Study on Big Data Based Decision Making Support System for Development of Software Education Model

Jihoon Seo and Kilhong Joo

Abstract—**The purpose of this paper is to study a big data–based decision making support system for Korean software education to analyze the performance of software education that has been carried out and to derive efficient software education model focusing on opinion mining visualization analysis technique based on big data collection data such as portal site news, SNS service, Internet cafe, other performance data related to software education collected from online media over the last 5 years.**

Index Terms—**Big data, opinion mining, software education, decision making support system.**

I. Introduction

Until now, the Korean government has selected and implemented R&D project plans related to major core technology items in preparation for the forthcoming Fourth Industrial Revolution. One of them is Big Data related projects, and technologies are combined in various fields such as Smart City, Knowledge Information Technology (KIT), and Culture Technology (CT). In line with that, the government does not hesitate to provide human and material support by establishing related laws and departments at the government level. In contrast, the utilization of big data technology in the education sector is at a moderate level compared to OECD countries. Fortunately, Korea has world-class IT capabilities, so if systematic related research is conducted even now, it is expected that Korea will be able to stand side by side with major developed countries. The result of review of the software education related polices that the government has been promoting for five years still show various and complicated technical problems. The Ministry of Education has been promoting the convergence talent development program through the cultivation of computing thinking and learners' creative solution ability for many years through software education, but it is true that there is a lack of systematic means of measuring the educational outcomes, resulting in inadequate analysis of visible unstructured data. In addition, there are not proper tools and systems that can specifically quantify this to enable mutual organic linkage and analysis.

In this paper, therefore, we are to present the big data based decision making system modeling technique that can develop a decision making support system that can quantitatively analyze the performance of software education which has been reorganized and implemented elementary and middle compulsory education courses since last year by developing a big data analysis-based decision making support system platform that can quantitatively measure the policy achievements of the software education that the government has been promoting, collect and reprocess all the online reputations that the people felt and experienced about education policies implemented by education-related government ministries into unstructured data to analyze big data centered on Opinion Mining, an unstructured data analysis technique based on this and quantitatively derive future prediction and software education model outcomes for the continuously changing Korean software education model.

II. Related Works

A. Application of Educational Big Data

According to previous studies related to large data utilization, the method of using big data for learning and the frequency of quality about educational big data were derived through the analysis of the relationship between various variables. However, this analysis model focuses only on the analysis of performance factors of creative education, and thus there is a limit to becoming a mid- to long-term development model [1]. Until now, the analysis method of education policy has enabled the related analysis only when the agenda was exposed in advance, so the policy has been implemented through subjective keywords to improve the educational environment. In addition, it is possible to derive the meaning of the education system when judging the importance of the education policy based on the inquiry about educational big data, but not inquired database is needed as a demand for reference data that can be applied or compared to educational policies [2]. An example of the actual use of big data in education is EDSS (Education Data Service System). Equipped with a system that collects, links, and processes education related data accumulated in the education office, metropolitan education office, education-related institutions, etc., this EDSS provides the researchers with the relevant data according to the designated censorship sequence for academic research purposes. The following Fig. 1 displays the basic concept of educational big data development by Google. As in the Fig. 1, the analysis of big data is handled by three steps. The steps are divided into three which are data, cloud based applications, and result.

Manuscript received October 20, 2018; revised July 5, 2019. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF- 2017S1A5A2A01025431) and this research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. 2017R1D1A1B03029292).

Jihoon Seo and Kilhong Joo are with the Department of Computer Education, Gyeongin National University of Education, Korea (Corresponding author: Kilhong Joo; e-mail: khjoo@ginue.ac.kr).

doi: 10.18178/ijmlc.2019.9.5.853

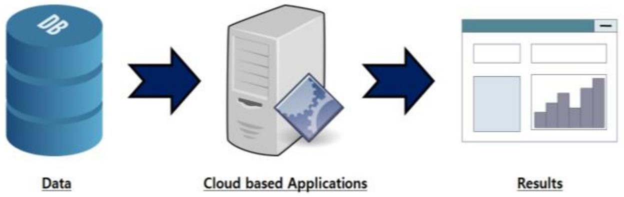


Fig. 1. The basic concept of educational big data developed by Google.

