**Development of prediction model for construction site accident**

**through web crawling and data mining**

Jaehong Kim1, Sangpil Youm2, A 3, A4

1 PhD Student, School of Civil & Environmental Engineering, Oklahoma State University, 207 Engineering South, Stillwater, OK, [jaehong.kim@okstate.edu](mailto:jaehong.kim@okstate.edu)

2

3

4

5

6

\*Corresponding author:

**ABSTRACT**

Recently,

1. 웹크롤링을 통한 데이터 콜렉션 (20%)
2. 수집된 데이터로 각종 분석 (사고유형별 / Fire에 집중) (40%)
3. 사고 예측 모델 개발 (40%)

**KEYWORDS:** Construction sites, Safety, Accidents, Web crawling, Data mining

1. **Introduction**

In recent years

1. **Background (Literature Review)**

* **각 항목별로 참고문헌 조사 및 글 작성**

***Construction site accidents***

* **이전과 같이 각 항목별 논문 또는 관련책 10개씩 넣으면 될 듯**

***Web crawling***

* **이전과 같이 각 항목별 논문 또는 관련책 10개씩 넣으면 될 듯**

***Data mining(예측모델과 관련된 모든 이론들)***

* **이전과 같이 각 항목별 논문 또는 관련책 10개씩 넣으면 될 듯**

1. **Methodology**

***Data collection using the web crawling***

* **데이터 콜렉션과 관계된 부분 작성**
* **파이썬과 같이 실제 사용한 개발 언어 및 라이브러리 관련도 작성**
* **5132 article , 5090410 words, 45*6*5 article**

***Pattern analysis by accident type***

* **수집된 데이터를 어떻게 분석할 것인지 방법론 작성**

***Accident prediction model***

* **예측에 사용된 모든 방법론 정리**

**(Regression model, data mining, machine learning 등)**

* **논문의 핵심 부분이 될 수 있으므로, 상세히 정리하는게 좋을 듯**

1. **Results**

***Web crawling***

* **웹크롤링을 통해서 수집한 기사 관련 내용**
* **데이터 클리닝 방법 및 총 개수 등 상세히 설명 (표 첨부 필수)**

***Accident pattern***

* **지금 우리가 고민하고 있는 분석 내용을 넣어야 될듯**
* **그래프나 관련 표 등이 많이 나와야 할 부분**

1. **20년 (2000~2019) 각 연도별 사고관련 기사 총 개수 및 유형별 정리 필요**
2. **사고 유형별 노출 빈도 분석 (Fall, Collapse, Fire, Equipment(또는 Crane), Explosion 정도)**
3. **사고 유형별 – 요일/월/계절/날씨~~~~~~~~ 등 시계열 분석**
4. **Fire 사고에 집중해서 다른 사고들과 어떠한 차이를 보이는지 그래프 등으로 분석**
5. **Time-series for 20year**
6. **ㅁㅁ**
7. **ㅁㅁ 아이디어 추가 해도**

***Prediction model***

* **카톡으로 말한 부분 분석 순서 및 결과를 나열 해야함 (모델을 만들어서 80을 가지고서 트레이닝 시키고 그걸로 나머지 20의 데이터에 대해서 테스트)**
* **수식으로 모델 구현**
* **구현된 모델로 테스트 또는 Case study 형식으로 한번 해보면 좋을 듯**

1. **Discussion**
2. **Conclusion**

By

**References**