

# A Proposal of Unstructured Document-based Safety Management Approach in Building Construction Projects

Sang Hyeong JEON<sup>1</sup>, Seung Ju WON<sup>2</sup>, Yoon Seok SHIN<sup>3</sup>, Wi Sung YOO<sup>4\*</sup>

<sup>1</sup> *Department of Architectural Engineering, Kyonggi University, Republic of Korea*, E-mail address: bro7827@kyonggi.ac.kr.

<sup>2</sup> *Department of Architectural Engineering, Kyonggi University, Republic of Korea*, E-mail address: wsj970406@kyonggi.ac.kr.

<sup>3</sup> *Department of Architectural Engineering, Kyonggi University, Republic of Korea*, E-mail address: shinys@kyonggi.ac.kr.

<sup>4</sup> *Department of Construction Economic & Finance Research, Construction & Economy Research Institute of Korea, Republic of Korea*, E-mail address: wsyoo@cerik.re.kr.

**Abstract:** About 70% of the data generated on building construction sites consists of unstructured data, such as text, photos, videos, etc. However, the text data, which constitutes the largest proportion of unstructured data, has been restrictively utilized. When using standardized data to evaluate safety performance, there are a few difficulties in addressing issues such as lack of data, omissions, and errors. This copes with limitations on the practical evaluation of safety performance on building construction sites. Despite generating extensive text-centric documents, the previous researches on evaluating safety performance levels using unstructured data are still in its infancy. This study proposes a framework for evaluating the safety performance by preprocessing and refining text-based construction supervision documents. In this framework, relevant keywords related to safety performance are extracted from supervision documents, tokenized, and analyzed for association rules among keywords. Based on the results of the association rule analysis, keywords are selected, and the unsatisfactory or satisfactory level of safety performance is quantified using logistic regression analysis, considering the frequency of their occurrence. While the proposed framework focuses on quantifying the safety performance levels of construction sites, it can be expanded to implement integrated performance diagnostics on-site by linking with tools that evaluate diverse performance levels. This extension will allow for a comprehensive assessment of on-site performance. Furthermore, the framework can serve as a tool supporting practical and proactive inspections and responses of safety managers by utilizing unstructured data alongside the traditional approach focused on standardized data for safety performance assessment.

**Key words:** Safety Performance, Unstructured Data, Association Rule Analysis, Logistic Regression Analysis

## ACKNOWLEDGEMENTS

This study is a basic research conducted in 2023 with the support of the Korea National Research Foundation with the funding of the government (Ministry of Education) (2021R1A2C2013841).

## REFERENCES

- [1] Sung, Y.K, Hur, Y.K, Lee, S.W, Yoo, W.S. "Development of Performance Indicators on Private Building Construction Sites using Supervisory Report." Korean Journal of Construction Engineering and Management, KJCEM, Vol. 23, no.6, PP.65-75. 2022.
- [2] Kim, J.Y, Ahn, S.E. "Developing a Bayesian Network Model for Real-time Project Risk Management." Korean Institute of Industrial Engineers, vol.2010, no.11, pp.970-978. 2010.