

텍스트마이닝을 활용한 건설현장 재해 유형별 예방 대책 분석

Analysis of Prevention Methods by Type of Construction Disaster
Using Text Mining Techniques조규필¹ · 이명도² · 신윤석³ · 김백중^{4*}Gyu Pil Jo¹, Myungdo Lee², Yoon-seok Shin³, Baek-Joong Kim^{4*}¹Researcher, Division of Architectural Engineering, Kyonggi University, Suwon, Republic of Korea²Professor, Department of Architectural Engineering, Jeju National University, Jeju, Republic of Korea³Professor, Division of Architectural Engineering, Kyonggi University, Suwon, Republic of Korea⁴Professor, Division of Fire Fight Safety and Management, Seojeong University, Yangju, Republic of Korea

*Corresponding author: Baek-Joong Kim, bjkim0714@seojeong.ac.kr

ABSTRACT

Purpose: This study provides prevention methods by type of construction disaster using text mining techniques. **Method:** Based on the database that analyzed the cases of critical disasters in the domestic construction sector, preventive measures and causes are analyzed by text mining techniques, and the contents of the analysis are visually shown. **Result:** This visual data represents the measures for preventing critical disasters of each process according to the importance. **Conclusion:** It is believed that the results will be helpful in identifying factors to be considered in preparing preventive measures for serious accidents in construction.

Keywords: Construction site, Major Disaster, Disaster Prevention, Text mining

요약

연구목적: 본 연구는 텍스트 마이닝 기법을 활용하여 유형별 건설재해의 원인을 도출하여 중대재해 사고의 예방대책 마련을 위한 주요 요소를 파악하는 것을 목적으로 한다. **연구방법:** 국내 건설분야의 중대 재해 사례를 분석한 데이터베이스를 기반으로 예방대책과 원인을 텍스트 마이닝 기법으로 분석하고, 분석 내용을 시각적으로 표현하였다. **연구결과:** 이 시각적 데이터는 중요도에 따라 공종별 중대 재해 예방에 필요한 요소의 파악을 용이하게 한다. **결론:** 본 연구의 결과는 건설현장 중대재해와 관련하여 예방 대책 마련 시 고려되어야 할 요소 및 요소간 명확한 상관관계의 파악에 기여할 것으로 사료된다.

핵심용어: 건설 현장, 중대 재해, 재해 예방, 텍스트 마이닝

Received | 15 November, 2023

Revised | 21 March, 2024

Accepted | 26 March, 2024

OPEN ACCESS



This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0>) which permits unrestricted noncommercial use, distribution, and reproduction in anymedium, provided the original work is properly cited.

© Society of Disaster Information All rights reserved.

서론

2021년 산업재해로 사망한 건설근로자는 총 417명으로 전체 산업재해 사고사망자의 50.4%를 차지한다(고용노동부, 2021). 사망한 건설근로자의 산업재해 유형은 떨어짐 248명(59.5%), 부딪힘 37명(8.9%), 물체에 맞음 30명(7.2%), 깔림·뒤집힘 26명(6.2%), 무너짐 25명(6.0%) 순으로 많이 발생하였다.

Table 1. Status of deaths of construction workers by type of disaster

구분	계	떨어짐		부딪힘		물체에 맞음		깔림 뒤집힘		무너짐	
		인원(명)	비율(%)	인원(명)	비율(%)	인원(명)	비율(%)	인원(명)	비율(%)	인원(명)	비율(%)
2020년	458	236	51.5	38	8.3	42	9.2	33	7.2	24	5.2
2021년	417	248	59.5	37	8.9	30	7.2	26	6.2	25	6.0
증감	▽41	12	8.0	▽1	0.6	▽12	▽2	▽7	▽1	1	0.8

건설근로자의 사망 사고 유형은 대부분이 예방이 가능한 재래형 사고임에도 불구하고 전체 산업재해 사고 사망자수 중 가장 높은 비율을 차지하고 있다. 이러한 상황 속에서 건설근로자의 예방 가능한 재해의 발생을 최소화하기 위해 각 사고 유형별 예방 대책을 분석하고 제시하는 것은 의미 있는 일이라 사료된다. 본 연구에서는 국내 건설업 중대재해 사례를 분석한 데이터베이스를 바탕으로 텍스트 마이닝 기법을 활용하여 예방 대책의 키워드를 도출한 후 WordCloud를 활용하여 분석 내용을 시각화하고, 이를 건설근로자의 사고를 예방하는데 기여할 수 있는 자료로 활용하고자 한다.

