ICIC Express Letters
Volume **12**, Number **12**, December **2018**

ICIC International ©2018 ISSN 1881-803X
pp. 1249–1257

EXTRACTING KEYWORD-RELATED AFFECTIVE WORDS USING THE TEXT MINING TECHNIQUE

Jinho Ahn and Kyungdoh Kim*

Department of Industrial Engineering
Hongik University
94, Wausan-ro, Mapo-gu, Seoul 04066, Korea
*Corresponding author: kyungdoh.kim@hongik.ac.kr

Received May 2018; accepted August 2018

Abstract. *Existing affective word extraction methods that are used in affective engineering typically include literature research, questionnaires, and interviews. While these techniques can effectively extract affective words, they are often costly and time consuming. This study proposes an affective word extraction method through text mining, which reduces these costs. The proposed process is as follows. First, web crawling is used to collect text data related to various topics. Then, the natural language processing tool is used to preprocess the collected data. Next, word candidates are obtained through the text mining technique. Lastly, the word candidates are evaluated in terms of their suitability as affective words, and the word selection is finalized. To verify the proposed method, a case study was conducted by considering the keyword "product" and extracting affective words. A comparison with data from a previous study showed that the present method extracts affective words as effectively as existing methods. The time and financial cost that incur through text mining are generally smaller than that involved in literature studies, questionnaires, and interviews. Therefore, the affective word extraction method proposed in this paper exhibits an equivalent performance but reduced time and financial costs compared to existing methods.*

Keywords: Affective engineering, Affective word, Text mining, Keyword, Web crawling

1. **Introduction.** The worldwide industrial advancement has surpassed the mass production era and has entered a period of mass customization. Specifically, the saturation of product supply generated a need to assess customer demand and preference as a strategy to promote purchases from consumers. This coincided with the birth of affective engineering, which converts a consumer's feelings toward a product into design factors [1]. The increased interest of the industrial world in affective engineering began with Mazda Motor Company of Japan in 1986, and many companies use affective engineering today to assess consumer demand.

Affection is "an aesthetic and psychological experience that occurs inside a human being through the emotion or perception that results from external physical stimulation" [2]. Because the emotions that are felt by the user of a product towards the product are extremely subjective and multi-dimensional, attempting to measure and analyze them may be a contradictory concept. However, there are continual efforts to quantify and measure affection using the biological and psychological measurement methods. The biological measurement method measures biological reactions (such as pulse, electromyography, and electroencephalogram) to external stimulation, whereas the psychological measurement method involves subjective evaluations of external stimulations and typically uses semantic differentials (SDs) that employ semantic affective words [3].

Affective engineering studies that use the psychological measurement method are being commonly conducted. Wu et al. [4] evaluated affective words related to smart watches in

DOI: 10.24507/icicel.12.12.1249

order to design smart watches; Mohamed and Mustafa [5] measured affection of vehicles' center stack through affective words; and Pambudi et al. [6] applied affective engineering techniques that used the psychological measurement method in order to design a legless chair that addressed consumer demands. In Korea, Jeong and Lee [7] used the psychological measurement method to develop a tool that can measure the affection that is experienced by consumers while using a product, and Lee et al. [8] extracted affective words that appeared on a three-dimensional web interface and compared them to a two-dimensional interface. Noh [9] derived affective words related to a university library space. Rhiu et al. [10] extracted affective words related to auditory sound. Most of these affective engineering studies that used psychological measurement methods involved the process of collecting large amounts of affective words through literature studies, user feedback data, interviews, and questionnaires, and then selecting an appropriate number of them through a professional review or questionnaire. This method effectively chooses only the required high-quality affective words among words that are used in a specific field. However, a significant amount of time and money is required to extract numerous affective words from various literature sources, recruit professionals and general participants, and then select an appropriate number of words.

