing that was newly added to the curriculum were not reflected in the textbooks, indicating that the terms used in the textbooks were not coherent.

B. Reflection status of the learning elements considering compound word forms

In general, word frequency or search is based on a single word consisting of one morpheme, not a compound word. However, in order to consider the characteristics of the learning elements with many compound words, this study used the following two criteria.

- Calculating the frequency of reflecting the learning elements at the sentence level.
- Searching compound words by separating them based on spaces, and including them into frequency if all of.

Applying the above two criteria to the analysis could identify whether the learning element semantically reflected the compound words even without their literal forms.

For instance, the learning element of 'operating system role' was difficult to be applied directly to sentences. On the other hand, "Operating system is software that have a role as an intermediary between a computer's hardware and a user so that it efficiently operates and manages the resources of the computing system and enables the user to conveniently use the computer."

Although this sentence describing the role of the operating system directly reflected the learning element of "operating system role", it was hard to calculate the frequency of whether this was a sentence reflecting the learning element due to the separate usage of each word. Nevertheless, the application of the above two criteria allowed the examination of whether the words 'operating', 'system', and 'role' all existed in this sentence, and the frequency calculation of the compound words. Therefore, it enabled the examination of whether the not-reflected learning elements in Table 2 were reflected in a space-separated state.

C. Extraction of terms other than the learning elements in textbooks

Among the words in the index term group, the study extracted the part-of-speech compositions of the compound words. In order to analyze the part-of-speech compositions of Korean, the study classified the part-of-speech by using the

TABLE II
NUMBER OF TEXTBOOKS NOT REFLECTED BY LEARNING ELEMENT

Area	Learning Elements	Number of textbooks not reflected
Information Culture	information science	4
	occupation group	
Data and Information	information security system	2
	information protection method,	4
	encoding	2
	Information visualization	2
problem-solving and programming	text-based programming environment	4
	input/output design	4
	algorithm efficiency	3
	overlapping control structure	2
	Software development	2
Computing System	Operating system role	4
	physical computing system design,	4
	physical computing system implementation	3
	wired network	1

pos function of the Okt model of the Korean natural language processing package called Konlpy.

Using the Okt model, the part-of-speech compositions of the index terms in the form of compound words were extracted. Then, word combinations with the extracted part-of-speech compositions were extracted. As data accumulates on wikipedia, various studies using the data are being conducted[15],[16],[17]. In this study, it was used to determine whether the text extracted through wikiapi is an existing term. After examining whether there was a Wikipedia page for the extracted word combination using wikiapi, the corresponding terms were extracted. Table 3 shows the part-of-speech compositions extracted from the indexes and their terms, which includes 235 terms of a total of 4 part-of-speachs. Thus, the four textbooks included 45 learning elements for textbook analysis, 400 index terms, and 235 extracted terms.

Relationship between the learning elements and other

In order to analyze the relationship between the terms extracted in the previous process and the learning elements, confidence of the association rule analysis was used. confidence refers to the probability that a specific term appears when another keyword appears[18]. The formula is (1).

$$\text{confidence}(A \rightarrow B) = \frac{P(A,B)}{P(A)} \quad (1)$$

$P(A)$ = A probability of the word A appearing in a sentence
 $P(A, b)$ = A probability of the words A and B appearing simultaneously in a sentence

The difference between confidence and the frequency of simultaneous occurrence of two words is its direction. Since it calculates the probability of occurrence of the keyword B when the keyword A appears in the sentence, the frequency of occurrence of the keyword B in case of the not-appearance of the keyword A does not affect the confidence. To analyze the relationship between terms used together based on the learning elements, its confidence was calculated by applying the keyword A as the learning element.