Hadoop, which is a free, Java-based programming framework supports the processing of large sets of data in a distributed computing environment. It is a part of the Apache project sponsored by the Apache Software Foundation. Hadoop cluster uses a Master/Slave structure [12]. Using Hadoop, large data sets can be processed across a cluster of servers and applications can be run on systems with thousands of nodes involving thousands of terabytes. Distributed file system in Hadoop helps in rapid data transfer rates and allows the system to continue its normal operation even in the case of some node failures. This approach lowers the risk of an entire system failure, even in the case of a significant number of node failures. Hadoop enables a computing solution that is scalable, cost effective, flexible and fault tolerant. Hadoop Framework is used by popular companies like Google, Yahoo, Amazon and IBM etc., to support their applications involving huge amounts of data. Hadoop has two main sub projects – Map Reduce and Hadoop Distributed File System (HDFS).

B. Unstructured Data Analysis Techniques

1) Text mining

Text mining is a data analysis technique aimed to extract and process useful information based on Natural Language Processing (NLP) technology. Through this technique, users can obtain results more than simple information search such as extraction of meaningful information from unstructured data, which is a massive text cluster, understanding of linkage with other information and discovery of categories of the text [3]. The main technical areas covered in text mining are as follows:

a) Document classification

In the past, librarians identified the contents of each book and categorized them manually according to a predetermined classification system. Nowadays, the development of digital technology and the activation of the Internet enabled the production and distribution of enormous information, and it is now almost impossible to classify and manage vast amounts of information inside and outside the organization [4].

b) Document clustering

This is a technology that identifies the characteristics of each knowledge content and clusters content which has similar contents or forms or is highly correlated. In addition to being able to group documents of interest together in order of their relevance through document clustering to review them effectively, users can quickly and easily access information hidden in massive documents through example-based queries. Conventional clustering techniques are implemented in a way of extracting differentiated important characteristics through linguistic analysis of the target document, comparing them with the characteristics of other documents, and bonding the documents with high similarity to each other [5].

c) Document summarization

This is a technique for helping each user to understand and utilize information quickly by identifying short and simple summary sentences by effectively reducing the complexity and length while maintaining the key meaning of the document. The document summarization technology is based on the character extraction and information extraction technology, and can be classified into an extraction summary method that selects, extracts and re-processes sentences that can represent the document throughout the text, and a creation summary method that creates sentences using important information [6].

2) Data mining

Data mining is an analysis technique that systematically and automatically analyzes statistical rules and patterns in large-scale stored data and finds necessary information. Here, finding information refers to the process of finding useful patterns and relationships by applying advanced statistical analysis and modeling techniques, which is the core technology of database marketing [7]. The main technical areas covered in data mining are as follows:

As in the Fig. 2, types of data mining technique are handled by two types: supervised learning and unsupervised learning. The supervised learning consists of two techniques – one is classification and regression. In addition, the unsupervised learning consists of two techniques – association rule and clustering.

a) Classification analysis

This is a technique that creates a model to find a value in the target field, generates a classification model by entering past data, and predicts the classification value for new data [8].

b) Association rule

This technique is called shopping cart analysis, and it is a technique to find out the rational products with a strong pattern of being sold together by analyzing the products purchased at once by customers at Internet shopping malls and offline stores. For example, it is possible to find a consumption pattern of "A person purchasing a product A purchases a product B together." Based on this, it is possible to recommend purchasing a product B to the customer who purchases the product A.

c) Continuous pattern

This technique is similar to Relational Rule, and is a technique for analyzing sequential purchase patterns by adding time information to Relational Rule. For example, it is possible to find a consumption pattern of "A person purchasing a product A additionally purchases a product B after 1 month." Based on this rule, it is possible to recommend related products to the customer [9].

3) Opinion mining

As a classification of original text mining technique. The opinion mining technique is called buzz monitoring. It discriminates positive, negative and neutral preferences of structure and unstructured text collected on SNS and is used for market size forecasts for specific services and products, consumer response analysis, word of mouth, etc. [10]. This technique extracts vocabulary information expressing positive and negative, recognizes sentences composed of opinions on the object to measure and determine positive and negative degrees by the sum of patterns including opinions. The opinion mining analysis technique can discover more valuable data information from unstructured data composed of many unspecified users. In another aspect, it can be called sentiment analysis, which can be interpreted in the broad sense of natural language processing and computer linguistics analysis and text mining [11]. The Fig. 3 shows examples of opinion mining techniques. As this figure, size, color, and length are depends on the relationship between the words analyzed.

