Journal of Theoretical and Applied Information Technology
31st August 2022. Vol.100. No 16

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

BIG DATA-BASED SW EDUCATION CURRICULUM RECOMMENDATION PLATFORM DEDIGN FOR LEARNERS

JI HOON SEO[1]

[1]Institute of General Education, Incheon National University, Incheon 22012, Republic of Korea

E-mail: [1]jihoon@inu.ac.kr

ABSTRACT

The current educational infrastructure environment is trying to change in line with the 21st century. As the importance of computational thinking, software, and artificial intelligence education gradually increases in the trend of using subject-centered curriculum, experience-oriented curriculum for learners is spreading. Since a general curriculum do not consider basic knowledge and difficulty of individuals but run the education by following the setup, it has limitations in realizing high satisfaction of learners and providing an effective education. Accordingly, this study uses online self-diagnostic data for learners based on AI education, which is the core of the future curriculum, to diagnose individual competency and difficulty, and develops a curriculum recommendation system that recommends personalized learning contents according to personal level of the learner.

Keywords: *Education System, Opinion mining, Recommendation system, Learning care platform.*

1. INTRODUCTION

The domestic educational infrastructure provides an efficient curriculum to learners based on various overseas advanced education model cases, and is attempting new changes through gradual reorganization of the curriculum [1]. In the era of the industrial revolution of the 19th century, which caused a transformation of the overall economic structure in the past, the focus was on nurturing talents to increase industrial productivity in a dense environment. However, in modern society, various technologies have been derived through scientific progress and discovery, and as a result, new studies have emerged, suggesting changes suitable for future education in the educational environment of the 21st century.

Current domestic education introduced smart education based on e-learning system, system education to nurture convergence-type talents, and coding education to develop problem-solving abilities to meet the demand for these changes. Such education is a newly emerged discipline, and when applying the general learning model pursued in the past, there are various constraints and difficulties in enhancing the learning effect of the learner. To improve this, various studies such as effectiveness analysis of educational performance are being conducted [2][3]. In other words, in the case of software and AI education, which are currently being emphasized, when applying the subject-centered curriculum and learning model that have been widely used in the past, customized education cannot be provided to learners [4][5][6]. It can be a chain factor that lowers the learning effect and the interest of the learner. Therefore, newly emerging academic or non-specialized programs require an experience-oriented curriculum and a new learning model for learners depending on whether that is their major or not. In addition, as the COVID-19 virus is rapidly spreading, the World Trade Organization (WTO) has declared a global pandemic, and major changes have begun to occur in all daily life, including the economy. Such changes have the greatest impact on education policies, and in order to minimize the spread of infection between instructors and learners, non-face-to-face classes are recommended, limiting the way to communicate smoothly with each other.

Therefore, this study builds an online platform centering on learners receiving AI education in order to improve these problems. By collecting self-diagnostic data of learners, analyzing difficulties for individual learners and difficulty level by individual competency according to the curriculum, and diagnosing problems, it searches a system that can be recommended for an individual's customized AI education curriculum according to the learner's field of interest and competency in artificial intelligence. Such a system can maximize the learner's performance using a personalized curriculum by

adopting an experience-oriented curriculum for learners. In addition, it can safely and effectively analyze learners' counseling data in a non-face-to-face environment through the online self-diagnostic recommendation system to clearly understand the counseling query and increase the accuracy of the curriculum recommendation system. Thus, it is expected to improve the quality of the future education environment in preparation for the 4th industrial revolution by providing support to learners to learn effectively through the curriculum recommendation system.

2. RELATED WORKS

2.1 Sentiment Dictionary

Sentiment dictionary means a corpus required for opinion mining analysis. Opinion mining, also called reputation analysis, is a study that analyzes whether the text presented in the document is positive or negative by deriving the positive and negative degrees of words from text documents, which are unstructured data [7][8].