III. RESULTS AND DISCUSSION

In order to identify whether the 4 textbooks following the information curriculum of Korean high schools reflected its learning elements, the current study considered the characteristics of the learning elements in the compound word forms

A. Reflection Status of the learning elements considering the compound word forms

When removing only the spaces of the texts and learning elements of the textbooks to examine the reflection status, it was found that 14 out of the 45 learning elements were not reflected in the textbooks. Since Korean uses various propositions and suffixes, it is difficult to search for compound words. To address this difficulty, the current study segmented compound words based on spacing and analyzed whether the segmented elements were included in the sentence unit. Table 4 presents the results. When examining the reflection status in consideration of the compound word forms, 8 learning elements, out of the 14 learning elements with any unreflected textbook, were reflected in all textbooks. 4 of the 6 learning elements (information science occupation

TABLE III
PART-OF-SPEECH COMPOSITIONS OF COMPOUND WORDS IN INDEXES

the part-of-speech composition in the index	The number of terms in the indexes	Terms extracted from the textbooks using the part-of-speech compositions	Number of extraction terms
Noun + Noun	140	vector image, file format, background music, output device, storage device, hard disk, Google spreadsheet., etc	188
Noun + Noun+ Noun	32	autonomous driving vehicle, Macintosh operating system, e-mail address, etc.	40
Noun + affix	8	-	0
Noun + affix + Noun	5	-	0
Noun + Noun+ Noun+ Noun	5	distributed denial of service, Point of sale information management, spring summer autumn winter, crime prevention environment design	4
Alphabet + Noun	4	-	0
Number + Noun + Noun+ Noun	4	-	0
Noun + Position +Noun	3	-	0
Noun+Postposition	3	-	0
Number + Noun + Noun	2	-	0
Noun + position + Noun + Noun	2	-	0
Modifier+ Noun	2	-	0
Noun + Noun + position	2	-	0
Noun + Noun + affix	2	-	0
Noun + Verb	2	Doxing, file reading, sit-ups	3
		total	235

group, input/output design, physical computing system design, and physical computing system implementation) that still had unreflected textbooks reduced the number of those textbooks.

The learning elements with the non-reflecting textbooks in spite of the consideration of the compound word forms included information science occupation group, coding, input/output design, physical computing system design, and physical computing system implementation. As a result of analysis of the non-reflecting textbooks directly, the words, 'information science occupation' and 'information science occupations', did not use affixes or used different affixes. The word "coding" of the data and information fields was not mentioned in the two textbooks, and was instead expressed as "encoding" that imported the pronunciation of a foreign language.

The term 'input/output design' in the domain of problem-solving and programming was reflected in the three textbooks not in a direct way but in an indirect way as the conceptual explanation and practice of file input/output or standard input/output. One textbook expressed 'superimposed control structure' only as 'superimposed selection structure' and 'superimposed conditional statement', respectively.

The domain of computing system was mainly used by subtracting the term "system" from the "physical computing system". In addition, sentences using various synonyms such as 'composition' instead of 'design' and 'composition' instead of 'implementation' were found.

In other words, all of the four textbooks mentioned the learning elements suggested by the curriculum in terms of contents, however, some textbooks did not reflect them in a presented form.

B. Relationship between the learning elements and terms

In order to analyze what contents were used in the textbooks along with the learning elements, the current study collected and analyzed 4 textbooks by domain. Based on the collected text sentences, the learning elements and confidence using the the learning elements and confidence and extracted terms were examined. Table 5 shows each learning element in the domain of information culture and the top 7 terms with high confidence.

The red cells in Table 5 indicate the learning elements within the same domain. The terms highly relevant to the learning element were mainly the learning elements in the same domain. The top 7 terms made it possible to guess what contents each learning element contained in the textbooks. For instance, it was assumed that the "information science occupation groups" would include the introduction and exploration of information science occupations related to information society or information science technology.

The examples of information science occupations included security experts at the end. The lower rankings had various occupations, such as virtual reality experts, network experts, robotics, web designers, and programmers. Because each textbook had different occupations to be explained or illustrated, they were assumed to be in the lower order. Even in terms of the 'information protection method', although not shown in this table, information patent law, GPL, vaccine program, firewall, and secure password were placed with lower confidence.