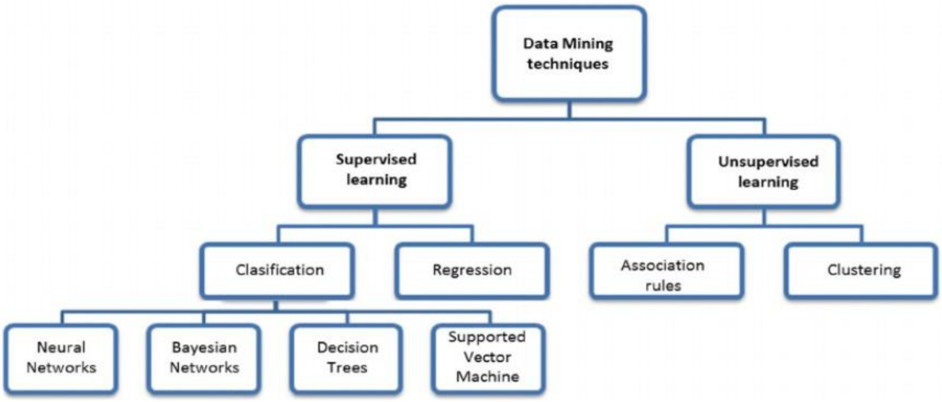


Fig. 2. Types of data mining technique (https://www.researchgate.net/).

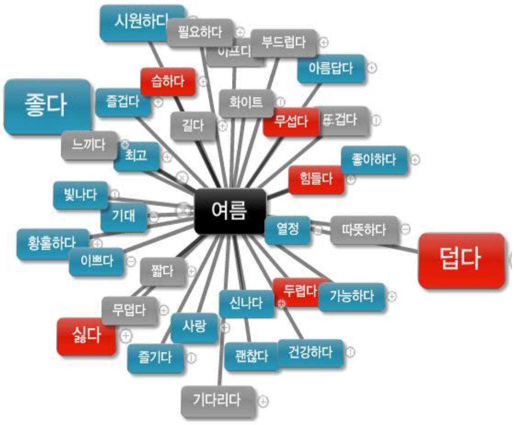


Fig. 3. Examples of opinion mining techniques.

III. Software Education Performance Analysis

As part of attempts to derive reputations in and trends of public opinions regarding software education policies included in writings written on social media or portal news comments based on unstructured data, the opinion mining analysis process proposed in this paper randomly collects unstructured data including the words “making software education mandatory”, which are a core keyword in relation to the Ministry of Education, the Ministry of Science, ICT and Future Planning, and the Korea Foundation for the Advancement of Science and Creativity from the three items, blogs, Twitter, and portal news distributed on online media to classify the polarity of sentiment words according to affirmation and negation and select sentiment data. The configuration of the overall system for the development of the sentiment dictionary, which is one of the important stages of this study, is classified into three models, a storage server for data collection, storage, and pre-processing, a natural language processing learning model for natural language processing and sentiment word morpheme analysis, and a sentiment dictionary construction server conversion stage for opinion extraction.

In order to analyze the most effective software education performance, we constructed a decision making system platform development process composed of three steps in total.

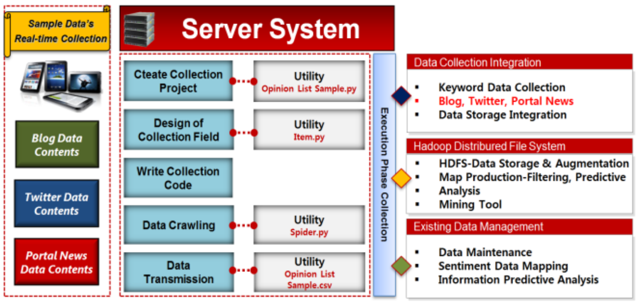


Fig. 4. Opinion analysis model.

