

Conflict Analysis in Construction Project with Unstructured Data: A Case Study of Jeju Naval Base Project in South Korea

Seungwon Baek^{1*}, Seung Heon Han¹, Changjun Lee¹, Woosik Jang², Jong Ho Ock³

¹ Department of Civil and Environmental Engineering, Yonsei University, Seoul 03722, South Korea

² Department of Civil Engineering, Chosun University, Gwangju 61452, South Korea

³ Department of Architectural Engineering, Seoul National University of Technology, Seoul 01811, South Korea

E-mail address: baeksw@yonsei.ac.kr

Abstract: Infrastructure development as national project suffers from social conflict which is one of main risk to be managed. Social conflicts have a negative impact on not only the social integration but also the national economy as they require enormous social costs to be solved. Against this backdrop, this study analyzes social conflict using articles published by online news media based on web-crawling and natural language processing (NLP) techniques. As an illustrative case, the Jeju Naval Base (JNB) project which is one of representative conflict case in South Korea is analyzed. Total of 21,788 articles and representative keywords are identified annually. Additionally, comparative analysis is conducted between the extracted keywords and actual events occurred during the project. The authors explain actual events in the JNB project based on the extracted words by the year. This study contributes to analyze social conflict and to extract meaningful information from unstructured data.

Key words: Social conflict, Jeju Naval Base project, Unstructured data, Web-crawling, Natural language processing (NLP)

1. INTRODUCTION

South Korea recorded higher social conflict level than that of OECD countries' average [1]. Social conflicts have a negative impact on not only the social integration but also the national economy as they require enormous social costs to be solved. In particular, the number of social conflicts occurred in construction projects has increased continually [2]. The JNB case is one of the representative project that social conflict occurred in South Korea. In 2007, public opinion polls were conducted twice and the Jeju local government approved the JNB construction project at Gangjeong village (Event #1 in Figure 1). Although government made efforts to arrange with residents through compensation agreement, conflict had deepened and a public recall vote was held to claim responsibility to the governor of Jeju-do in 2009 (Event #2 in Figure 1). The construction was initiated in 2010. Various parties such as politics, environment and religious protested against construction execution. Conflict level increased rapidly as citizens and parties were accused and conflict had spread out nationally. As a result, the construction temporally stopped in May, 2011 and resumed in Sep., 2011 (Event #3 in Figure 1). The number of articles reached its peak in 2012 with blasting of the Gureombi rock which is a symbol of environment preservation (Event #4 in Figure 1). In addition, the 19th general election and the 18th presidential election were held in 2012 (Event #5 and #6 in Figure 1) and each political party suggested commitment related to the JNB project. In this period, the JNB project were expressed in the news steadily. The number of articles had suddenly increased in 2015 due to implementation of administrative vicarious execution against the construction opposite assembly (Event #7 in Figure 1). Lastly, the JNB construction was completed in 2016 (Event #8 in Figure 1).