연구의 범위 및 방법

본 연구를 위해 한국산업안전보건공단(KOSHA)에서 제공하는 1999년부터 2021년까지의 약 3909개의 국내 건설 중대 재해사례에 대한 텍스트 데이터를 수집하였다. 하지만, 한국산업안전보건공단(KOSHA)에서 제공하는 데이터는 대부분 비정형 텍스트 형태로 이루어져 있어 종합적인 분석에 어려움이 있다. 이에 본 연구에서는 텍스트 마이닝(Text Mining)을 통해 비정형 텍스트 데이터에서 유의미한 정보를 얻고자 하였다. 이를 위해 한국어 형태소 분석 패키지인 Konlpy를 활용하여 수집한 데이터에서 형태소를 추출하고, t.nouns 기능을 사용해 텍스트 데이터 내 명사들을 분류한 후 분류된 명사 데이터의 빈도수를 파악하였다. 또한, 도출된 데이터를 기반으로 WordCloud를 생성하여 사고 유형별 주요 예방대책 키워드를 시각화한 자료를 제시하였다.

선행 연구 고찰

텍스트 마이닝 기법의 발전으로 비정형 텍스트 데이터로부터 유의미한 정보를 추출하고 이를 분석하는 연구가 증가하고 있다. 텍스트 마이닝은 대량의 빅데이터에서 유용한 정보를 발굴하는 데이터 분석 기법으로, 이를 활용한 연구는 다수의 텍스트에서 핵심어나 연관 관계를 산출하는 방식이 주로 이룬다(Kil, 2019). 고은정의 연구에서는 원본 문서를 각 문장 단위로 분리하여 문장의 집합으로 구성된 데이터로 재구성한 후, 이를 토픽 모델링하여 미리 정해진 개수의 토픽을 추출함으로써 문서를 효과적으로 요약하였다(Ko et al., 2018). Sajid의 연구에서는 문장 단위로 텍스트를 분할하고 단어의 빈도, 명사 빈도, 대명사동사형용사명사 빈도를 나누어 각각 분석한 후, 명사를 대상으로 했을 때 가장 신뢰도가 높음을 보였다(Sajid et al., 2017).

2000년대 이후로 건설 현장에서 발생하는 각종 사고를 예방하여 중대재해 발생을 저감시키기 위한 연구에 텍스트 데이터를 분석하고자 하는 연구가 지속적으로 수행되고 있다. Tixier의 연구에서는 건설사고 보고서에서 유의미한 데이터들을 95% 이상의 정확도로 추출하는 방법론을 개발하였다(Tixier et al., 2016). 또한, 김준수의 연구에서는 계절별 건설현장 추락

사건의 특징을 분석하기 위한 텍스트 마이닝을 실시하여 계절별 추락사고의 직·간접적 요인을 분석하였다(Kim et al., 2019).

건설 중대재해 데이터 분류

본 연구에서는 Python 기반의 웹 스크래핑을 활용하여 한국산업안전보건공단(KOSHA)에서 제공되는 보고서를 사건 제목, 사건 개요, 사고 유형, 핵심예방대책, 상세 예방대책으로 내용을 나누어 Table 2와 같이 엑셀 파일로 데이터를 정리하였다. 또한, 사고유형별 대책에 대한 텍스트마이닝을 진행하기 위해 건설 중대재해 상세예방대책을 사고 유형에 따라 분류하였다. 본 연구에 사용된 데이터는 사고 유형을 발생 빈도수가 높은 8가지 유형(감전, 붕괴 및 도괴, 협착, 비래, 전도, 충돌, 질식, 추락)으로 분류되었다.

