

## 사회재난 및 안전사고 데이터 분석을 위한 표준 구조 연구

## A Study of the Standard Structure for the Social Disaster and Safety Incidents Data

이창열<sup>1\*</sup> · 김태환<sup>2</sup>Chang Yeol Lee<sup>1\*</sup>, Taehwan Kim<sup>2</sup><sup>1</sup>Professor, Department of Computer Engineering, Donggeui University, Busan, Republic of Korea<sup>2</sup>Professor, Department of Security Service, Yongin University, Yongin, Republic of Korea

\*Corresponding author: Chang Yeol Lee, lcy@deu.ac.kr

## ABSTRACT

**Purpose:** In this paper, we propose a common dataset structure which includes the incidents investigation information and features data for machine learning. Most of the data is from the incidents reports of the governmental part and restricts on the social disaster and safety areas. **Method:** Firstly, we extract basic incidents data from the several incident investigation reports. The data includes the cause, damage, date, classification of the incidents and additionally considers the feature data for the machine learning. All data is represented by XML standard notation. **Result:** We defined the standard XML schema and the example for the incidents investigation information. **Conclusion:** We defined the common incidents dataset structure for the machine learning. It may play roles of the common infrastructure for the disaster and safety applications areas

**Keywords:** Social Disaster, Incident Investigation, Machine Learning, Risk Prediction, Incident Data

## 요약

연구목적: 본 논문은 사회재난 및 안전사고 발생에 따른 재난 유형별 조사 분석 정보에 대한 공통 데이터 도출과 머신 러닝 기반 사고 예측을 지원하는 특성화 데이터를 통합한 사회재난 및 안전사고 데이터셋 구조를 도출하는 연구에 초점을 맞추었다. 연구방법: 기존 조사 분석 보고서의 사고 분류, 원인, 피해 등을 표시할 수 있는 데이터를 중심으로 머신 러닝에 활용할 수 있는 특성화 데이터 도출과 이에 대한 XML 기반의 표준 체계를 도출한다. 연구결과: XML 기반의 표준 스키마 도출과 사례 제시를 하였다. 결론: 본 논문에서 도출된 표준안을 사회재난 및 안전사고 데이터셋 구축에 활용하고, 이를 기반으로 여러 분야에서 재난 사고 및 안전의 위험을 예측할 수 있는 응용 기술을 개발할 수 있게 지원한다.

핵심용어: 사회재난, 사고 분석, 머신 러닝, 위험 예측, 사고 데이터

## 서론

도시화, 산업화, 지구온난화 등으로 재난 발생 빈도가 급격하게 증가하고 있으며, 이에 따라 발생 재난에 대한 원인 분석과 해결을 위하여 범 부처간 많은 노력이 진행되고 있다. 이러한 노력의 일환으로, 2011년 현재의 행정안전부에서 재난 사고 예측을 위한 재난 전조 정보

Received | 25 October, 2021

Revised | 16 November, 2021

Accepted | 19 November, 2021

 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 any medium, provided the original work is properly cited.

© Society of Disaster Information All rights reserved.

기반의 위험 예측 시스템(Lee et al., 2001)을 구축하여 활용하고 있다. 본 시스템에서는 하인리히 법칙에 따라 재난 발생 전조가 되는 징후를 도출하기 위해 상황 정보를 온톨로지 기반으로 구축하였으며, 매일 새울전자민원창구(전국 230개 민원 사이트), 지킴이(구 소방방재청 제보 사이트), 포털뉴스 사이트(네이버 등록 뉴스 267, 지역신문사 307개)를 대상으로 재난 징후가 되는 사건을 추출하였다. Fig. 1은 추출된 사건 정보를 보여주는 화면에 해당된다.



Fig. 1. Disaster symptom management system

이렇게 자동으로 추출된 재난 전조 정보를 기반으로 점검과 사전 조치를 통하여 발생될 재난에 대한 사전 조치 업무를 수행하였다.

자연재난의 대부분이 재난관리주관기관이 행정안전부로서 해당 정보 접근이 쉽지만, 사회재난의 경우, 재난 유형에 따라 부처별로 재난관리주관기관이 상이하기 때문에 실제적인 사고 정보를 관리하기에는 한계가 있는 것이 현실이다. 따라서 사회재난에 대한 통합적 관리와 이를 기반으로 종합적인 대책을 마련하는 것이 절실한 실정이다. 이러한 문제를 해결하기 위하여 행정안전부는 재난사고관리운영지원시스템을 1차, 2차에 걸쳐서(2016년 - 2019년) 구축하였다(MiraeIt Ltd., 2016). Fig. 2는 행정안전부가 구축한 재난사고관리 운영지원시스템의 초기 화면을 보여주고 있다. 이 시스템에서는 중앙 행정기관과 행정망 이용기관 10 곳, 그리고 국가재난원인조사기관협의회에 참여하는 14개 부처와 소속 산하 23개 기관이 참여하였으며, 초기에 통합적 데이터를 구축하였지만, 재난 사고 정보 통합에 대한 법적 지원이 부족하여 시범적 운영에서 종료되었다.