Table 6 shows the top 7 confidence terms in the domain of computing system. The yellow cells represent the learning elements in the same domain; the table showed that the relevance between the learning elements in the same domain was lower than the domain of information culture. In particular, the learning elements used crossly between the learning elements of different sub-domains were absent. This was perhaps because the physical computing sub-domain newly added to the current curriculum had fewer pages and learning elements than other domains, and had weak relevance with other domains. The sub-learning elements related to physical computing included Entry and Arduino, the programming environments using physical computing. As the textbooks with higher confidence could recognize the theoretical contents that deal with the same contents in common, such as the domain of information culture, the textbook with lower confidence confirmed explanatory examples or details of the relevant learning elements.

TABLE IV
THE NUMBER OF NON-REFLECTING TEXTBOOKS CONSIDERING THE COMPOUND WORD FORMS OF THE LEARNING ELEMENTS

Area	Learning Element	Number of textbooks not reflected	
		Remove whitespace	Segmentat ing Searching
Information Culture	information science occupation group	4	<u>2</u>
	information security system	2	0
	information protection method,	4	0
Data and Information	encoding	2	<u>2</u>
	Information visualization	2	0
	text-based programming environment	4	0
problem-solving and programming	input/ouput design	4	<u>3</u>
	algorithm efficiency	3	0
	overlapping control structure	2	<u>2</u>
	Software development	2	0
Computing System	operating system role	4	0
	physical computing system design,	4	<u>3</u>
	physical computing system implementation	3	<u>1</u>
	wired network	1	<u>0</u>

TABLE V
TOP 7 CONFIDENCE TERMS IN INFORMATION CULTURE AREA

Area	Detail Area	Learning elements	Top 7 confidence terms (confidence)						
Information Culture	Information society	information science	Science and technology (0.083)	information science and technology (0.081)	information science occupation group (0.072)	Information society (0.052)	Software (0.032)	problem (0.024)	exploration (0.019)
		information science occupation group	information science (0.192)	Information society (0.056)	Science and technology (0.044)	Software (0.044)	information science and technology (0.041)	exploration (0.033)	security expert (0.021)
	Information protection system	information protection (0.098)	Information sharing (0.082)	copyright (0.06)	Cyber ehctic (0.049)	information protection method, (0.044)	software copyright (0.044)	Software (0.044)	
		information protection method,	Information sharing (0.08)	information protection (0.069)	copyright (0.048)	Information protection system (0.043)	Software (0.037)	software copyright (0.032)	information security (0.027)
	Information Ethics	Information sharing	information protection (0.052)	personal information (0.035)	data (0.035)	problem (0.03)	network (0.024)	copyright (0.023)	information science (0.02)
		information security	information protection (0.045)	Software (0.034)	program (0.028)	Information sharing (0.028)	information science (0.025)	copyright (0.025)	security expert (0.023)
		software copyright	copyright (0.215)	Software (0.215)	program (0.034)	license (0.029)	copyright law (0.027)	copyright infringement (0.022)	Open source (0.022)
		Cyber ehctic	cyberspace (0.12)	copyright (0.044)	problem (0.04)	information protection (0.04)	Information protection system (0.036)	Information sharing (0.036)	software copyright (0.032)

TABLE VI
TOP 7 CONFIDENCE TERMS IN COMPUTING SYSTEM CULTURE AREA

Area	Detail Area	Learning elements	Top 7 confidence terms (confidence)						
Computing System	Principle of computing system operation	operating system role	operating system (0.173)	Computing system (0.094)	Resource management (0.072)	resource (0.072)	software (0.043)	hardware (0.043)	Memory device (0.043)
		resource management	resource (0.201)	operating system (0.128)	Computing system (0.108)	program (0.065)	network (0.04)	Memory device (0.03)	operating system role (0.025)
		wired/wireless network	wireless network (0.157)	network (0.157)	Computing system (0.096)	operating system (0.042)	Resource management (0.03)	resource (0.03)	wired network (0.03)
	IP address	network (0.199)	Computing system (0.141)	Domain name (0.084)	DNS server (0.058)	bit (0.052)	Wireless network (0.037)	Internet address (0.031)	
	Physical computing	physical computing system design	Physical computing (0.236)	problem (0.153)	program (0.056)	Problem solving (0.056)	Problem analysis (0.056)	physical computing system implementation (0.042)	Computing system (0.042)
physical computing system implementation		Physical computing (0.223)	Computing system (0.116)	program (0.074)	Output device (0.05)	problem (0.05)	hardware (0.05)	sensor (0.041)	