A. Collection and Storage Processing Technology of Unstructured Data

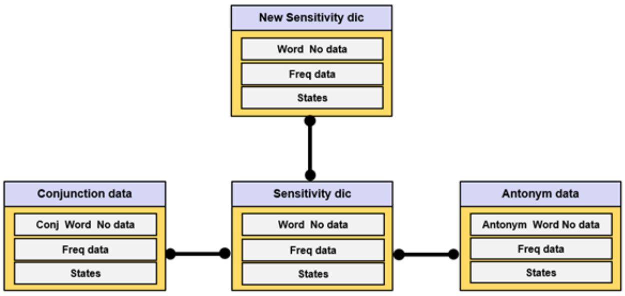
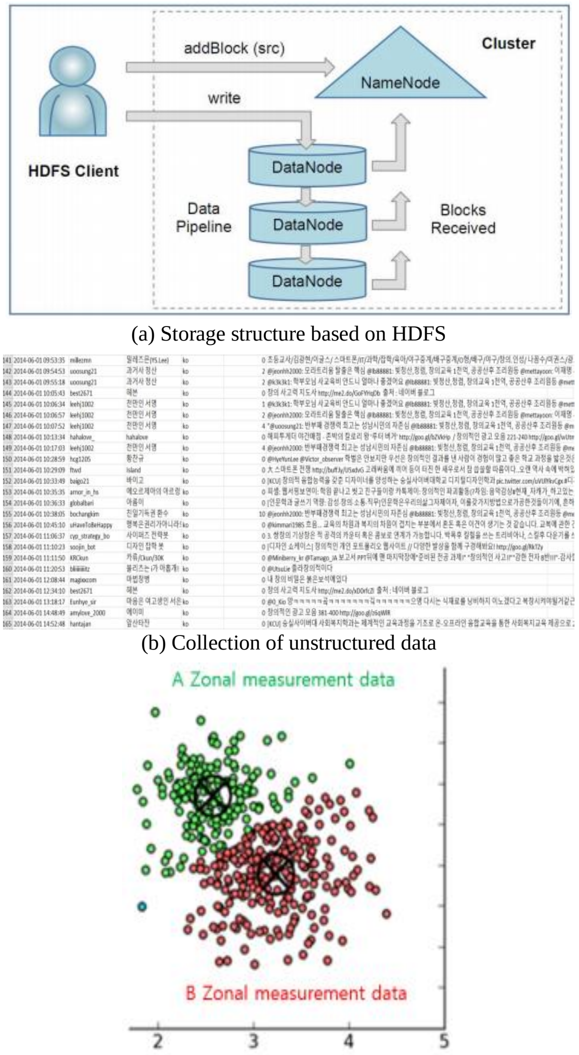


Fig. 5. Sentiment Dictionary DB join design.

In this step, Python-based crawling technique was performed to lower the dependency of real-time data collection in this paper as much as possible and to collect raw data based on the accumulated unstructured data. To conduct opinion mining analysis, a sentiment dictionary must be constructed. In order to improve the accuracy and reliability of this sentiment dictionary, the collection of unstructured data is presented as an important element. First, the times and scales of the collected data should be clear and the larger the quantity of collected data, the higher the achievable accuracy of the sentiment dictionary. The Fig. 6 displays storage structure, collection data and clustering model of unstructured data. In this process, unstructured data on software education policy related issues were collected randomly from Internet news portal articles, blogs, online communities, and SNS from 2012 to 2017.