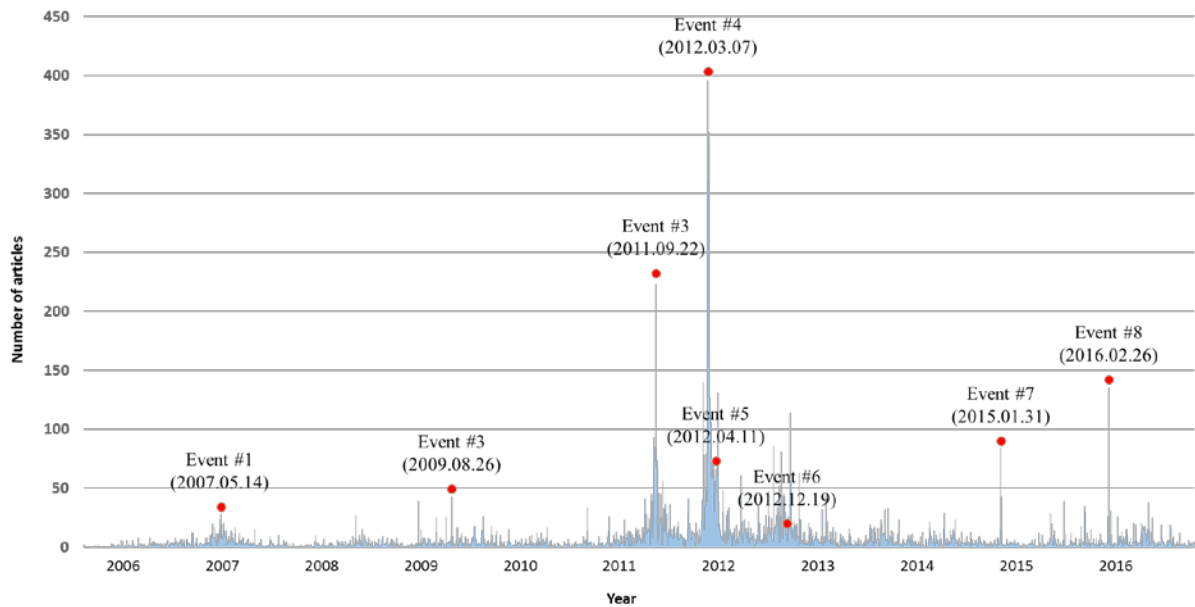


Fig. 1. Published articles related to the JNB construction project

2. METHODOLOGY

2.1. Web-crawling

As the internet and smartphone have proliferated, present generation constantly takes a lot of information day by day. Internet news is one of main information sources that modern people can easily access. Because human perception is limited to available information, news contents have significant impact on forming public opinion [3]. This study employed web-crawling technique to acquire extensive news article data. Web-crawling is a method that connects to web pages for gathering and storing specific information in a local database automatically [4].

The JNB case is one of the representative project that social conflicts occurred in South Korea. Thus, there is a lot of content in Korean, but contents in other languages are insufficient at both quantitative and qualitative perspective. Against this backdrop, the authors determined to limit source of news articles to South Korea and portal site ‘NAVER’ (www.naver.com) was selected as search engine. NAVER has provided news service since 2000 and it has the largest market share by 55.4% in South Korean portal sites that provide internet news service [5], and NAVER currently has a partnership with 208 news media in South Korea.

Table 1 shows crawling conditions used in this research. Published year was limited from 2006 to 2016 because a task force team for impact analysis of the JNB construction was set up in 2006 and the construction project was completed in 2016. In addition, the political tendency of article varies depending on news media and each of articles describe a single event in various point of view [6]. With the purpose of minimizing political bias, all news media that provide news articles through NAVER were targeted as crawling pages.

Table 1. Web-crawling condition

Search word	Jeju Naval Base (In Korean: 제주해군기지)
Document type	News Article
Published year	2006 - 2016 (11 years)
Crawling target	208 news media that have partnership with NAVER
Language	Korean

2.2. Natural language processing (NLP)

Natural language processing (NLP) is a set of various language techniques for analyzing linguistic structure and extracting information from unstructured data [7]. In this research, NLP is performed in two steps; tagging words and word vectorization. First, a body of article breaks down into words and the extracted words are tagged with morpheme. Because the purpose of this study aims to identify representative keywords, only nouns were extracted and other morphemes were excluded. Then, frequency of each word appeared in a document is counted. Each word and frequency is stored as a set of data for word vectorization. Second, word vectorization is performed in order to transform unstructured data to structured data. In this research, the bag-of-words (BoW) concept is applied for word vectorization as shown in Figure 2. The BoW represents a document as a bag of textual words in matrix form to make a computer recognize unstructured data. As a NLP tool, KoNLPy which is an open source Python package for Korean language is used [8].

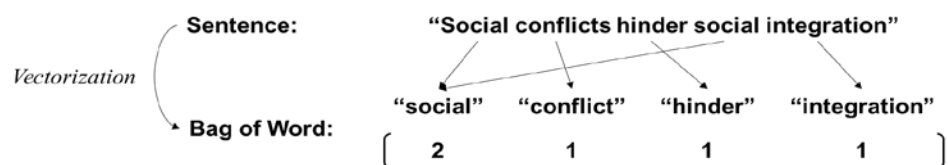


Fig. 2. Example of word vectorization

3. RESULT

3.1. Overall

From the web-crawling process, total of 29,112 news articles were collected from 93 news media. As a preprocessing, the authors removed redundant articles such as photo news without text, duplicate articles. Finally, 21,788 news articles were used for NLP. Table 2 shows the number of posted articles by year. Specifically, articles related to the JNB project rapidly increased in 2011 when the construction was stopped because of civil protest. This event attracted interest of not only news media but also citizens over the country. In particular, the JNB project had been issued by many politicians while the 19th general election and the 18th presidential election in South Korea. After the elections, the number of published articles relatively decreased. In light of news media's characteristic, a certain event that attracted journalist's interest become an article in the news media. In other words, the number of published articles represent how many citizens are interested in the JNB project and it is somewhat correlated to conflict level.

