**Analysis of Fire Accident Factors on Construction Sites**

**Using Web Crawling and Word Embedding**

Jaehong Kim1, Sangpil Youm2, A 3, A4

1 PhD Student, School of Civil & Environmental Engineering, Oklahoma State University, 207 Engineering South, Stillwater, OK 74078, PH (405) 780-3866; Email: [jaehong.kim@okstate.edu](mailto:jaehong.kim@okstate.edu)

2

3

4

5

6

\*Corresponding author:

**ABSTRACT**

Recently,

**KEYWORDS:** Construction sites, Safety, Fire accidents, Data mining, Machine learning

**Introduction**

In recent years

**Background (Literature Review) – 일단 아래 3개 테마로 해서 각 10개씩 논문 찾아서 정리하면 될 듯**

***Construction Safety***

Construction site accidents

***Data collection - Web crawling***

The

***Word Embedding (Word2Vec)***

To

**Methodology – 일단 대략적인 가이드라인을 잡아 놨음. 혹시 더 추가해야 되는 방법론이나 분석방법이 있으면 자유롭게 추가 가능**

***Data collection procedures and sample***

Based on

***Web crawling***

* ***Selenium***
* ***Beautiful Soup***
* ***Data facts***

|  |  |
| --- | --- |
| * Issued duration | 2000.01.01 – 2020.04.22 |
| News Source | The New York Times |
| Number of Article | 861 |
| Number of words | 453,283 |

Table : news data information

* ***Data preprocessing***

The

***Word Embedding - Word2Vec***

***- Why Word2Vec***

***- Theoretical grounds in Word2Vec***

***-- neural network***

***-- skip-gram vs cbow***

***- Setting of Word2Vec***

***- cosine similarity***

The

***UMAP***

* ***Why UMAP***
* ***Dimensional Reduction***

The

**Results – 결론 부분에는 그래프, 표, 이미지 다수 포함. 소제목들은 일단 그래프 표 등 정리한 후에 결정**

1. ***Basic statistical facts***Figure : frequency of keywords
2. ***Word Embedding with word2vec***
   1. ***Retrieve similar words by cosine similarity***

***2.1.1 similarity between keywords***

Table : similarity between keywords

***2.1.2 Visualizing with UMAP***



Figure 2 : word2vec result overview with umap

Subgraphs for keywords

A close up of text on a white background

Description automatically generated

Figure 3 : subgraphs for keywords

Figure 3 shows graph confining the scope of keywords.

***2.1.3 Similar words of keywords***

***- network analysis (Threshold : cosine similarity 0.5)***



Figure 4 : network of keyword's similar words

The size of nodes varies depend on degree.

* ***Top 20 list of similar words of keywords***



Table 3 : top 20 similar words

**Discussion**

**Conclusion**

By investigating firefighters' perceptions of construction site fire accidents that require consideration of various characteristics and by exploring the structural relationships between the elements, it is possible to make useful data in developing more realistic and differentiated policy proposals.

**REFERENCES**