(a) Storage structure based on HDFS

141	2014-06-01 09:53:35	millermn	밀레즈문(YS.Lee)	ko		0	초등교사/김광현/이글스/스마트폰/IT/과학/잡학/육아/야구중계/배구중계/0형/배구/야구/창의.인성/나꼼수/미권스/꽁
142	2014-06-01 09:54:53	uoosung21	과거사 청산	ko		2	@jeonhh2000: 모라트리움 탈출은 핵심 @lb88881: 빛청산,청렴, 창의교육 1천억, 공공산후 조리원등 @mettayoon: 이재명
143	2014-06-01 09:55:18	uoosung21	과거사 청산	ko		2	@k3k3k1: 학부모님 사교육비 안드니 얼마나 좋겠어요 @lb88881: 빛청산,청렴, 창의교육 1천억, 공공산후 조리원등 @mett
144	2014-06-01 10:05:43	best2671	해본	ko		0	창의 사고력 지도사 http://me2.do/GoFYHqDb 출처 : 네이버 블로그
145	2014-06-01 10:06:34	leehj1002	천만인 서명	ko		1	@k3k3k1: 학부모님 사교육비 안드니 얼마나 좋겠어요 @lb88881: 빛청산,청렴, 창의교육 1천억, 공공산후 조리원등 @mett
146	2014-06-01 10:06:57	leehj1002	천만인 서명	ko		2	@jeonhh2000: 모라트리움 탈출은 핵심 @lb88881: 빛청산,청렴, 창의교육 1천억, 공공산후 조리원등 @mettayoon: 이재명
147	2014-06-01 10:07:52	leehj1002	천만인 서명	ko		4	"@uoosung21: 반부패 경쟁력 최고는 성남시민의 자존심 @lb88881: 빛청산,청렴, 창의교육 1천억, 공공산후 조리원등 @m
148	2014-06-01 10:13:34	hahalove_	hahalove	ko		0	해피투게더 야간매점 - 존박의 칼로리 왕 '쿠터 버거' http://goo.gl/bZVkHp / 창의적인 광고 모음 221-240 http://goo.gl/wUtn
149	2014-06-01 10:17:03	leehj1002	천만인 서명	ko		4	@jeonhh2000: 반부패경쟁력 최고는 성남시민의 자존심 @lb88881: 빛청산,청렴, 창의교육 1천억, 공공산후 조리원등 @me
150	2014-06-01 10:28:59	hcg1205	황찬규	ko		0	@HyeYunLee @Victor_observer 학벌은 안보지만 우선은 창의적인 결과를 낸 사람이 경험이 많고 좋은 학교 과정을 밟은것은
151	2014-06-01 10:29:09	ftwd	island	ko		0	大 스마트폰 전쟁 http://buff.ly/U5xdvG 고래싸움에 끼어 등이 터진 한 새우로서 참 씁쓸할 따름이다...오랜 역사 속에 박혀있
152	2014-06-01 10:33:49	baigo21	바이고	ko		0	[KCU] 창의적 융합능력을 갖춘 디자이너를 양성하는 숭실사이버대학교 디지털디자인학과 pic.twitter.com/uVUfYkvCgx #디
153	2014-06-01 10:35:35	amor_in_hs	에오르제아의 아르링	ko		0	피생: 웹서핑보면이: 학원 끝나고 씻고 친구들이랑 카톡제이: 창의적인 파괴활동(?차임: 음악감상#현재_자캐가_하고있는
154	2014-06-01 10:36:33	globalbari	아롬이	ko		0	[인문학과 글쓰기 역량: 감성·창의·소통·직무]인문학은우리의삶그자체이자, 이를갖가지방법으로가공한것들이기에, 흔히
155	2014-06-01 10:38:05	bochangkim	친일기득권 환수	ko		10	@jeonhh2000: 반부패경쟁력 최고는 성남시민의 자존심 @lb88881: 빛청산,청렴, 창의교육 1천억, 공공산후 조리원등 @me
156	2014-06-01 10:45:10	uHaveToBeHappy	행복은권리가아니라!	ko		0	@kimman1985 흐음... 교육의 차원과 복지의 차원이 겹치는 부분에서 혼돈 혹은 이견이 생기는 것 같습니다. 교복에 관한 경
157	2014-06-01 11:06:37	cyp_strategy_bo	사이퍼즈 전략봇	ko		0	3. 쌍창의 기상창은 적 공격의 카운터 혹은 콤보로 연계가 가능합니다. 박목후 칼힘을 쓰는 트리비아나, 스킬후 다운기를 쓰
158	2014-06-01 11:10:23	soojin_bot	디자인 잡학 봇	ko		0	[디자인 쇼케이스] 창의적인 개인 포트폴리오 웹사이트 // 다양한 발상을 함께 구경해봐요! http://goo.gl/RkT2y
159	2014-06-01 11:11:50	KRCkun	카쿠/Ckun/30K	ko		0	@Miniberry_kr @Tamago_JA 보고서 #PPT뒤에 맨 마지막장에"준비된 전공 과제!" "창의적인 사고!!"강한 전자 8번!!!"-감사합
160	2014-06-01 11:20:53	bliiiiiitz	블리츠는 i가 아홉개!	ko		0	@UtsuLie 졸라창의적이다
161	2014-06-01 12:08:44	magiocom	마법장병	ko		0	내 창의 비밀은 붉은보석에있다
162	2014-06-01 12:34:10	best2671	해본	ko		0	창의 사고력 지도사 http://me2.do/xD0rfcZl 출처 : 네이버 블로그
163	2014-06-01 13:18:17	Eunhye_sir	마음은 여고생인 서은	ko		0	@0_Kio 앙ㅋㅋㅋㅋㅋ곰ㅋㅋㅋㅋㅋㅋㅋㅋ깈ㅋㅋㅋㅋㅋㅋㅋ으앙 다시는 식재료를 낭비하지 아노겠다고 복창시키야될거같군
164	2014-06-01 14:48:49	amylove_2000	에이미	ko		0	창의적인 광고 모음 381-400 http://goo.gl/z6qWRR
165	2014-06-01 14:52:48	hantajan	암산타잔	ko		0	[KCU] 숭실사이버대 사회복지학과는 체계적인 교육과정을 기초로 온-오프라인 융합교육을 통한 사회복지교육 제공으로 2

(b) Collection of unstructured data

(c) Clustering model

Fig. 6. Unstructured data collection storage structure.

User applications access the file system using the HDFS client, a code library that exports the HDFS file system interface. Similar to most conventional file systems, HDFS supports operations to read, write and delete files, and operations to create and delete directories. The user references files and directories by paths in the namespace. The user application generally does not need to know that file system metadata and storage are on different servers, or that blocks have multiple replicas. When an application reads a file, the HDFS client first asks the NameNode for the list of DataNodes that host replicas of the blocks of the file. It then contacts a DataNode directly and requests the transfer of the desired block. When a client writes, it first asks the NameNode to choose DataNodes to host replicas of the first block of the file. The client organizes a pipeline from node-to-node and sends the data. When the first block is filled, the client requests new DataNodes to be chosen to host replicas of the next block. A new pipeline is organized, and the client sends the further bytes of the file. Each choice of DataNodes is likely to be different. The interactions among the client, the NameNode and the DataNodes are illustrated in Fig. 6(a). Fig. 6(b) and Fig. 6(c) show the collection of unstructured data crawling from the Web and the clustering model for processing the collection of unstructured data respectively.