Therefore, a sentiment dictionary is required for opinion mining, and the data in the sentiment dictionary is classified into positive and negative words and tagged. To build the sentiment dictionary, it takes a considerable amount of time because meaningful words are extracted based on a large amount of text data and undergoes a pre-processing process. In addition, since each language has a unique position of part-of-speech, it is a field that requires a lot of research. In the case of overseas, "SentiWordNet" is used for sentiment analysis, but in this study, sentiment data based on Korean grammar was constructed in consideration of the learner's target [9].

2.2 Text Mining Technology

The text mining technique is a technology aimed at extracting and processing useful information based on technologies to process natural languages, which are unstructured data.

That is, in cases where unstructured data were collected, the repository contains both meaningful information and unnecessary information, and the unnecessary information accounts for at least 80% of the entire information. In the case of such data, more resultant values than simple information searches can be obtained such as extracting meaningful information from the vast text corpus through text mining technology, understanding connectivity with other pieces of information, and finding out the category of the text [10]. To analyze the languages used by humans and discover the information hidden in the languages, computers use massive language resources and statistical and regular algorithms.

The areas of application include document classification, document clustering, information extraction, and Document Summarization. Areas related to text mining include the technology called opinion mining or reputation analysis (Sentiment Analysis).

2.3 Opinion Mining Technology

Opinion mining is also called sentiment analysis and is interpreted broadly as natural language processing, computer language analysis, and text mining. The theory of opinion mining as such has been gradually developed from the early 2000s, and has been extensively studied through the analyses of reputation and reviews among consumers conducted in e-commerce. Seol Yong-soo (2013) defined that texts in modalities can be used to recognize the immanent emotions of humans that are not exposed to the outside. Kim Yu-sin (2012) collected stock-related news data published on portal sites for three months and analyzed the data. He gave weights to the analyzed vocabularies to extract positive, neutral, and negative data values and predicted the amplitudes of stock price rises and falls based on two news data.

2.4 Big data processing analysis Technology

In the case of big data analysis, since the amount of data that must be processed is huge and the proportion of unstructured data is high, the complexity is shown to be high correspondingly. Most analysis techniques improve their algorithms to fit large-scale data processing based on techniques already used in the field of data mining, including statistics, and apply the algorithms to big data processing analysis.

Currently, those analysis technologies that are used to process vast amounts of data in real time include text mining, which extracts meanings contained in un- structured sentences and establishes assumptions among those pieces information that have been extracted, opinion mining, which is reputation index analysis that discriminates users' opinions on certain services and products, social network analysis that grasps those users (Influencers) who are the center of the word of mouth, and cluster analysis, which analyzes the dissimilarities between objects belonging to the target group with high similarities and objects

belonging to other clusters to derive new user groups.

2.5 Recommendation System

The recommendation system refers to a system that recommends information items suitable for a corresponding user by using information filtering. For the recommendation system, it has a similar context to decision-making support system but with different meanings. The decision-making system analyzes large-capacity data, extracts the knowledge required for decision-making, and provides it, and mainly supports decision-making to cope with organizational risks.

On the other hand, the recommendation system recommends an algorithm based on a user information profile such as an item that fits the user's interest, preference, or aptitude. Such a system is used in various fields and can be applied to search queries, product searches, and content recommendations tailored to each taste.

2.6 Characteristic of Domestic AI Education

The Korean AI education focuses on curriculum cases of other countries, strengthening SW and AI education as core competency for the future society. Since 2022, the Korean government has been adopting AI education as a core subject in the elementary and secondary education. Since 2020, it has been assisting AI education in training and appointment of teachers, making AI education a requisite for teaching qualifications; education colleges designated AI education as a mandatory training course while graduate schools of education are drawing its participation by opening up an AI-based convergence education. Still, the domestic environment lacks standardized models and various framework applications to perform AI education. Although it is using some AI education programs such as Python-based TensorFlow and Teachable Machine 2.0 by Google, those independent programs have limitations for learners to study in depth.