Table 2. The number of posted articles by year

Year	2006	2007	2008	2009	2010	2011
Num	375	1,241	547	1,014	713	3,743
Year	2012	2013	2014	2015	2016	Total
Num	8,030	2,311	1,220	1,126	1,468	21,788

As a result of NLP, total of 58,779 words were identified for whole period (2006 – 2016). Obviously, 'Jeju' was the most frequent word with 84,453 times followed by 'Naval Base' with 61,698 times because those two words were the search word for web-crawling. Table 3 represents 30 most frequent words. High ranked keywords are representative of the JNB conflict case generally. Based on those words, the JNB conflict case is summarized as follow. There was a *problem* during the *JNB construction project at Gangjeong village. Residents opposed to the navy to preserve Jeju-do as a peace island.* The conflict attracted interest of *citizen* over the whole country. Many politicians such as *congressman* and *candidates for the presidency* mentioned the *JNB* .

Some non-nouns words such as ‘about’, ‘our’ and ‘for’ were ranked in high position of frequency analysis because these words have a homonym which is a noun in Korean. In case of ‘about’, it simultaneously means ‘Korea’ and ‘our’ has another meaning that stands for ‘cage’. These homonyms need to be distinguished to extract frequency precisely in a future study.

Table 3. The top 30 words in order of frequency

Rank	Keyword	Frequency	Rank	Keyword	Frequency	Rank	Keyword	Frequency
1	Jeju*	84,453	11	Execution	22,669	21	President	15,593
2	Naval Base	61,698	12	Region	20,839	22	South Korea	15,249
3	Gangjeong*	39,966	13	About	20,765	23	Related to	15,059
4	Village	37,577	14	Navy	20,461	24	Seoul	14,913
5	Resident	33,857	15	Peace	20,386	25	Citizen	14,692
6	Candidate	31,327	16	Project	18,677	26	For	14,632
7	Construction	26,706	17	Our	17,186	27	Police	13,823
8	Jeju-do*	25,980	18	Congressman	16,997	28	Last	13,657
9	Problem	25,061	19	Opposition	16,679	29	Politics	13,240
10	Government	24,830	20	People	15,651	30	Saenuri Party**	12,467

* Name of a local area in South Korea

** Political party in South Korea

3.2. Frequency analysis

The JNB project has three stages in terms of conflict; pre-construction stage, construction stage and post-agreement stage. Pre-construction stage is from 2006 to 2009. Main conflict parties at pre-construction stage were residents and the governor of Jeju local government. During the pre-construction stage, conflict was limited in Jeju-do. Residents of Jeju-do had a dispute over whether accept or not the proposal of government and navy to construct the naval base in own village. Opinion polls were conducted twice and the results showed that there were more favorable votes than the opposite, however, residents who oppose to the JNB construction raised an objection that the governor of Jeju manipulated public opinion. Consequently, the public recall vote was held to claim responsibility to the governor of Jeju-do who decided to attract the naval base. Against this backdrop, identified keywords in pre-construction stage has close connection with resident opinion for attracting the JNB and the opinion pool as shown in Table 4. Negative words such as ‘Problem’, ‘Opposition’, ‘Recall’ and ‘Vote’ were extracted, however, any positive word was not frequently used in articles. This indicates that negative public opinion to the project were formed by news media whereas legitimacy and necessity of the naval base in Jeju were not appropriately delivery to the public at the early phase of the project.

Table 4. Identified keywords during the pre-construction stage

Rank	2006	2007	2008	2009
1	Region	Jeju-do*	Jeju-do*	Resident
2	Jeju-do*	Resident	Village	Recall
3	Islander	Islander	Gangjeong*	Jeju-do*
4	Construction	Region	Resident	Governor
5	Jeju Province*	Construction	Construction	Vote
6	Peace	Peace	Government	Region
7	Plan	Village	Business	Construction
8	Seoul	Seoul	Region	Islander
9	Problem	Opposition	Islander	Business
10	Governor	Daily-news	Promotion	Village

* Name of a local area in South Korea

Conflict level rapidly increased during the construction stage and especially it got into peak in 2011 and 2012. Residents who opposed to the construction protested and eventually the construction execution was stopped in May, 2011. Many trials related to the JNB construction were submitted and dispute between the government and residents were severe. The keyword ‘police’ appeared in the first time in 2011 and it represents the suppression by the government. In addition, some political related words were identified in 2012 due to the 19th general election and the 18th presidential election such as ‘Candidate’, ‘Saenuri Party’ and ‘President’ as shown in Table 5. As many politicians mentioned the JNB project during campaign. Consequently, budget compilation to the JNB project was issue in the national assembly as ‘Budget’ recorded the 10th frequent word in 2013.