B. Data Integration System of Unstructured Data

In this step, the data warehouse (DW) requires a stable and physically appropriate modeling system for effective decision making analysis. The Fig. 7 shows the build of data warehouse model for performance analysis of software education. As in the Fig. 7, we applied three types of construction methods to distribute, process and manage unstructured data collected on the Internet by appropriately using the storage server for big data analysis.

1) Based on the ETL tools of Extraction, Transformation, and Loading, we construct a disk sharing method to manage and refine past raw data

2) A relational database management system (RDBMS) is used to build big data based on Map Reduce and Stream

3) Using the Opinion Sentiment Dictionary through training for each unstructured text document, we extract the positive and negative reputation with high reliability in unstructured data related to software education and build the technique with improved accuracy.

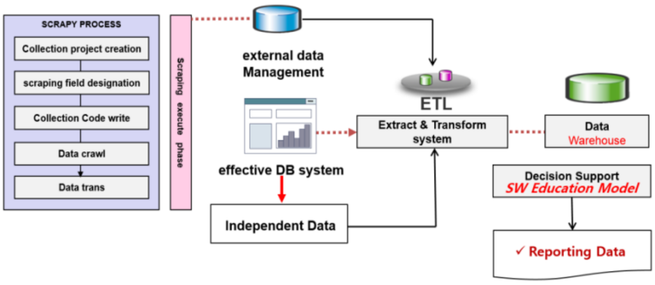


Fig. 7. Build of data warehouse model for performance analysis of software education.

C. Opinion Mining Analysis Platform of Unstructured Data

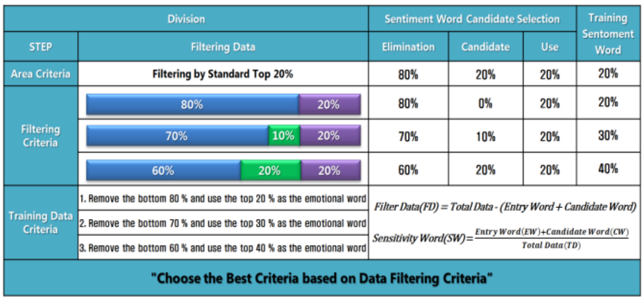
In this step, the interface only for a Korean grammar-based sentiment dictionary (API) was separately designed to perform more efficient Opinion Mining analysis and the data filtering technique was used to maximize the reliability and accuracy of the collected data. In addition, an Opinion Mining Automatic Analysis Platform was developed for the purpose of facilitating the derivation of software education agenda.

TABLE I: WORD STEMMING FILTER PROCESS

***Filtering rules applied for the construction of a sentiment dictionary**
1. Remove special characters, English words, and disused vocabularies
2. Remove meaningless terms and one-letter texts
3. Separate the same words in the form of a conjunction to classify the natures of sentiment words
4. Distinguish homonyms and synonyms from each other
5. In the case of abbreviations and newly coined words, reflect only those that have been registered in WiKipedia or a Korean language dictionary

In this paper, outliers, missing values, or wrong values are not shown in the data values because the preprocessing was performed based on the collected unstructured data. In particular, filtering was conducted to extract highly important words in sentences or important words that could be sentiment words. Therefore, the proposed filtering process derive sentiment words from sentences in Korean grammar efficiently and this rules as shown in Table I were applied to perform word filtering.

The words derived through the preprocessing of meaningful words are classified into sentiment word candidate groups. In this paper, to improve the accuracy and reliability of the resultant values of opinion mining analysis, the top 20% of the words selected as the sentiment word candidate groups were extracted to finally select them as sentiment words. Although the data distributed in the lower groups are composed of the sets of words with high weighted values or meaningful words, most of them fall under neutral words and words close to other polarities in cases where they are tagged as positive, negative, neutral, and other polarities and polarity classification is carried out. Therefore, the word sets that are less useful as such were removed in advance. On the other hand, words that correspond to neutral and other polarities also exist among the top 20% words finally selected as sentiment words but they have very high weighted values or correspond to higher word groups with very high frequencies in documents.