3. PLATFORM DESIGN FOR CURRICULUM RECOMMENDATION SYSTEM

3.1 Sentiment Dictionary System

The overall system block diagram for the sentiment dictionary construction presented in the present study is classified into three models, that is, an area of a data collection and storage place for data collection, storage, and processing, a main server where sentiment words for natural language processing and morphemes, and finally a web system server for opinion extraction. The curriculum recommendation system presented in this study is built as a platform that is processed in two ways using self-diagnostic data submitted by learners in an online form.

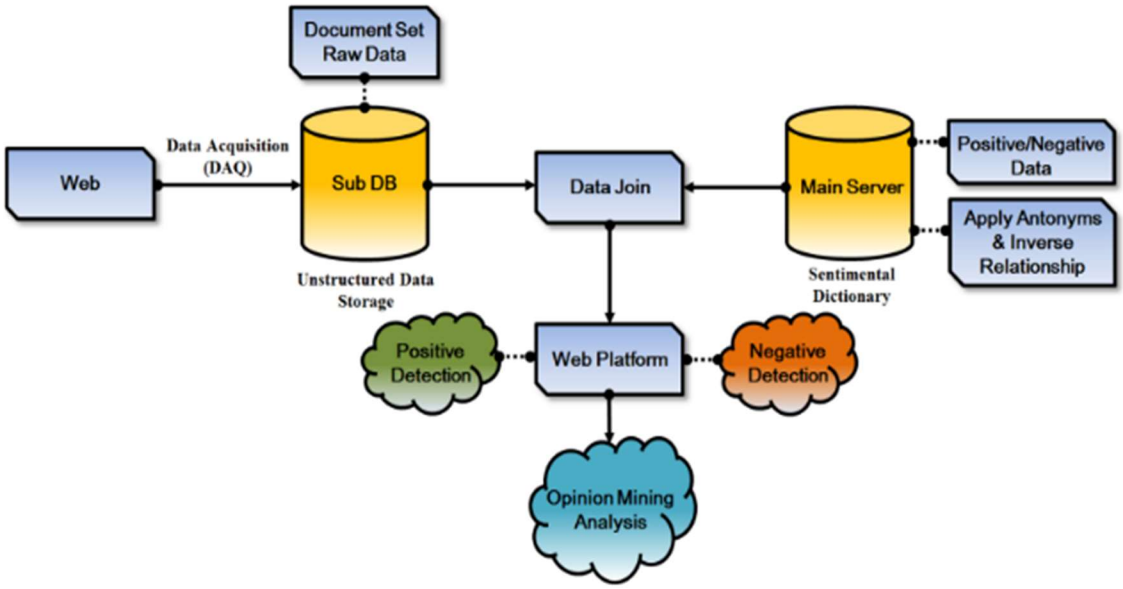

Figure 1: Design of Sentiment Dictionary Process

The main server consists of OM System (Opinion Mining System) and Survey System. Self-diagnostic data of learners is collected using an online form, and the collected data is stored on the main server. The

platform was constructed with a configuration suitable for a big data environment by building a data warehouse as a whole.

3.2 Curriculum Recommendation System based on Opinion Mining

The first method, the process of the OM System, is a technique for processing unstructured data, and the natural language data submitted by the learner through self-diagnosis is stored in the server. The stored data extracts patterns for positive and negative degrees by using opinion mining analysis, and recommends a curriculum suitable for learners based on matching decision-making scenarios [3]. The composition of the opinion mining platform was designed as an element for extracting sentiment words used for the analysis of unstructured data. It was constructed as the sentiment dictionary by labeling positive and negative words after going through a preprocessing process that removes meaningless words, stop words, single-letter words, and special symbols through data filtering to extract the most important meaningful data in big data analysis. In this study, the self-constructed sentiment dictionary consists of more than 170,000 emotional words, including positive and negative Korean-based words.

