**The 6ᵗʰ International Conference on Construction Engineering and Project Management (ICCEPM 2015)**
Oct. 11 (Sun) ~ 14 (Wed) 2015 • Paradise Hotel Busan • Busan, Korea
www.iccepm2015.org

# Big Data Analytics Applied to the Construction Site Accident Factor Analysis

Joon-soo KIM[1] Ji-su Lee[2] Byung-soo KIM[3]

***Abstract:** Recently, safety accidents in construction sites are increasing. Accordingly, in this study, development of 'Big-Data Analysis Modeling' can collect articles from last 10 years which came from the Internet News and draw the cause of accidents that happening per season. In order to apply this study, Web Crawling Modeling that can collect 98% of desired information from the internet by using 'Xml', 'tm', "Rcurl" from the library of R, a statistical analysis program has been developed, and Datamining Model, which can draw useful information by using 'Principal Component Analysis' on the result of Work Frequency of 'Textmining.' Through Web Crawling Modeling, 7,384 out of 7,534 Internet News articles that have been posted from the past 10 years regarding "safety Accidents in construction sites", and recognized the characteristics of safety accidents that happening per season. The result showed that accidents caused by abnormal temperature and localized heavy rain, occurred frequently in spring and winter, and accidents caused by violation of safety regulations and breakdown of structures occurred frequently in spring and fall. Plus, the fact that accidents happening from collision of heavy equipment happens constantly every season was acknowledgeable. The result, which has been obtained from "Big-Data Analysis Modeling" corresponds with prior studies. Thus, the study is reliable and able to be applied to not only construction sites but also in the overall industry.*

***Keywords:** Big-data, R, Web Crawling, Datamining, Textmining*

## 1. INTRODUCTION

Along with the advancement of information & communications technology, 'Big Data' is receiving attention. Since 2007, the era of data flooding began in which the amount of data generated across the world surpasses the applicable storage volume. It is also estimated that the amount of data will exponentially increase by 50 times compared to current amount by 2020. Accordingly, 'Big Data' referred to as the flooding and sharp increase of data has become a main topic throughout society rather than being restricted to a particular area. In accordance with such trend, large volumes of information are being required and generated in construction sites that are becoming larger in scale and complex in nature while requiring specialization. Accordingly, there is an increasing interest in developing systems to manage and process such information and data. Using R, a statistical analysis program, this paper aims to collect Internet news on construction site safety accidents that occurred during the last decade from the SBS NEWS website to deduce main types of construction site accidents and season-specific accident type factors based on text mining & data mining. In addition, the method used in this paper will be verified by comparing with existing method.

## 2. CASE STUDY

Safety accident. In a study on safety accident analysis, H

. S. Jin et al. analyzed the effects of workers' individual characteristics on safety behavior to establish a comprehensive modeling path of in/direct effects between individual characteristics and safety behaviors [2]. D. W. Lee et al. developed a safety management information system for each stage from planning and design stages to bidding, construction and post-construction stages [1].

In recent, a USN-based construction site safety management monitoring system applicable to construction site has been developed by integrating IT [5]. Additionally, learn-centric learning tools that can be used for university-level construction safety education have been developed. Since there has not been any case study conducted on construction site safety accident using Big Data analytical method, the significance of this study can be found in that it presents a new case study methodology.
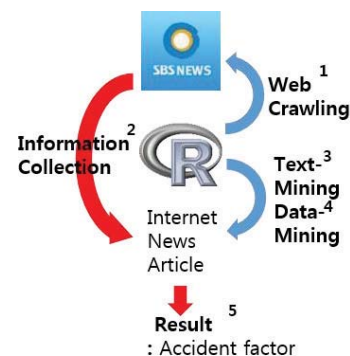


Figure 1: Big-data Analysis Process

[1] Big Data Analytics Applied to the Construction Site Accident Factor Analysis, 80 Daehakro, Bukgu, Daegu, 702-701, Korea, kimjoonsoo@knu.ac.kr
[2] Big Data Analytics Applied to the Construction Site Accident Factor Analysis, 80 Daehakro, Bukgu, Daegu, 702-701, Korea, dlwltnaa@naver.com
[3] Big Data Analytics Applied to the Construction Site Accident Factor Analysis, 80 Daehakro, Bukgu, Daegu, 702-701, Korea, bskim65@knu.ac.kr (*Corresponding Author)

**The 6ᵗʰ International Conference on Construction Engineering and Project Management (ICCEPM 2015)**
Oct. 11 (Sun) ~ 14 (Wed) 2015 • Paradise Hotel Busan • Busan, Korea
www.iccepm2015.org

### 3.2 Information Collection

From the SBS NEWS website, 7,384 out of 7,534 Internet news articles, which represent 98%, that have been posted during the period of Jan. 2005 – Oct. 2014 regarding 'construction site' were collected.