Table 2. Example of data cleanup

사건 제목	사고 유형	사건 개요	예방대책	
			핵심예방대책	상세예방대책
강관 파일과 파일 사이에 협착	협착	2005.3.9. 전북 김제시에서 강관 적치를 위한 멈춤 썸기를 설치하는 작업 중 페이로다로 내려놓은 파일이 굴러서 파일과 파일 사이에 협착되어 사망한 재해임.	경사면에서의 중량물 취급방법 준수, 작업계획서 작성	원형으로 된 강관파일 등 중량물 취급시 굴러가는 등 이동을 방지하고 취급재료의 하중에 충분히 버틸 수 있는 구름 멈춤대, 또는 썸기 설치, 중량물의 종류 및 형상, 취급방법 및 순서, 작업장소의 넓이 및 지형 등이 포함된 작업계획서 작성.

형태소 추출 및 빈도수 확인

분류된 텍스트 데이터를 기반으로 Fig. 1과 같이 Python을 사용하여 한국어 형태소 분석 패키지 Konlpy를 통해 텍스트 데이터에서 형태소를 추출하였다. 다음으로 Fig. 2와 같이 형태소 데이터의 두 글자 이상의 명사만을 정렬하여 빈도수를 확인한다. Fig. 3은 도출된 형태소의 빈도의 예시로서 추락 및 붕괴와 관련된 형태소의 빈도수이다.

```
In [9]: tokens = t.morphs(df_data)
print('# 토큰 수 : ', len(tokens))
print('# 토큰 예시 : ')
print(tokens[:1070])

'물', '매물', '감전', '추락', '추락', '추락', '추락', '낙하', '낙하', '
'멈', '깔림', '깔림', '추락', '전도', '전도', '추락', '추락', '깔림', '
'괴', '붕괴', '붕괴', '붕괴', '화재', '깔림', '추락', '추락', '협착', '
'깔림', '추락', '전도', '추락', '의사', '전도', '추락', '깔림', '추락', '
'하', '붕괴', '교통사고', '화재', '\n', '중독', '붕괴', '붕괴', '붕괴', '
'붕괴', '전도', '교통사고', '추락', '붕괴', '붕괴', '\n', '추락', '추
'괴', '붕괴', '붕괴', '폭발', '감전', '붕괴', '붕괴', '감전', '\n', '국
'통사고', '협착', '충돌', '\n', '추락', '화재', '맞음', '\n', '전복', '
'추락', '전도', '\n', '협착', '전도', '\n', '협착', '추락', '추락', '
'협착', '추락', '협착', '추락', '추락', '추락', '협착', '추락', '추락
'사고', '화재', '\n', '중독', '붕괴', '추락', '추락', '추락', '의사', '
'붕괴', '추락', '추락', '붕괴', '붕괴', '붕괴', '질식', '질식', '붕괴
'락', '붕괴', '협착', '붕괴', '붕괴', '붕괴', '붕괴', '낙하', '배', '
'전도', '추락', '추락', '전도', '추락', '붕괴', '폭발', '붕괴', '추락'
```

Fig. 1. Morpheme extraction

```
In [14]: tokens = [token for token in my_words if len(token) > 1]

In [15]: my_dict = {}
for word in tokens:
    my_dict[word] = my_dict.get(word, 0) + 1

sorted_dict = sorted(my_dict, key=my_dict.get, reverse=True)
for key in sorted_dict[:300]:
    print(f'{repr(key)} : {my_dict[key]}')
print()
```