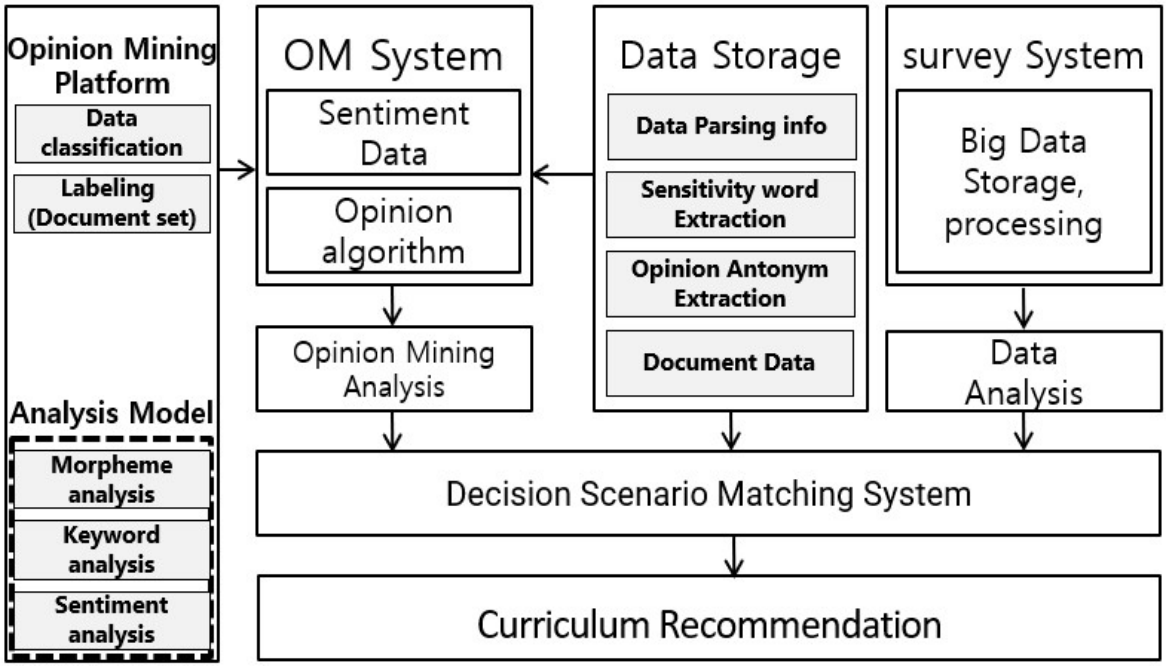

Figure 2: Design of Curriculum Recommendation System

In the course of action for learners to receive curriculum recommendations using the OM System, the online form-type item list provided by this system is written in descriptive form. The descriptive self-diagnostic data submitted by the learner is regarded as unstructured data and joined with the sentiment dictionary to extract positive and negative words.

It recommends a curriculum suitable for the learner's difficulty through the matching process of decision-making scenarios based on the words extracted from the learner's self-diagnostic process, Because the OM System proposed in this study recommends the curriculum by processing the text, which is the unstructured data submitted by the learner, it enables quantitative analysis of the learner's level and competency, and can improve the learning effect in terms of increasing the accuracy in judging which part of the learner is lacking.

3.3 Curriculum Recommendation System based on Matching Level

The second method, the Survey System, is a survey method-based recommendation system that is processed based on structured data. It is a technique

that stores the four multiple-choice questionnaire submitted by the learner in the main server and recommends customized learning content through the survey result. This technique has the advantage of being able to self-diagnose in real time and recommending a curriculum suitable for the learner by identifying the learner's propensity or level. In the recommendation server of the Survey System, there are k survey questions built on the platform, and the survey questions are composed of multiple-choice surveys including n selectable texts, respectively. Here, k questionnaire items and n texts mean two or more natural numbers. In the case of k questionnaire items, there may be multiple items in one questionnaire document, If the value of n is 4 in the question of k, it means a multiple-choice question with four choices.