IV. CONCLUSION

Textbooks are the most easily accessible educational medium for students. The current study aimed to objectively analyze whether the contents of informatics textbooks in Korea reflected the learning contents suggested by the curriculum, and to explore the relationship between what terms were used to teach the learning contents. The conclusion and suggestions of the study are as follows.

First, consistency of expressions of the learning elements between textbooks is necessary. The 4 textbooks of the research subjects reflected all of the learning elements in terms of their contents and meanings. However, there were 14 learning elements that did not follow the form of terms suggested by the curriculum. The nationally standardized curriculum in Korea provides the same knowledge to all students. Differences in expressions of the learning elements, the core knowledge of textbooks, can pose a risk of confusion and misunderstanding to students. The textbooks need to avoid such a risk by using the the exact terms of the learning elements suggested in the curriculum at least once and then using different terms.

Second, terms for common theories including the learning elements in the same domain had higher confidence with the learning elements. When the four textbooks were collected and applied to the analysis by domain, the higher confidence terms tended to include the basic contents that were commonly explained, and the lower confidence terms included examples and details of each textbook to explain the learning elements.

Third, this study was able to extract terms used in the textbooks in a basis of part-of-speech compositions of the terms constituting indexes, verify them using wikiapi, and extract terms in compound word forms used in informatics textbooks. Due to the complex nature of Korean language composition, various propositions or affixes are not distinguished by spaces but transform into a form attached to words, which makes the accuracy of part-of-speech tagging not so great. However, this study made it possible to extract the terms regardless of the accuracy of part-of-speech tagging because the current study determined whether a word combination of the same composition as the corresponding composition existed in the text regardless of the accuracy of part-of-speech compositions. As the terms in the form of new compound words can be extracted when new definitions are uploaded on Wikipedia pages, the performance of term extraction will also increase with the number of pages.

This study has the significance of objectively analyzing textbooks by using natural language processing and text mining that takes into account the characteristics of Korean, which has a variety of word transformations. Utilizing indexes and part-of-speech compositions, the study automatically extracted the terms used in informatics textbooks. In this study, we used wikipedia to determine whether the extracted text is a term for computer science. However, wikipedia doesn't have many words in languages other than English[19]. As the Korean data on wikipedia increase over time, the term discrimination performance will naturally improve[20]. Overall method is applicable to other subjects or textbooks other than informatics, and has the generalability of automatically extracting general terms and analyzing them based on the learning elements. Also, there is

still room for improvement in this method. The relationship was described based on the frequency of two words appearing at the same time. We need a way to show the relationship between words that are more performant.

ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. 2020R1F1A1066518)