Fig. 2. Check the quantity of morpheme



Fig. 3. Quantity of morpheme

건설 중대재해 유형별 Word Cloud 생성 및 분석

건설 중대재해 유형별로 도출된 형태소의 빈도를 기준으로 빈도수 상위 70개의 단어를 활용하여 Fig. 4와 같이 Word Cloud를 생성하였다. Word Cloud는 각 단어의 크기가 빈도 또는 중요성을 나타내는 텍스트 데이터 시각화 기술이다. 즉, Word Cloud는 문서의 키워드, 개념 등을 직관적으로 파악할 수 있도록 핵심 단어를 시각적으로 돋보이게 하는 기법이다. 주로 방대한 양의 정보를 다루는 빅데이터를 분석할 때 데이터의 특징을 도출하기 위해 사용한다. Word Cloud 분석의 결과는 출현 빈도수가 높은 단어는 크게 표시되고 각 단어들은 색상으로 구분 표시된다. 여기서 출현 빈도수가 높은 단어는 일반적으로 중요도가 높고 관심도가 높은 것으로 전제한다. 본 연구에서는 Word Cloud를 건설 중대재해 데이터베이스의 각 사고유형별 예방대책에 적용한 것으로서 사고유형별 주요 예방대책이 시각적으로 제시된 것이다.

추락 관련 Word Cloud

추락사고의 예방 대책과 관련한 주요 키워드는 안전, 설치, 발판, 난간, 부착 등으로 나타났다. 이를 통해 추락사고의 주요 예방대책은 안전난간, 발판의 설치 및 부착이라 할 수 있다. 안전난간이 미설치되거나 견고히 설치가 되지 못한 사례가 종종 발생하며, 사고로 직결됨을 알 수 있다.

붕괴 관련 Word Cloud

붕괴사고의 예방 대책과 관련한 주요 키워드는 굴착, 설치, 지반, 해체 등으로 나타났다. 붕괴사고의 경우, 굴착이나 해체 작업 중 주로 발생한다. 따라서, 지반에 대한 흠막이벽과 같은 구조물의 설치 필요성이 대두된다. 붕괴 사고의 특성상 대형사고로 이어질 수 있는 점을 염두하고 사고발생을 사전에 예방하는 것이 중요하다.

감전 관련 Word Cloud

감전사고와 관련한 주요 키워드는 절연, 충전, 설치 등으로 나타났다. 감전사고는 설비의 충전 과정에서 발생 가능성이 높음을 알 수 있고, 감전 사고는 갑작스러운 사고가 많이 발생하는 유형으로 절연기구 등의 사용을 통해 사고 발생의 가능성을 낮추고 설치 작업을 진행하는 것이 중요하다고 판단된다.



Fig. 4. Word cloud

협착 관련 Word Cloud

협착사고는 기계, 차량, 건설, 근로자, 유도 등의 키워드가 크게 표시되어 있다. 협착사고는 주로 건설 기계 혹은 차량으로 인해 주로 발생한다고 볼 수 있다. 이러한 사고 예방을 위해 건설기계 유도근로자 배치의 필요성이 요구되는 것으로 사료된다.

비래 관련 Word Cloud

비래사고의 예방대책 주요 키워드는 인양, 낙하, 사용중량 등으로 나타났다. 낙하사고의 대부분은 자재 인양 중에 발생하고 대부분의 자재가 중량물이기 때문에 사망사고로 직결되는 사고 유형들 중 하나이다. 비래 사고의 대부분은 인양하는 자재의 중량이 과도할 때 발생하기 때문에 비래의 사고 예방대책 중 가장 중요한 것은 사용 중량 준수이다.

전도 관련 Word Cloud

전도사고의 예방대책 키워드는 설치, 방지, 전도, 사용, 기계 등이 있다. 전도사고의 경우, 주로 크레인이나 차량 등과 같은 건설기계의 전도사고가 대부분이다. 전도사고는 설치나 해체 작업 등에서 주로 발생하며, 항시 방지대책과 안전수칙을 준수하는 것이 중요하다.

충돌 관련 Word Cloud

충돌사고의 사고예방대책 주요 키워드는 근로자, 기계, 차량, 유도, 건설이다. 위를 통해, 충돌사고는 주로 건설기계나 차량과의 충돌로 주로 발생함을 알 수 있다. word cloud에서 근로자가 가장 크게 표시되었듯, 근로자의 안전수칙 준수가 가장 중요하며 추가적으로 사고방지를 위해 건설 기계 유도근로자의 배치 또한 중요하다.