Table 1: Survey Data Table.

K=3	If the value of n is 4			
Item1	1 text	1 text	1 text	1 text
	Text 1	Text 2	Text 3	Text 4
Item2	1 text	1 text	1 text	1 text
	Text 5	Text 6	Text 7	Text 8
Item3	1 text	1 text	1 text	1 text
	Text 9	Text 10	Text 11	Text 12

In addition, a plurality of preset learning curriculums are managed by a matching table corresponding to each. In the matching table, the relative rankings between texts representing how well the n texts included in the k survey questions are matched to the learning curriculum are stored separately for each survey question.

Table 2: An Example of Matching Level Rank Table.

	K=3	If the value of n is 4			
Curriculum 1	Question 1	Matching LV.1	Matching LV.2	Matching LV.3	Matching LV.4
		Rank = 1	Rank = 2	Rank = 3	Rank = 4
	Question 2	Matching LV.5	Matching LV.6	Matching LV.7	Matching LV.8
		Rank = 1	Rank = 2	Rank = 4	Rank = 3
	Question 3	Matching LV.9	Matching LV.10	Matching LV.11	Matching LV.12
		Rank = 4	Rank = 3	Rank = 2	Rank = 1

	K=3	If the value of n is 4			
Curriculum 2	Question 1	Matching LV.1	Matching LV.2	Matching LV.3	Matching LV.4
		Rank = 3	Rank = 4	Rank = 1	Rank = 2
	Question 2	Matching LV.5	Matching LV.6	Matching LV.7	Matching LV.8
		Rank = 4	Rank = 3	Rank = 2	Rank = 1
	Question 3	Matching LV.9	Matching LV.10	Matching LV.11	Matching LV.12
		Rank = 1	Rank = 2	Rank = 4	Rank = 3

	K=3	If the value of n is 4			
Curriculum 3	Question 1	Matching LV.1	Matching LV.2	Matching LV.3	Matching LV.4
		Rank = 2	Rank = 3	Rank = 4	Rank = 1
	Question 2	Matching LV.5	Matching LV.6	Matching LV.7	Matching LV.8
		Rank = 3	Rank = 4	Rank = 1	Rank = 2
	Question 3	Matching LV.9	Matching LV.10	Matching LV.11	Matching LV.12
		Rank = 2	Rank = 1	Rank = 3	Rank = 4

[Table 2] is an example that shows matching level rank of the three curriculum and the above data is applied to a platform by the number of each subject. For each curriculum, you may assign relative ranks of four different texts, which represents their matching level with the curriculum, to three

questions. For instance, let's suppose that curriculum 1 is "Understanding of Artificial Intelligence," question 1 asks "Which one are you interested in among the fourth industrial revolution related technologies?," and text 1 is "AI", text 2 "Big Data", text 3 "IoT", and text 4 "3D Printer." Then, in recording relative ranks of the texts representing matching level of "Understanding of Artificial Intelligence" curriculum, it may be organized in this order from rank 1: AI, Big Data, IoT, and 3D Printing.

Using such an algorithm enables to generate matching level vector that corresponds to each curriculum: we asked the first one among the learners to submit the questionnaire, referred to the matching level table corresponding to curriculum, and constructed matching level vector in k dimension, which holds a certain rank corresponding to the answering text of k questions. This recommendation system calculates Manhattan Norm of the matching level vector, which fits with individual curriculum, as the first selection reference value. The system recommends learners a certain curriculum that has the minimum selection reference value. Manhattan Norm is LI norm that represents the size of vector or a matrix, and can be defined by the equation below;

$$\| X \|_1 = \sum_{k=1}^{n} |x_k| \quad (1)$$

Here, $\| X \|_1$ represents Manhattan Norm, while X_K signifies vector or K_{th} component of the matrix. In other words, if a learner selects text 1 in Q(question)1, text 6 in Q2, and text 11 in Q3 for curriculum 1, it draws out the matching level vector of curriculum 1 as [1 2 2] with "Rank 1, rank 2, rank 2" corresponding to each text in order.