Table 5. Identified keywords during the construction stage

Rank	2010	2011	2012	2013
1	Resident	Gangjeong*	Candidate	Gangjeong*
2	Jeju-do*	Village	Gangjeong*	Village
3	Governor	Resident	Village	Construction
4	Village	Execution	Problem	Congressman
5	Gangjeong*	Police	Government	Government
6	Business	Construction	Construction	Problem
7	Construction	Peace	Saenuri Party**	Our
8	Region	Jeju-do*	President	About
9	Candidate	Navy	Execution	Resident
10	Islander	Government	Citizen	Budget

* Name of a local area in South Korea

** Political party in South Korea

In 2013, the civilian-military agreement was signed to use the JNB as a complex port for both public and military purpose and conflict level seemed to be decline. The keyword ‘pope’ recorded as the 4th frequent word in 2014 because Pope Francis had visited in South Korea and the pope mentioned the JNB project. Korean Catholic officially opposed to the JNB project and the pope encouraged priests who participated in the opposition movement. As the JNB construction was terminated in 2016, identified keywords were related to other events that linked to social conflict issue. Because the JNB conflict case was mentioned in articles that describe other social conflicts rather than the JNB project. Especially, 2016 was conflict era in South Korea due to the presidential impeachment issue. Representative conflict event such as Sewol ferry case, THAAD (Terminal High Altitude Area Defense) issue were mentioned by news media along with the JNB project as shown in Table 6.

Table 6. Identified keywords during the post-agreement stage

Rank	2014	2015	2016
1	Candidate	Gangjeong*	Resident
2	Gangjeong*	Village	Gangjeong*
3	Village	Resident	Village
4	Pope	Navy	Navy
5	Korea	Execution	Sewol-ferry
6	Resident	Peace	Government
7	Problem	Problem	Region
8	About	Our	Problem
9	People	Opposition	THAAD
10	Our	About	Jeju-do*

* The name of a local area in South Korea

4. CONCLUSIONS

This study attempted to extract meaningful information through web-crawling and NLP based on news articles as big data. As an illustrative case, the JNB project executed by Korean government from 2006 to 2016 was selected. Total of 21,788 articles were collected and representative keywords were identified by the year. As a result, frequent keywords were somewhat related to events that occurred at that period. The authors described relationship between extracted words and actual events that affected the JNB project. Extracted keywords were analyzed in comparison with actual events qualitatively. At pre-construction stage, negative words toward the JNB project appeared in most frequent keywords, however, any positive word rarely identified. In other words, government and navy didn't respond to residents' requirement properly at initial stage of the project. During construction stage, the JNB project was used as a political issue by congressman who had the general and presidential ahead. Because of the elections, not only the number of articles were rapidly increased, but also most high-ranked words in order of frequency in 2012 were related to politics. Lastly, the JNB project was mentioned by news media along with other national event or social conflict problems during the post-agreement stage. Recently, South Korea had severely suffered from social conflict in 2016 and the JNB conflict case was introduced as representative conflict case in most articles. Although this study suggested correlation between identified word and actual event qualitatively based on frequency analysis, it is limited in extracting meaningful information through quantitative analysis. In a future study, the authors will analyze interrelation between words based on word embedding quantitatively.

ACKNOWLEDGEMENTS

This work was supported by a Korea Science and Engineering Foundation (KOSEF) grant funded by the Korea government (MOST) (NRF-2015R1A2A1A09007327).

REFERENCES

1. Lee, B. H., and Park, Y. J. (2016). "Estimation of economic effects of social conflict and implications." *Hyundai Research Institute*, 718(45), 1-12. (In Korean)
2. Tazelaar, F., and Snijders, C. (2010). "Dispute resolution and litigation in the construction industry. Evidence on conflicts and conflict resolution in The Netherlands and Germany." *Journal of Purchasing and Supply Management*, 16(4), 221-229.
3. Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., and Ngo, D. C. L. (2015). "Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment." *Expert Systems with Applications*, 42(1), 306-324.
4. Kovacevic, M., Nie, J.-Y., and Davidson, C. (2008). "Providing answers to questions from automatically collected web pages for intelligent decision making in the construction sector." *Journal of Computing in Civil Engineering*, 22(1), 3-13.
5. Opinion Concentration Investigation Committee (2015). "Opinion concentration investigation report." Korea Ministry of Culture, Sports and Tourism. (In Korean) <https://www.mcst.go.kr/web/s_data/research/researchView.jsp?pSeq=1602> (May 15, 2017).
6. Druckman, J. N., and Parkin, M. (2005). "The impact of media bias: How editorial slant affects voters." *Journal of Politics*, 67(4), 1030-1049.
7. Bird, S., Klein, E., and Loper, E. (2009). *Natural language processing with Python: analyzing text with the natural language toolkit*, O'Reilly Media, Inc."
8. Park, E. L., and Cho, S. "KoNLPy: Korean natural language processing in Python." *In Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology*, 133-136.