This paper proposes an affective word collection method that addresses the issue of the financial and time cost of the affective word collection process for psychological affection measurements. Our method uses the web crawling and text mining techniques, which are used in data information technology. Web crawling enables the collection of large amounts of information in a short period without requiring much effort, while the text mining technique allows the quick extraction of meaningful affective words from a large amount of text data [11]. Therefore, the application of these methods eliminates the requirement to determine individual affection-related words in the literature or use questionnaires or interviews. Thus, the associated time and cost are reduced.

This study proceeds as follows. In Section 2, the framework of the proposed affective word collection method is presented and a brief description of the techniques in use is provided. Section 3 presents a case study that employs the proposed affective word collection method, and Section 4 uses these results to evaluate the performance of the proposed method by comparing them with those of the existing affective word collection method. Lastly, Section 5 summarizes this study and discusses the direction of future studies.

2. **Methods.** It is known that no standard affective word collection process exists for psychological affection measurements. The examination of relevant studies shows that they generally follow the process of collecting affective word candidates, removing inappropriate words and combining similar words, extracting appropriate affective words, and selecting the final words [4-9]. For example, in a study conducted by Jeong and Lee [7], affective words were collected through product usage review data and Think Aloud data regarding a series of product usage tasks in a laboratory environment. Subsequently, duplicate or similar words were removed and two rounds of suitability testing were performed to extract 88 affective words. Then, the 88 affective words were reduced to 32 words through factor analysis.

This affective word collection process is similar to a typical text mining process. The text mining process can be divided into text preprocessing, the text mining operation, and text postprocessing [12]. In the text preprocessing stage, the required data are selected from the collected data, categorized accordingly, and then converted into a form that can be used for analysis. In the text mining operation stage, a text-mining tool is used for text clustering, correlation analysis, pattern analysis, and other analyses that are appropriate for this purpose. In the text postprocessing stage, the results of the analysis are evaluated, selected, interpreted, and visualized. In analogy with the affective word collection process, the text preprocessing stage is similar to the removal of inappropriate words and the combination of similar words that is performed during affective word collection. Furthermore, the text mining operation stage is similar to the extraction of appropriate affective words during affective word collection and the text postprocessing stage is comparable to the final affective word selection stage. Therefore, if the text mining process is applied to fitting the existing affective word collection process, affective word extraction may be possible through text mining.

In this study, the following process is proposed to derive appropriate affective words through text mining. First, to collect affective word candidates, a large amount of topical text data is obtained by using web crawling instead of literature data or interviews. Next, the natural language processing (NLP) tool is used to preprocess the collected data, and word candidates are extracted through the text mining technique. Lastly, the word candidates are evaluated for their suitability as affective words and the final words are selected. Figure 1 shows the detailed process.