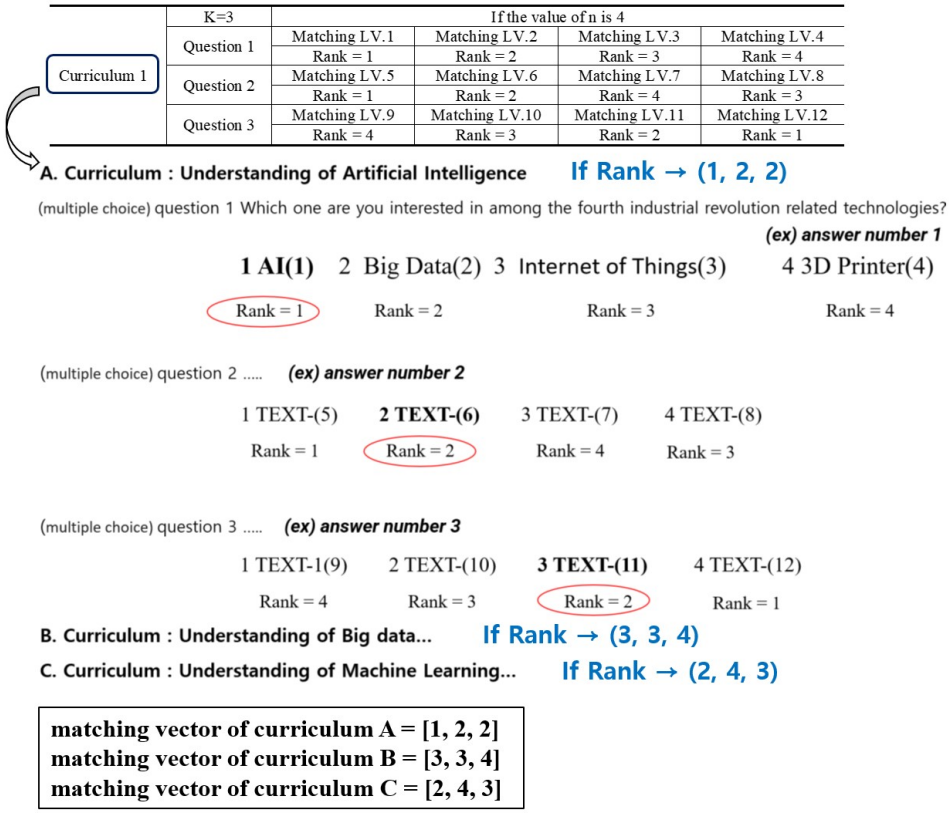

Figure 3: Flow of Curriculum Recommendation

Likewise, when assuming the overall matching level vectors drawn out from the questionnaire submitted by learner are [1 2 2], [3 3 4], and [2 4 3] for curriculum 1, 2, and 3, respectively, the system selects Manhattan Norm 5, 10, 9 of matching level vector as standard values, and recommends

curriculum 1 that has the minimum value to the learner.

Table 3: An Example of Matching Level Rank Vector

Curriculum	Metrix	matching level vector	order of priority
Understanding of Artificial Intelligence	[1, 2, 2]	5	1
Understanding of Big data	[3, 3, 4]	10	3
Understanding of Machine Learning	[2, 4, 3]	9	2

4. SUPPLEMENTARY CURRICULUM RECOMMENDATION SYSTEM

Also, the system offers a supplementary curriculum recommendation to learners. The recommendation is determined by calculating the Euclidean distance between matching level vectors that correspond to not-recommended curriculums, and one that correspond to the finally recommended curriculum. The Euclidean distance refers to the distance between two vectors and can be defined as below;

$$D = \sqrt{\sum_{i=1}^{n} (A_i - B_i)^2} \qquad (2)$$

Here, D represents the Euclidean distance, while A_i and B_i show the i^{th} components in two vectors. In general, the shorter the Euclidean distance between two vectors, the more similar two vectors are, and the reverse is also true. By adopting this algorithm, calculation of the Euclidean distance between the matching level vector of curriculum 2 and 3 and that of curriculum 1, the final selection, draws out two values in total. Among the two curriculums, the one that has the shorter Euclidean distance with the matching level vector corresponding to the finally selected curriculum 1 is transmitted to the learner as a recommendation for a supplementary curriculum.