Division		Sentiment Word Candidate Selection			Training Sentoment Word
STEP	Filtering Data	Elimination	Candidate	Use	
Area Criteria	Filtering by Standard Top 20%	80%	20%	20%	20%
Filtering Criteria	80% 20%	80%	0%	20%	20%
	70% 10% 20%	70%	10%	20%	30%
	60% 20% 20%	60%	20%	20%	40%
Training Data Criteria	1. Remove the bottom 80 % and use the top 20 % as the emotional word	$Filter\ Data(FD) = Total\ Data - (Entry\ Word + Candidate\ Word)$ $Sensitivity\ Word(SW) = \frac{Entry\ Word(EW) + Candidate\ Word(CW)}{Total\ Data(TD)}$			
	2. Remove the bottom 70 % and use the top 30 % as the emotional word				
	3. Remove the bottom 60 % and use the top 40 % as the emotional word				
"Choose the Best Criteria based on Data Filtering Criteria"					

Fig. 8. Data filtering results for opinion analysis model.

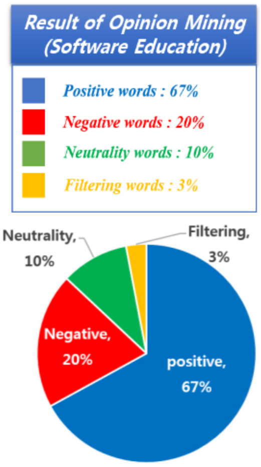


Fig. 9. Visualization analysis of opinion mining sentiment dictionary for software education.

IV. Experiment Result and Analysis

In this paper, the Opinion Mining analysis based on Big Data was conducted from 2012 to 2017 focusing on the key keyword of "Mandatory Software Education", and the polarity classification part of reputation analysis was divided into 4 kinds of positive, negative, neutrality and others. The Fig. 9 and Fig. 10 display the analysis of opinion mining sentiment dictionary for software education. As this figure, the positive takes 67% of total words. On the other hand, the negative takes 20% of total words. This means that the effects of software education in Korea is a very positive change.

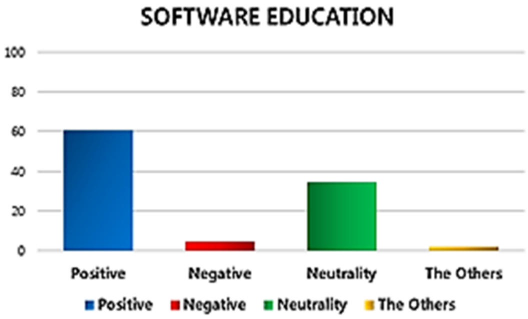


Fig. 10. Analysis of opinion mining for software education.

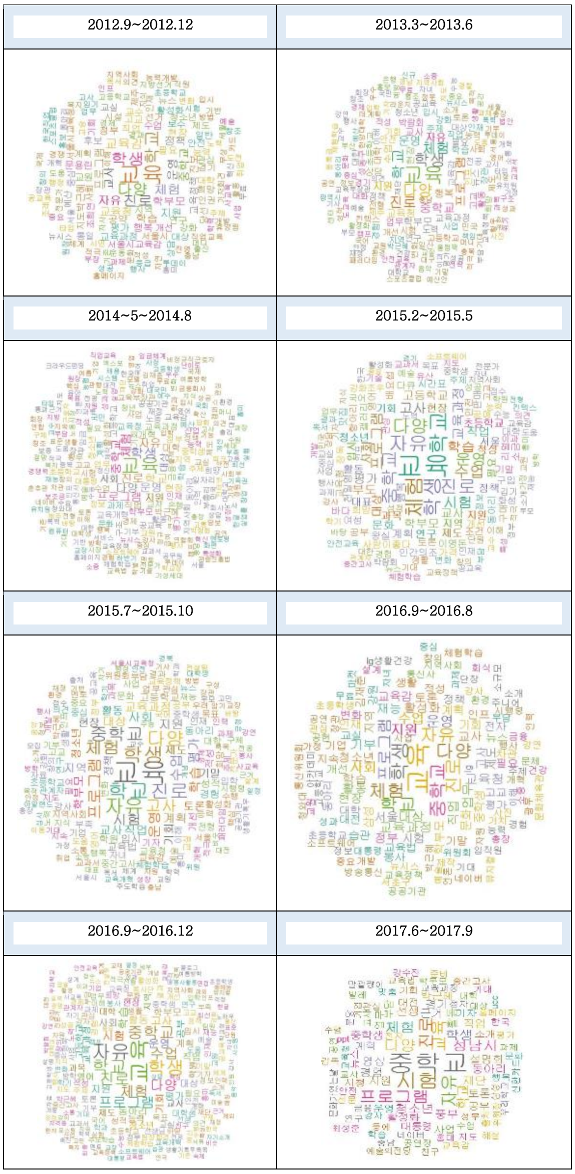


Fig. 11. Analysis of word frequencies on the internet for software education.