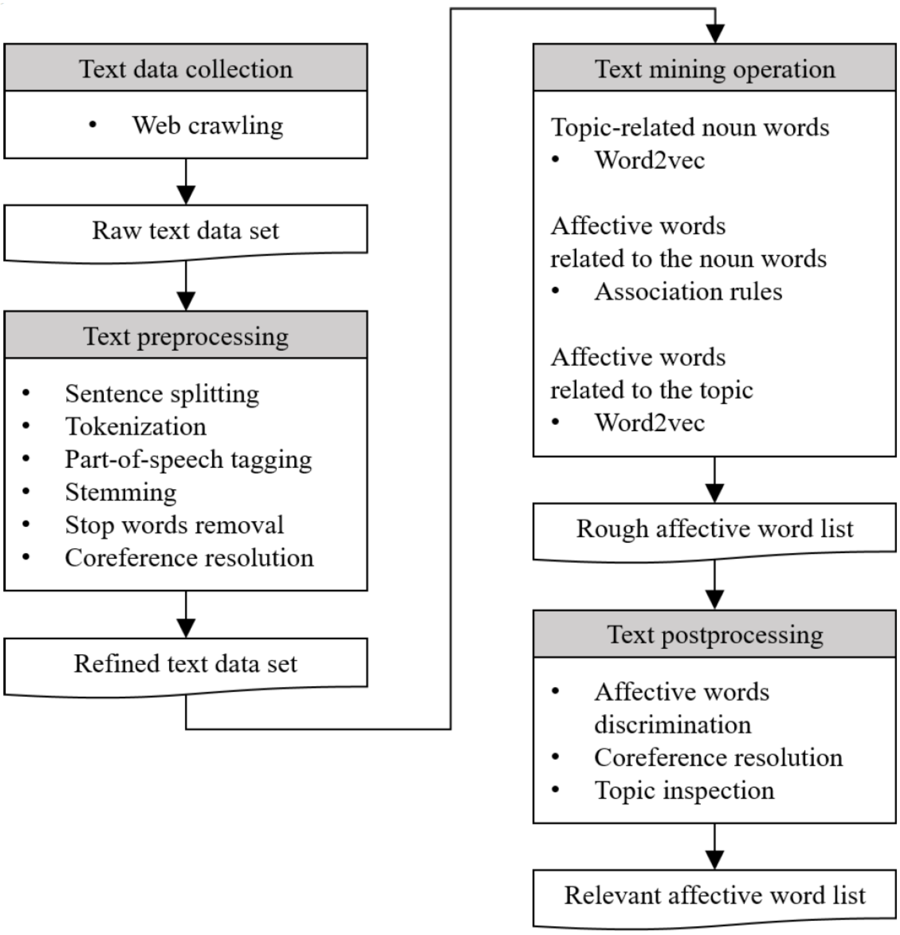

FIGURE 1. A new framework for extracting affective words

2.1. **Text data collection.** To collect affective word candidates, the topics of interest and the relevant literature are investigated to determine which affective words are frequently used. In the proposed method, web crawling is used to quickly collect a large amount of data related to the topic instead of manually investigating the literature and conducting interviews. Web crawling uses a search engine to collect pages from the web and a web crawler to quickly collect large amounts of information related to the topic. Appropriate text data related to the subject of interest are collected from product review boards, blogs, newspaper articles, social media, and other text information sources because a sufficiently large amount of data must be collected to increase analysis reliability.

2.2. **Text preprocessing.** Since the collected data are in a natural language form, the data undergo preprocessing and are refined into single word units to facilitate analysis. Because the objective is to extract affective words related to the topic, the NLP tool is used to conduct word-by-word syntactic analysis for data preprocessing. This process may involve tokenization, sentence splitting, part-of-speech tagging, morphological analysis, named entity recognition, syntactic parsing, coreference resolution, and other annotation tasks [13].

This study performs sentence splitting, tokenization, part-of-speech tagging, morphological stemming, stop-word removal, and coreference resolution. Sentence splitting is performed to divide the range of a word's influence into sentence units. Tokenization is conducted to enable single-word analysis and the extraction of affective words. Part-of-speech tagging is the function of assigning the relevant part of speech to all words; this operation is necessary because 1) it identifies synonyms of different parts of speech and 2) determines adjectives which affective words typically appear. Morphological stemming converts each word into its original form and is performed to prevent derivatives of the same word from being calculated separately. Stop-word removal removes punctuation, symbols, or particles that are not required in the analysis process; this process is performed because it increases calculation speed and ensures that insignificant words do not appear in the results. Because this study's objective was to extract affective words through text mining, numbers, symbols, and the words besides parts of speech such as nouns, verbs, and adjectives that frequently constitute affective words were processed as stop words. Lastly, coreference resolution combines synonyms or antonyms within the collected data into a single word and is performed to prevent affective words with the same meaning from appearing separately in the results. If this process is ineffective, it may be performed during the text postprocessing stage. Antonyms are combined because the affective word that is used in the SD technique creates a pair with the antonym [14].

