Use of Text Mining Technique in Doing Trend Analysis
of the Internet Articles for Nuclear Energy

So Yun Jeong, Jae Wook Kim, Young Seo Kim, Han Young Joo, and Joo Hyun Moon[*]
Dankook Univ., 119, Dandae-ro, dongnam-gu, cheonan-si, Rep. of Korea, 31116
[*]Corresponding author: jhmoon86@dankook.ac.kr

1. Introduction

Because use of nuclear energy is highly influenced by the public acceptance, it is necessary to identify the people's perception in establishing the policy for the use of nuclear energy [1]. However, there is actually no way to check if the public's first perceptions are kept without changes or how they vary as time goes on. Since the public's perceptions are subject to change by the mass media, such as SNS (Social Networking Service), newspapers and news, the trend analysis of those media could be a useful method to predict the people's perception of nuclear energy [2].

This study analyzed the internet articles posted on 'NAVER', a Korean internet portal site, to figure out the trend of those articles on nuclear energy for the past four years from January 1, 2016 to December 31, 2019. For this, we used a big data analysis program 'R' and performed text mining technique.

2. Methods and Results

Text mining is one of the big data analysis techniques, which is a series of procedure to find meaningful information by extracting interesting patterns or relationships from atypical text in mass media [3]. Fig. 1 shows the analysis procedure. First, we extracted the 15 words related to nuclear energy each year, and selected the major annual keywords. Second, we investigated the relationship between the occurrence date of major nuclear issues and the number of internet articles reported monthly. Finally, we selected positive/negative words and checked the trends of the public opinions.

2.1 Text mining

We chose 'NAVER' as a portal site to be analyzed because it is the portal site Korean people most visited every day. The analysis period was from January 1, 2016 to December 31, 2019. The articles including the words of 'Nuclear Power Plant (NPP)' and 'Nuclear energy' in their titles or contents more than once were extracted. For the period, a total number of the articles including 'NPP' and 'Nuclear energy' was 26,718, and the monthly average number was about 557. The all articles were those for our analysis.

After decomposing the articles into sentences, we removed special symbols and analyzed morpheme by using the Korean morphological analyzer, 'KoNLP' package [4, 5]. Then, we extracted nouns and excluded unnecessary words. Among those, the top 10 words most quoted a month were selected.
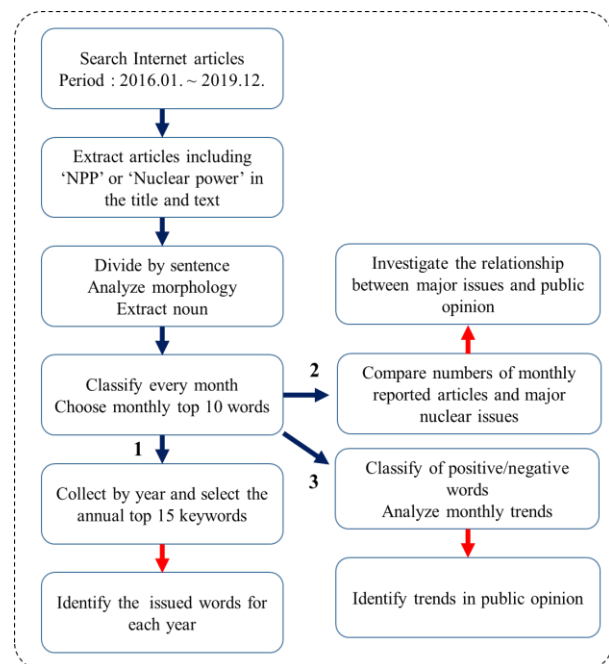

Fig. 1. Schematic diagram of analysis procedure.

2.2 Annual frequency analysis

To figure out the words most quoted a year, we selected the all-time most quoted 15 words for the past four years from the collection of the monthly most quoted 10 words. Table I shows the all-time most 15 keywords from 2016 to 2019. During the period, 'NPP', 'Energy' and 'Power generation' were the words most frequently quoted each year. The order varies year to year, but the words such as 'Earthquake', 'Publicized', 'Denuclearization' and 'Renewable' were along the top ranks. It shows that the words with top ranks were mainly related to the whole nuclear industry or to the social issues raised in the specific year.

Table I: Top 15 words most quoted from 2016 to 2019.

Top15	2016	2017	2018	2019
1	NPP	Energy	Energy	Energy
2	Power generation	NPP	NPP	NPP

3	Energy	Power generation	Problem	Economy
4	Technology	Policy	Power generation	Power generation
5	Development	Problem	Business	Policy
6	Business	Construction	Industry	Technology
7	Occurrence	Business	Policy	Industry
8	Earthquake	Publicized	U.S	Problem
9	Safety	Technology	North Korea	Nuclear
10	Region	Safety	Denuclearization	Safety
11	Scale	Nuclear	Economy	Renewable
12	Problem	Electricity	Technology	Corporation
13	Industry	Discontinue	Corporation	Business
14	Research	Nation	Nuclear	U.S
15	Nuclear	City	UAE	Research

2.3 Review of the relevance to the nuclear issues

Fig. 2 shows the trends in the number of the articles reported a month for the past four years and the occurrence date of the major nuclear issues. This figure showed that sudden increase in the number of the articles about the issue right after occurrence of a specific nuclear issue.