The Fig. 11 shows analyzed result of word frequencies on the internet for software education. In the case of the neutral part, the noun phrase and the analysis element were excluded from the sentiment word dictionary because they correspond to meaningless items. In the case of the others part, all of the words are unusual words with ambiguous meanings, which were classified into meaningless data without analysis value.

The result of the opening mining analysis on software education of Korea showed that positive reputation is measured to be more than 60%, showing relatively strong response to learners and the positive index was measured to indicate a high rate of growth each year as the software education in the Korean elementary and middle curriculum has been implemented as a regular curriculum since last year.

V. Conclusion

The use of the decision making support platform for the effective data analysis of Korean software education performance developed in this study enables stable predictions of variable factors in the ever-changing education environment, and it can not only be a way to develop the future national education and to find effective solutions, but it will be an opportunity to recognize the necessity of studying a Korean software education model suitable for the Korean situation, and it is expected that we will be able to lay the foundations for developing and fostering creative talents with international competitiveness in the era of the forthcoming Fourth Industrial Revolution. In addition, when comparing the unstructured data-based big data analysis decision making system such as Internet media, past year policy report with existing unstructured data-based analysis techniques such as brainstorming, Delphi, expert channel, it is expected to develop into a versatile platform that can be widely used in IT fields and other areas such as the field of culture technology (CT) as a highly reliable and accurate system in terms of technology.

References

[1] Y. Sun and K. Jia, “Research of word sense disambiguation based on mining association rules,” in *Proc. Third International Symposium on Intelligent Information Technology Application Workshops*, NanChang, China, 2009, pp. 86-88.

[2] X. Wu, X. Zhu, G.-Q. Wu, and W. Ding, “Data mining with big data,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 26, no. 1, pp. 97-107, 2014.

[3] C. Tang and C. Liu., “Method of Chinese grammar rules automatically access based on association rules,” in *Proc. the. Computer Science and Computational Technology*, Shanghai, China, vol. 1, pp. 265-268, 2008.

[4] I. A. Khan and J. T. Choi, “An application of educational data mining (EDM) technique for scholarship prediction,” *International Journal of Software Engineering and Its Applications*, vol. 8, no. 12, pp. 31-42, 2014.

[5] Y. Xu, Y. Li, and G. Shaw, “Reliable representations for association rules,” *Data & Knowledge Engineering*, vol. 70 no. 6, pp. 555-575, 2011.

[6] B. Pang, L. Lee, and S. Vaithyanathan, “Thumbs up?: Sentiment classification using machine learning techniques,” in *Proc. the ACL-02 Conference on Empirical methods in Natural Language Processing*, 2002, vol.10, pp.79-86.

[7] J.-H. Seo, E. Cho, and K.-H. Joo, “Analysis of agenda prediction according to big data based creative education performance factors,” *Lecture Notes in Electrical Engineering*, vol. 474, pp. 1876-1100, 2018.

[8] E. Courses and T. Surveys, “Using sentiment SentiWordNet for multilingual sentiment analysis,” in *Proc. IEEE 24th International Conference on Data Engineering Workshop*, Cancun, Mexico, 2008, pp. 507-512.

[9] H.-J. Woo, N.-H. Park, K.-H. Joo, “OLAP analysis based on clustering over big data streamers,” *Asia Life Sciences*, vol. 11, pp. 223-232, 2015.

[10] M. Stonebraker, “SQL databases v. NoSQL databases,” *Communications of the ACM*, vol. 53, pp. 10-11, 2010.

[11] J. Manyuka *et al*., “Big data: The next frontier for innovation, competition, and productivity,” McKinsey Global Institute, America, 2011.

[12] L. Huang, T.-T. Hu, and H.-S. Chen, *Research on Hadoop Cloud Computing Model and its Applications*, Hangzhou, China: 2012, pp. 59 – 63.

Jihoon Seo received his bachelor degree in 2008 at Seoul National University of Science and Technology from department of Safety Engineering. He finished his MS and Ph.D. at Incheon National University from Department of Computer Science and Engineering in 2010 and 2015 respectively. His research interest includes data mining, database management, big data analysis and software education

Kilhong Joo received his M.S. and Ph.D. degree in computer science from Yonsei University, Seoul, Korea, in 2000 and 2004. He is currently a professor of Department of Computer Education at Gyeongin National University of Education, Korea. His current interests include mining data streams, data analysis, big data, smart learning, and software education.