Thus, the proposed survey system can recommend customized education for a learner by determining personal tendency, difficulty level, and capabilities based on the transmitted responses based on structured data. Although its quantitative analysis may be lacking compared to the recommendation based on opinion mining analysis suggested by Opinion Mining (OM) system, the former has a clear advantage that it allows the learner to conduct self-diagnosis and receive recommendations in a short time.

5. CONCLUSION

The thesis proposes a self-diagnostic system to recommend a customized curriculum to learners by using two methods based on unstructured and structured data.

this platform enables learners to use educational content in connection with the e-learning system. Also, this study proposes more valuable and future-oriented education infrastructure by converging artificial intelligence that has a foundation on a hyper-connected society and big data, the central pillars of the fourth industrial revolution.

The study is expected to have large applications and enhance learners' satisfaction with their academic achievements in the pandemic era of COVID-19 and the like by recommending an educational curriculum that suits personal level and aptitude.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(No. 2020R1I1A1A01055323)

REFERENCES:

[1] Ji-Hoon Seo, EunMi Cho, Kil-Hong Joo, "Analysis of Agenda Prediction According to Big Data Based Creative Education Performance Factors," Advances in Computer Science and Ubiquitous Computing, CUTE 2017, CSA 2017, pp. 1314-1319, December 2017.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[2] Ji-Hoon Seo, Seok-jin Im, "Designing a Learning Model for an Artificial Intelligence Curriculum," REVIEW OF INTERNATIONAL GEOGRAPHICAL EDUCATION, CUTE 2017, CSA 2017, pp. 1972-1977, 11(8), SPRING, 2021.

[3] Seo Ji-Hoon, Joo Kil-Hong (2018) Analysis of the elements of future development of korean style software education through the opinion mining technique. Lect Notes Electr Eng 474:1410–1415.

[4] Miyoung Ryu, Sungwan Han, "A Study of SW Education Contents based on Computational Thinking", JOURNAL OF The Korean Association of information Education, 13(2), pp.521–528, 2019.

[5] Ministry of Education, Education Policy Direction and Core Tasks in the Age of Artificial Intelligence, 2020.

[6] Ministry of Science and Technology Information and Communication, "Artificial Intelligence National Strategy", 2019.

[7] Ji-Hoon Seo, Ho-Sun Lee and Jin-Tak Choi, "Classification Technique for Filtering Sentiment Vocabularies for the Enhancement of Accuracy of Opinion Mining", International Journal of u- and e- Service, Science and Technology, Vol.8 No.10, 2015, pp.11-20.

[8] D. Kim, T. Cho and J. H. Lee, "A domain adaptive sentiment dictionary construction method for domain sentiment analysis," Proceedings of the Korean Society of Computer Information Conference, Vol. 23, No. 1, pp. 15-18, 2015.

[9] Courses, E., Surveys, T.: Using Sentiment SentiWordNet for multilingual sen- timent analysis, IEEE 24th International Conference on Data Engineering Workshop, Cancun, Mexico, 2008, pp. 507-512.

[10] Irfan Ajmal Khan, Junghyun Woo, Ji-Hoon Seo, Jin-Tak Choi, "Text Mining : Extraction of Interesting Association Rule with Frequent Itemsets Mining for Korean Language from Unstructured Data", International Journal of Multimedia and Ubiquitous Engineering SCOPUS , Vol.10 No.11, pp.11-20, 2015.