2.3. **Text mining operation.** Once preprocessing is completed, the data undergo an analysis process that derives affective words with high relevance to the subject of interest. Affective words must be relevant to the subject that is being addressed and must be able to express human senses, impressions, feelings, and emotions relevant to the subject. To achieve this, the meaning of words that are related to the subject must be assessed and the semantic relationship between words must be utilized. This study used Word2vec, which creates a model of the semantic relationship of words [15]. The Word2vec technique places words in a multi-dimensional space to show their semantic relationship. The meaning of each word is quantified by vectors, which can be used to assess the relationship between words. For example, the calculation Vector ("Asia") − Vector ("Rice") + Vector ("Bread") results in Vector ("West"). This method shows the semantic relationship of words that are not shown by existing methods, such as latent semantic analysis [16] or latent Dirichlet allocation [17], where one word corresponds to one vector.

To use the word semantic calculation formula, the keywords, corresponding nouns, and affective words related to these words must be extracted. First, nouns that correspond to the topic and are highly similar to the keywords are extracted. Because a vector value is assigned to each word when Word2vec is used, a meaningful similarity can be obtained through a cosine similarity. Among the words with high similarly, words for which the part of speech was tagged as a noun during preprocessing are identified and extracted.

Next, affective words that are related to the noun are obtained through an association rule analysis. The association rule analysis is to discover a concurrence rule describing how frequently certain items appear simultaneously with other items in the data [18]. This is used to determine affective words that are frequently present when certain nouns exist in each sentence. Because only affective words with a high frequency of appearance are determined, their meaningful relevance may be low; however, obtaining a large number of nouns and relevant affective word combinations to use in the above formula is important at this stage. Affective words are selected from the set of frequency items that have a lift value of at least 1 using the affective word dictionary. When nouns related to the topic and affective words related to the noun are selected, Vector ("Affective word") − Vector ("Noun") + Vector ("Topic word") is calculated for all nouns and affective word combinations, and words with a high cosine similarity are extracted. The overall text mining process is shown in Figure 2.

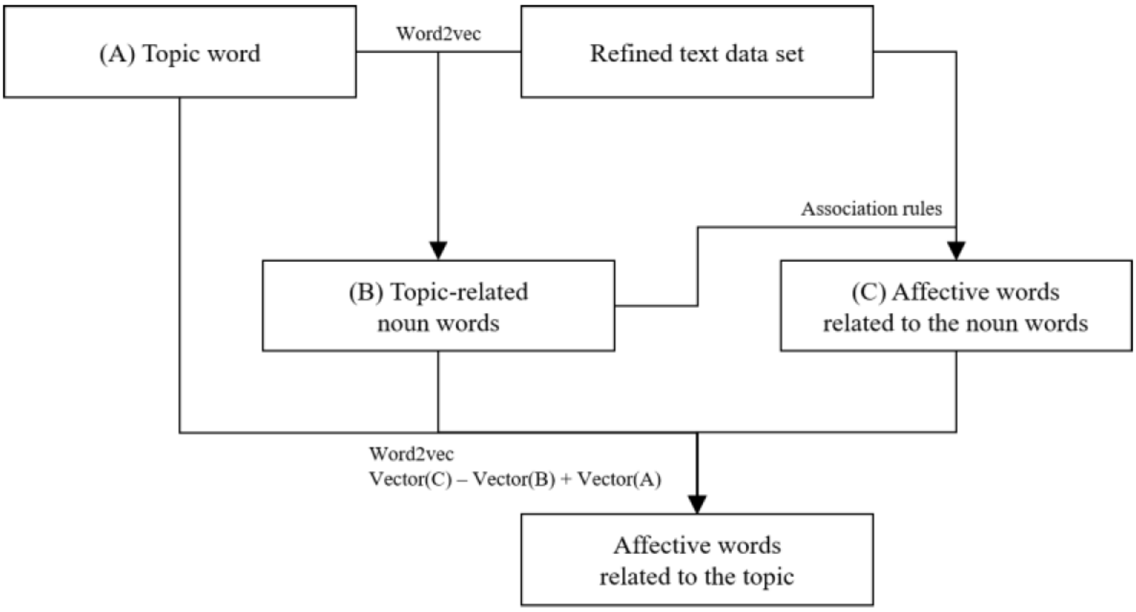

FIGURE 2. A detailed text mining operation process

2.4. **Text postprocessing.** Text postprocessing is simple. First, the process establishes whether the words that were extracted in the previous stages are affective words or not using the affective word dictionary. Additionally, studies related to these affective words are referenced and the affective words are finalized from the extracted words and examined for synonyms and antonyms. Synonyms and antonyms can be confirmed through the synonym dictionary. Lastly, the words are reviewed one last time to ensure that they are relevant to the subject. When the processed words are used to create an affective scale, the process is complete.