질식 관련 Word Cloud

질식사고는 산소, 농도, 가스, 환기, 사용 등의 주요 키워드가 사고 예방대책으로 뽑혔다. 질식사고는 대부분 작업장소의 산소 저하로 발생한다. 주로 가스에 의한 산소 농도 저하가 주로 발생하기에 환기를 통해 산소 농도를 맞추는 것이 가장 중요한 질식사고 예방대책 중 하나이다.

결론

본 연구에서는 국내 건설현장의 중대재해 사고의 패턴을 분석하기 위해 한국산업안전보건공단의 국내 건설업 재해사례 3909건을 수집하였다. 국내 건설업 중대재해 사례 텍스트 데이터를 활용하여 국내 건설업에서 실제 가장 많이 발생하는 재해 유형 빈도수를 도출하고, 가장 많이 언급된 8가지의 사고 유형에 대한 상세예방대책의 텍스트 데이터를 정리하여 재해 유형 간의 예방대책에서 많이 언급된 키워드를 도출하여 사고유형별 예방대책 키워드를 Word Cloud를 통해 시각화하여 제공하였다.

연구 결과 국내 건설현장에서 많이 발생하는 재해 유형은 추락, 감전, 붕괴(도괴), 협착(끼임, 깔림), 낙하(비래), 전도, 충돌(맞음, 강타), 질식 순이었다. 유형 별 예방대책 키워드를 Word Cloud로 도식화하였을 때, 추락사고에서는 빈도수가 큰 단어로 ‘설치’, ‘안전’, ‘발판’, ‘난간’, ‘비계’ 등이 분석되었다. 감전사고에서는 ‘절연’, ‘충전’, ‘설치’, ‘접지’ 등이 분석되었고 붕괴사고에서는 ‘설치’, ‘굴착’, ‘해체’, ‘흙막이’, ‘지반’ 등의 키워드 단어를 도출해 낼 수 있었다. 협착에서는 ‘기계’, ‘유도’, ‘근로자’, ‘차량’ 등의 단어를 분석하였고 낙하-비래에서는 ‘인양’, ‘낙하’, ‘중량’, ‘크레인’ 등의 단어들에 중요 키워드가 되었음을 알 수 있었다. 전도에서는 ‘방지’, ‘설치’, ‘조치’, ‘기계’ 등을 그리고 충돌은 ‘기계’, ‘차량’, ‘배치’, ‘근로자’ 등의 단어들 키워드로 분석할 수 있었다. 마지막으로 질식사고에서는 ‘산소’, ‘가스’, ‘농도’, ‘사용’, ‘환기’ 등이 단어들 분석하였

다. 이렇게 도출된 8가지 사고유형별 상세 예방대책 관련 핵심 키워드들을 통해 중대재해 사고의 예방대책을 마련에 있어 고려해야할 요소를 파악하는데 도움이 될 것이라 사료된다. 또한, 도출된 요소에 대한 다양한 관점에서의 해석을 통해 안전 예방 연구로의 확장을 도모할 수 있을 것이라 사료된다.

Acknowledgement

본 연구는 2023년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기본연구임(2021R1A2C2013841).

References

- [1] Kil, H.-H. (2019). "A study on the analysis method using text mining in the content structure of single text." *Journal of Korea Reading Association*, Vol, 52, pp. 91-134.
- [2] Kim, J.-S., Kim, B.-S. (2019). "Characteristics analysis of seasonal construction site fall accident using text mining." *Korean Journal of Construction Engineering and Management*, Vol. 20, No. 3, pp. 113-121.
- [3] Ko, E.-J., Kim, N.-G. (2018). "Automatic quality evaluation with completeness and succinctness for text summarization." *Journal of Intelligence and Information Systems*, Vol. 24, No. 2, pp. 125-148.
- [4] Sajid, A., Jan, S., Shah, I.A. (2017). "Automatic topic modeling for single document short texts." In 2017 *International Conference on Frontiers of Information Technology*, Pakistan, pp. 70-75.
- [5] Tixier, A.J.P., Hallowell, M.R., Rajagopalan, B., Bowman, D. (2016). "Application of machine learning to construction injury prediction." *Automation in Construction*, Vol, 69, pp. 102-114.