Table 1 : Collection Result

| Year | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|
| Article | 491 | 667 | 606 | 619 | 600 | 644 | 1,019 | 832 | 1,127 | 917 |
| Collection | 485 | 667 | 602 | 616 | 595 | 641 | 1,017 | 826 | 1,030 | 905 |
| % | 98.7 | 100 | 99.3 | 99.2 | 99.3 | 99.5 | 99.8 | 99.2 | 91.3 | 98.6 |

### 3.3 Text Mining

For data analysis, preprocessing is required for the Internet articles collected. In the 7,384 articles collected, sentence components such as conjunction, predicate and adverb are included. These components are considered as unnecessary noise for the analysis and the task of filtering by using the library of R and the task of extracting only words were performed.

Table 2 : Key word

| Word | Fire | Collapse | Explosion | Crash | Heavy Equipment | Suffocation |
|---|---|---|---|---|---|---|
| Counting | 1,114 | 903 | 439 | 438 | 355 | 78 |

Among the 7,384 articles collected regarding construction site safety accident, the frequency of mentioning 'fire' is 1,114 times representing 15.0%. As for the frequency of mentioning 'collapse' resulting from ground subsidence and structural collapse, it was 903 times, which represent 12.2%. In terms of the frequency of mentioning 'explosion' and 'crash', it was respectively 439 and 438 times, representing 5.9%. As for 'heavy equipment', frequency was 355 times with 4.8% and 'suffocation' was 78 times with 1.0%. The result shows fire and collapse represented 27.2% of total number of safety accidents in construction site as most frequently occurring safety accidents.

### 3.4 Data Mining

With the six major accident types deduced from text mining, search, web crawling and text mining were performed for each season from the SBS NEWS website to save the results as text file. In the data saved, both necessary and unnecessary information for this study exist. Accordingly, it requires a filtering process of extracting only necessary information for this study. Accordingly, 168 words with high frequency and highly related to safety accident were deduced and filtering was performed for the collected text file to identify the frequency of corresponding words. In addition, main season-specific accident type factor-analysis and principal component analysis were performed.
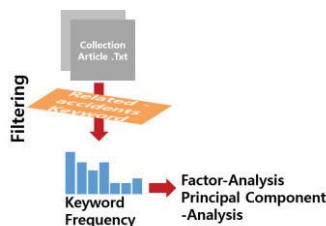


Figure 2 : Data-mining Process

### 3.5 Result

The result of performing principal component analysis of Internet articles regarding safety accidents for each season showed that the weather factor of abnormal temperature and localized heavy rain has significant effects in winter and summer even with the same accident type. In spring and fall, violation of safety regulations and construction process beyond a reasonable level have significant effects on safety accidents. In addition, common factors such as heavy equipment collision, defective structure and faulty construction are highly related to construction site safety accident.
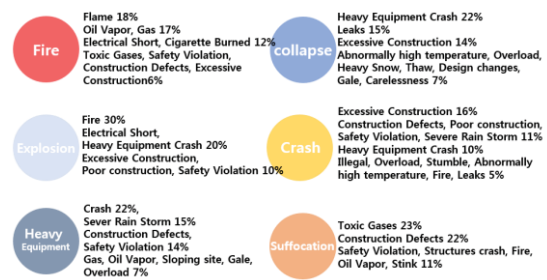


Figure 3 : Accident Factor

### 3. CONCLUSION

Based on the analysis of 7,384 Internet articles related to safety accidents that occurred at construction site during the last decade, this study deduced relative importance of types of accidents frequently occurring at construction site, as well as season-specific accident type factors through principal component analysis. The safety accident factors deduced in this study are not very different from results that can be obtained from previous studies. This is a proof that the methodology used in this study is effectively. The fact that it can deduce results similar to that obtained by using existing method proves that the methodology of this study, which is the application of text & data mining methods through the Internet safety accident article search, is valid. For a follow-up study, it is expected that more reliable results can be obtained by performing data mining upon adding descriptive variables such as temperature, humidity, weather and construction market growth rate to collected articles on construction site safety.

### REFERENCES

[1] D. W. Lee, Y. S. Kim, "A Study on the Cost risk Analysis for Construction Projects using fuzzy-AHP Method", AIK, Vol. 9, No. 4, pp.169-176, 2003

[2] H. S. Jin, D. E. Lee, "Developing the Path Model defining the Relationship between Construction Worker's Personal Characteristics and Safety Behaviors", KIC, Vol. 12, No. 2, pp.169-180, 2013

[3] S. H. Kwon, J. H. Lee, D. H, Cho, J. W. Kang "Design and Implementation of a High Performance Web Crawler", KSIIS, Vol. 8, No.4, 2003

[4] D. M. Seo, H. M. Han, "Intelligent Web Crawler for Supporting Big Data Analysis Services", IJOC, vol. 13, No.12, pp.575-584, 2013

[5] K. T. Kim, "A Study on the Implementation of USN Technologies for Safety Management Monitoring of Architectural Construction Sites, KIC, Vol. 9, No. 4, pp.103-109, 2009