3. **Case Study.** An experiment was performed to confirm that the proposed affective word extraction process extracts affective words that are suitable for the subject. The experiment used the proposed method to extract affective words that had already been derived in a previous study. Thirty two representative affective words derived in a study conducted by Jeong and Lee [7] were selected as the existing study data. After removing words with duplicate meanings through the synonym dictionary, 23 words were adopted as the comparison affective words for this study.

3.1. **Text data collection.** Because the keyword was "product", a crawler was created to collect text data from posts and replies on product review boards of electronic product club sites. The posts focused on product usage information. Because of the large number of reviews and descriptions of the specific product, the posts were useful in selecting affective words related to the product.

3.2. **Text data preprocessing.** Part-of-speech tagging and morphological stemming was performed on the collected data by using the twitter-korean-text processor [19,20], and words other than nouns, verbs, and adjectives were processed and removed as stop words. These preprocessed data were analyzed through the Word2vec algorithm and 300 words related to "product" were extracted in the order of the highest cosine similarity.

3.3. **Text mining operation.** The vector calculation principle was used to ensure the relationship between various random nouns and affective words, and this relationship was applied to "product" to selecting affective words related to the product. For this purpose, nouns in the previously collected set that had a high correlation with "product" and affective words that were frequently used to modify these nouns were identified. The continuous bag of words (CBOW) algorithm of Word2vec was used and ten nouns with a high cosine similarity with "product" were collected. Affective words that modified nouns that were obtained in this manner were acquired through the association rule analysis [18]. Affective words that existed in sentences that included each noun were determined, and those with the highest lift value above 1 were given priority for selection. Up to three affective words were collected per noun; if the lift value was too low, fewer or no words were selected.

The relationship between the nouns and the affective words that were collected using the vector calculation principle was examined to assess their relationship to "product". For example, words that were calculated by Vector ("Light") − Vector ("Laptop") + Vector ("Product") to have a high vector and cosine similarity were extracted. Results with more meaningful correlations were derived by using the skip-gram algorithm of Word2vec, and only combinations of nouns and affective words with a cosine similarity above 0.5 were extracted.

3.4. **Text data postprocessing.** The study conducted by Jung and Nah [21] was used as a reference to determine whether the extracted affective words were indeed affective words. After the affective words were selected, duplicate words or words with similar meanings were removed and antonyms were combined. This resulted in 25 final affective words, shown in Table 1.

Table 1. The affective words derived from our new approach

*Pretty	*Cute	*Luxurious	*Disappointing	Strong
Cool	Sturdy	*Neat	*Meticulous	*Remarkable
*Pleasing	*Natural	Warm	*Refined	*Suitable
Ample	Low	*Excellent	Soft	Awful
*Unique	Extravagant	Absolute	Safe	Rewarding

*The words obtained from the study of Jeong and Lee [7]

4. **Assessment.** To verify the proposed affective word extraction method, the performance of the technique was evaluated in comparison with the existing method. For this purpose, it was examined whether the extracted words were affective words related to the subject and the overall grouping of the affective words. The first question provided affective words with the proposed and existing methods and users were asked whether these were affective words that were relevant to the product. Users were asked to answer "yes" if a word was an affective word, and "no" if it was not. The second question provided all the affective words that were obtained with the two methods, and users were asked to state their degree of agreement with the overall affective word list, which provided a score using the Likert 5-point scale. The survey was conducted through social media and on a university's community website, and 64 responses were received.