REFERENCES

- [1] Qadir, Junaid, et al, "Engineering Education, Moving into 2020: Essential Competencies for Effective 21st Century Electrical & Computer Engineers," *2020 IEEE Frontiers in Education Conference (FIE)*, IEEE, pp. 1-9 October, 2020.
- [2] Ministry of Education, "Introduction to the elementary and secondary curriculum the 2015 revised curriculum overview", *Ministry of Education.*, Seoul, South Korea, REP. 2015-74, Sep. 2015.
- [3] Ministry of Education, "2015 Curriculum (Added) Direction of Core Competencies in the Age of Artificial Intelligence", *Ministry of Education.*, Seoul, South Korea, REP. 2020-248, Sep. 2020.
- [4] *2022 College Admission System Reform and High School Education Innovation Direction*, Ministry of Education., Seoul, South Korea, 2018.
- [5] Smith, J. M, "Presenting Basic CS Concepts: A Content Analysis of AP CSA Textbooks", *Koli Calling'20: Proceedings of the 20th Koli Calling International Conference on Computing Education Research*, pp. 1-2, Novembe, 2020.
- [6] Kirk, D., Tempero, E., Luxton-Reilly, A., & Crow, T, "High School Teachers' Understanding of Code Style", *Koli Calling'20: Proceedings of the 20th Koli Calling International Conference on Computing Education Research*, pp. 1-10, November, 2020.
- [7] Jamee Kim, et al, "Inquiry trend analysis in the field of 'information society and information technology' in the middle school'information' subject.", *Journal of the Korean Industrial-Academic Technology Society*, vol. 12, no.7, pp. 3022-3029, July, 2011.
- [8] Choi Hyun-jong, "Analysis of the core concepts of the'Problem Solving and Programming' section of the 2015 revised middle school information textbook", *Journal of the Korean Digital Contents Society*, vol. 21, no. 1, pp. 63-70, Jan, 2020, DOI. 10.9728. /dcs.2020.21.1.63
- [9] Sekiya, T., Matsuda, Y., & Yamaguchi, K, "Analysis of computer science related curriculum on LDA and Isomap". In *Proceedings of the fifteenth annual conference on Innovation and technology in computer science education*, June, 2010, pp. 48-52.
- [10] HoSung Woo, JaeHong Kim, JaMee Kim, WonGyu Lee, "Exploring the AI Topic Composition of K-12 Using NMF-based Topic Modeling", *International Journal on Advanced Science, Engineering and Information Technology*, vol. 10, no. 4, 2020, pp.1471-1476.
- [11] Jung, H., & Lee, B. G, "Research trends in text mining: Semantic network and main path analysis of selected journals", *Expert Systems with Applications*, vol. 162, December, 2020, DOI. 10.1016/j.eswa.2020.113851
- [12] Akimov, M., Loginova, E., & Musin, M, "A Graph-based Approach for Learner-tailored Teaching of Korean Grammar Constructions", In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, IEEE., November, 2018, pp. 349-354.
- [13] Lee, H., & Song, J, "Understanding recurrent neural network for texts using English-Korean corpora.", *Communications for Statistical Applications and Methods*, vol. 27, no. 3, 2020, pp.313-326
- [14] Jin, G., & Yu, Z, "A Korean named entity recognition method using bi-LSTM-CRF and masked self-attention", *Computer Speech & Language*, 65, 101134, 2021.
- [15] Macdonald, E., & Barbosa, D, "Neural Relation Extraction on Wikipedia Tables for Augmenting Knowledge Graphs.", In *Proceedings of the 29th ACM International Conference*, pp. 2133-2136, October, 2020
- [16] Mirza, A., Nagori, M., & Kshirsagar, V. "Constructing Knowledge Graph by Extracting Correlations from Wikipedia Corpus for Optimizing Web Information Retrieval", In *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE., July, 2018, pp. 1-7.

- [17] Zhou, Y., & Xiao, K, "Extracting prerequisite relations among concepts in wikipedia", In *2019 International Joint Conference on Neural Networks (IJCNN)*, IEEE., July, 2019, pp. 1-8.
- [18] Kjin Jeon, "Technology management research topic network analysis and R&D commercialization performance evaluation model development." Domestic master's thesis Seoul National University of Science and Technology, 2017.
- [19] Wu, Tianxing, et al., "Knowledge graph construction from multiple online encyclopedias.", *World Wide Web*, vol. 23, no. 5, September, 2020, pp.2671-2698, DOI. 10.1007/s11280-019-00719-4
- [20] Cheon, J., & Ko, Y, "Parallel sentence extraction to improve cross-language information retrieval from Wikipedia", *Journal of Information Science*, 2021, DOI. 10.1177/0165551521992754.