Specifically, in September 2016, NPP safety was the biggest social issue because of the earthquake in Gyeongju. The number of the articles on nuclear energy in 2017 was larger than those in the other years. That was mainly because the new government launched in 2017 made clear 'energy transition policy'. In June 2017, the new president said that plans for new power reactors will be cancelled and the operating periods of existing units will not be extended beyond their design license at the ceremony of the permanent shutdown of the Kori unit 1. In October 2017, when the public deliberation about resumption of the construction of Shinkori units 5 and 6 was ongoing, the number of the articles was surged. Since 2018, on average, 520 articles were released, though there was no hot issue like those events.
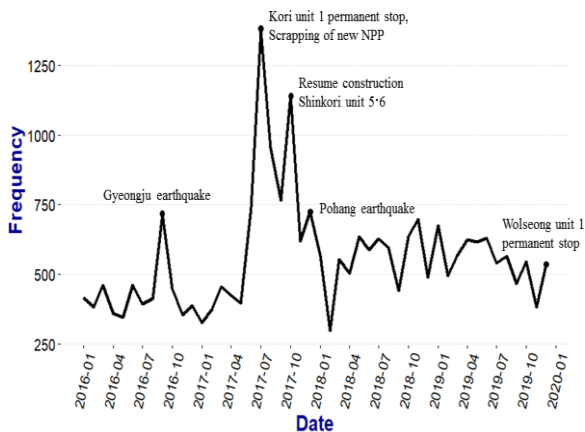
2.4 Trend analysis of the nuances of the nuclear energy related-articles

Each word composing a sentence has various emotional meanings depending on their context [6]. To understand the real intentions, the articles wanted to deliver, we analyzed the nuances of the words extracted previously. After excluding the unnecessary words such as the country name, we classified the remaining words into the 3 groups, positive, negative or neutral, for the nuclear energy as shown in Table II.

Table II: Classification for the nuances of the words.

Classification	Words
Positive	Technology, Economy, Export, Development, Advance, Resumption, Construction, Safety
Negative	Problem, Earthquake, Denuclearization, Disuse, Contamination, Nuclear test, Restriction, Nuclear armament, Discontinuity
Neutral	Policy, Radioactivity

Fig. 3 shows the difference between the numbers of the positive and negative words by month. The monthly average of the positive words was 2.52, and that of the negative words was 0.98. To quantify the overall nuances of the articles, +1 was assigned to the positive words, and -1 was to the negative words. If the same word was repeated more than once in an article, it was regarded as once appeared in the article. If the difference between the two different nuance words is zero, the article is considered as a neutral. And, the net value of the differences between them in the article is positive or negative, then, the article is regarded as positive or negative nuance.
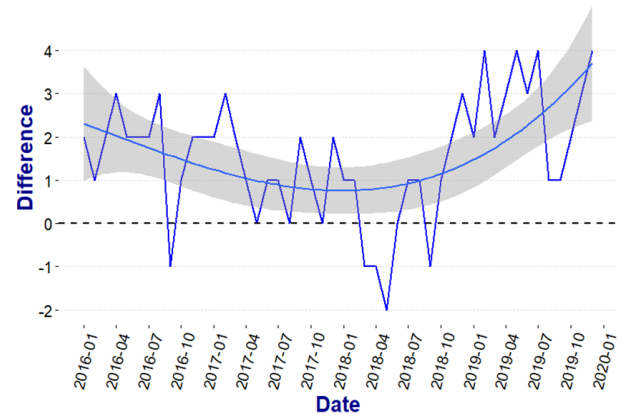

Fig. 2. Number of the articles including 'Nuclear energy' and 'NPP' released each month.

Fig. 3. Overall nuances of the nuclear energy related-articles released each month.

3. Conclusions

This study analyzed the internet articles posted on 'NAVER', a Korean internet portal site, to figure out the trend of those articles on nuclear energy for the past four years from January 1, 2016 to December 31, 2019 by month. For this, we used the text mining technique.

As a result, we identified the top 15 words most quoted in the article, as shown in Table I. The most quoted words were 'NPP', 'Energy', and 'Power generation', etc. These words were mainly related to the nuclear issues in each year. Second, we found surge in the number of the articles following the occurrence of the major nuclear issues. Finally, we found that, as shown in Fig.3, the articles with positive nuance have been more released, though the articles with negative nuance were more released in the early phase of the new government.

Acknowledgements

REFERENCES

[1] K. R. Lee, News trend frame analysis on nuclear power plant issues -A topic modeling and semantic network analysis approach-, Sungkyunkwan Univ., 2018.

[2] K. D. Ham, Social recognition on a multicultural minority group portrayed by mass media: mews article analysis through text mining, Korea Univ., 2015.

[3] E. B. Song, Analysis of news on bitcoin using text mining, Ewha Womans Univ., 2017.

[4] H. W. Jeon, KoNLP: Korean NLP Package. R Package Version 0.80.2, https://github.com/haven-jeon/KoNLP, 2016.

[5] D. S. Kim, J. W. Kim, Public Opinion Mining on Social Media: A Case Study of Twitter Opinion on Nuclear Power, Advanced Science and Technology Letters, Vol.51, pp.224-228, 2014.

[6] N. Farra et al, Sentence-level and Document-level Sentiment Mining for Arabic Texts, IEEE international conference on data mining workshop, pp.1114-1119, 2010.