Regarding the first question, an analysis was performed to determine the difference between the percentage of affective words determined by the present and previous studies. Because the evaluation standards of each participant may differ, the paired t-test was selected as the analysis method. The analysis results showed that there was no difference between the two affective word fractions at a significance level of 0.05 (Table 2). In other words, the data extracted using the proposed method contained as many affective words as the existing method.

Table 2. Paired t-test results for affective word agreement rates

Mean (%)		t value	P value
Ours	Jeong and Lee [7]		
30.0	29.9	0.0772	0.94

Regarding the second question, an analysis was performed to determine the difference between the degrees of agreement regarding the full affective word list for the two data sets. The paired t-test was again used as the analysis method. The analysis results showed that there was no difference in the degree of agreement for the affective word list of the two data sets at a significance level of 0.05 (Table 3). In other words, the data that were extracted using the proposed method were sufficiently accepted as an affective word list compared to the data obtained with the existing method. This confirms that the affective word extraction method proposed in this study can generate an affective word list as suitable as that produced with the existing method.

Table 3. Paired t-test results for affective word list agreement scores

Mean (scores)		t value	P value
Ours	Jeong and Lee [7]		
3.4	3.3	1.2104	0.23

The advantage of the proposed method is that it does not cost more and requires less time to extract affective words than the existing method. The existing affective word extraction method requires significant time and effort because affective word candidates must be determined one by one by searching various literature sources during affective word candidate collection. However, the proposed method crawls through subject-related text data, thus reducing time. Because the existing method extracts suitable words through questionnaires, it requires additional time and cost for the questionnaires. However, because the proposed method performs a computerized analysis and extracts subject-related words using Word2vec, only a small amount of time and effort is required for the analysis. Although computer calculations are time consuming, especially for the large amounts of data typically involved in big data analysis, Word2vec requires less than one day to learn 1.6 million words [15]. On the other hand, questionnaires involve much more time and cost because of the participant recruitment period, questionnaire creation costs, and participant reward costs.

5. **Conclusion.** This study proposes a new affective word extraction method that uses text mining. The existing affective word extraction method involves the process of collecting affective word candidates, removing unsuitable words and combining similar words, extracting suitable affective words, and selecting the final words. On the other hand, the proposed method involves text data collection, which is a general text mining process, text data preprocessing, a text mining operation, and text data postprocessing. Although the two procedures are similar, the tasks involved at each stage are different. A case study was conducted to select affective words with the proposed method using "product" as

the keyword. User evaluations were conducted to confirm the suitability of the extracted affective words by comparing the resulting data with those obtained with the existing method. The results showed that the affective word extraction method proposed in this study could extract suitable affective words related to the subject.

The limitation of the proposed method is that the affective word selection process is complicated. Despite the reduced time and cost, a considerable amount of work is required to select a subject-related word list for tasks such as crawling, text data preprocessing, Word2vec modeling, and association rule analysis. Therefore, the proposed method must be canonicalized. Because NLP for Korean text data preprocessing is incomplete, some data may not be processed correctly, which may also be considered a limitation. However, this can be overcome with advancements in the NLP tool for Korean. Another possible limitation is that the average agreement rate of the first survey question during the evaluation stage was low, approximately 30%, for both this study and the study conducted by Jeong and Lee [7]. This seems to result from the different standards of the evaluation participants for affective word selection, also they tended to judge affective words at a particularly high standard.

The proposed affective word extraction method showed considerably good performance and could reduce time and cost compared to the existing method. Therefore, it can probably be used instead of the existing affective word extraction method in study fields that require affective engineering techniques. Because we perform a case study that uses text mining in affective engineering research, it can also enable the integration of text mining and affective engineering. To enhance the affective word extraction method that is proposed in this study, the method must be normalized to facilitate its application. Other affective word extraction studies that use various text-mining algorithms other than Word2vec must be conducted in the future to create a faster and stronger affective word extraction method.

Acknowledgment. This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (Grant No. 2015R1C1A1A01053529). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] M. Nagamachi, Affective engineering: A new ergonomic consumer-oriented technology for product development, *International Journal of Industrial Ergonomics*, vol.15, no.1, pp.3-11, 1995.

[2] J. W. Kim, *Human Computer Interaction Introduction*, Ahngraphics, Seoul, 2005.

[3] S. Nagasawa, Affective and business, *International Journal of Affective Engineering*, vol.3, no.3, pp.3-12, 2002.

[4] Y. Wu, J. Cheng and X. Kang, Research on the functional relation between affective evaluation and smart watch, *The 8th International Symposium on Computational Intelligence and Design (ISCID)*, pp.287-290, 2015.

[5] M. S. Mohamed and S. Mustafa, Affective engineering implementation on car center stack designs, *International Journal of Education and Research*, vol.2, no.4, pp.355-366, 2014.

[6] A. T. Pambudi, M. R. Suryoputro, A. D. Sari and R. D. Kurnia, Design of lesehan chair by using affective engineering method and anthropometry approach, *IOP Conference Series: Materials Science and Engineering*, vol.105, 2016.

[7] S. H. Jeong and K. P. Lee, Development of tool for measuring the user's emotions expressed while using a product, *Journal of Korean Society of Design Science*, vol.19, no.2, pp.343-354, 2006.

[8] Y. Y. Lee, H. Y. Ryoo and S. H. Jeong, User's emotion while browsing 3D web interfaces, *Korean Society for Emotion and Sensibility*, vol.13, no.4, pp.659-668, 2010.

[9] D. J. Noh, A study on the emotional vocabulary based on space assessment of the academic library, *Journal of the Korean Biblia Society for Library and Information Science*, vol.26, no.4, pp.83-104, 2015.

[10] I. Rhiu, S. Kwon, M. H. Yun and D. C. Park, Analysis of relationship between brand personality and customer satisfaction on a vehicle exhaust sound, *International Journal of Industrial Engineering*, vol.23, no.1, pp.68-82, 2016.

[11] B. Carrera, J. Lee and J. Y. Jung, Discovery of gatekeepers on information diffusion flows using process mining, *International Journal of Industrial Engineering*, vol.23, no.4, pp.253-269, 2016.

[12] Y. Zhang, M. Chen and L. Liu, A review on text mining, *The 6th International Conference on Software Engineering and Service Science*, pp.681-685, 2015.

[13] C. D. Manning, M. S. Sista, J. Bauer, J. Finkel, S. J. Bethard and D. McClosky, The Stanford CoreNLP natural language processing toolkit, *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp.55-60, 2014.

[14] C. E. Osgood, The nature and measurement of meaning, *Phychological Bulletin*, vol.49, no.3, p.197, 1952.

[15] T. Mikolov, G. Corrado, K. Chen and J. Dean, Efficient estimation of word representations in vector space, *Proc. of the International Conference on Learning Representations (ICLR 2013)*, pp.1-12, 2013.

[16] T. K. Landauer, P. W. Folt and D. Laham, An introduction to latent semantic analysis, *Discourse Processes*, vol.25, nos.2-3, pp.259-284, 1998.

[17] D. M. Blei, A. Y. Ng and M. I. Jordan, Latent Dirichlet allocation, *Journal of Machine Learning Research*, vol.3, nos.4-5, pp.993-1022, 2012.

[18] R. Agrawal, T. Imielinski and A. Swami, Mining association rules between sets of items in large databases, *Proc. of the 1993 ACM SIGMOD International Conference on Management of Data*, vol.22, pp.207-216, 1993.

[19] E. L. Park and S. Cho, KoNLPy: Korean natural language processing in Python, *Proc. of the 26th Annual Conference on Human & Cognitive Language Technology*, pp.4-7, 2014.

[20] H. W. Ryu, *Twitter-Korean-Text*, https://github.com/twitter/twitter-korean-text, 2014.

[21] H. W. Jung and K. Nah, A study on the meaning of sensibility and vocabulary system for sensibility evaluation, *Journal of the Ergonomics Society of Korea*, vol.26, no.3, pp.17-25, 2007.