Fig. 2. Disaster accident management operation support system

이로인해, 사회재난에 대한 통합적 접근에 대한 법적 지원의 필요성이 대두되었다. 관련하여, “재난 현장 대응 및 상황 정보 관리 모델에 관한 연구”(Lee, 2018), 사회재난 위험지수 개발연구”(MOIS., 2018), “사회재난 위험성 분석 모델 개발 및 지자체 시범 적용”(MOIS., 2020). 연구들이 수행되었다. 연구 결과물로 사회재난 관련 법안 초안을 마련한 상태이며, 이를 기반으로 행정안전부가 입법화를 추진하고 있는 상태이다.

다양한 사회재난을 통합적으로 관리하기 위하여 이들 사고 정보에 대한 표준화된 데이터 체계가 필요하며 본 논문에서 그 표준안을 제시한다. 제시되는 표준은 추후 사회재난 통합 플랫폼 등에 활용될 수 있으며, 나아가 머신러닝 등에서 활용할 수 있는 특성 정보를 표준에 포함시켜서 제안하고 있다.

## 기존 연구 분석

### 사회재난 및 안전 연구 분석

#### 행정안전부 사회재난 유형 분류

행정안전부가 고시한 사회재난은 총 43종이며, 이를 관리하기 위하여 지역안전지수, 지역안전도 평가, 재난 관리 평가를 수행하고 있으며 이러한 평가는 주요 재난 7종을 대상으로 하고 있다. 그러나 모든 사회 재난을 분석하기에는 광범위하며 분석의 한계가 상존하기 때문에, 대부분의 연구에서는 직접 관련된 사회재난 유형을 한정하고 수행하고 있다. 예를 들어 시·도가 관리하는 중점 사회 재난 유형에 관하여 수행한 연구(Lee et al., 2021)에서는 사회재난 유형을 다음과 같이 설정하여 연구를 수행하였다.

- 감염병 재난, 가축질병, 수질분야 환경오염분야, 공연장사고, 다중밀집시설재난, 해양선박사고, 해양분야 환경오염사고, 사업장 인적사고, 지하철사고, 유해화학물질유출사고, 산불

### 국립재난안전연구원의 재난 분류

국립재난안전연구원(NDMI; National Disaster Management Research Institute)에서 수행한 “비정형 텍스트 데이터 기반 weak signal 탐색 방법론 개발 및 탐색 사례연구”의 결과(MOIS., 2020)로 재난 이슈에 대한 정보를 쉽게 획득할 수 있는 “실시간 재난안전뉴스 모니터링 시스템(<http://rscanner.ndmi.go.kr/scanning/>)”을 개발하여 제공하고 있다. 본 시스템에서 재난 분류는 자연재난 사회재난 구분없이 발생하는 재난 및 안전 유형으로 다음과 같이 57개의 재난 유형을 분류하여 사용하고 있다.

태풍, 미세먼지, 홍수, 수질오염, 강풍, 해양오염, 풍랑, 에너지마비, 해일, 통신마비, 조수, 교통마비, 산사태, 금융마비, 가뭄, 의하는 중점 사회 재난 유형에 관하여 수행한 연구(Friend, 2013)에서는 그 유형을 통합하여 다음과 같이 선정하여 수행하고 있으며, 이는 기존 사회재난 분류를 일부 통합하여 새로 정의한 체계이다. 료마비, 지진, 식용수마비, 황사, 공공청사마비, 녹조, 쓰레기매립지마비, 적조, 감염병, 대설, 가축전염병, 한파, 수산생물전염병, 폭염, 승강기사고, 낙뢰, 전기가스사고, 화산폭발, 등산레저사고, 자연우주물체, 물놀이사고, 화재, 생활제품사고, 산불, 사업장산재, 붕괴, 농어업사고, 자동차사고, 식품사고, 자전거사고, 의료제품사고, 지하철사고, 범죄, 고속철도사고, 자살, 항공교통사고, 테러, 선박사고, 취약계층, 방사능사고, 해충, 유해화학물질사고

### 중대재해처벌법의 재난 유형

2021년 1월에 제정되고 2022년 시행을 앞둔 중대재해처벌법은 다음과 같이 중대산업재해와 중대시민재해로 구분된다. 각 재해에 대한 분류는 다음과 같다.

- 중대산업 재해: 산업안전보건법에 따른 산업재해(건설물, 설비, 원재료, 가스, 증기, 분진 등 작업과 관련된 재해)
- 중대시민재해: 원료 또는 제조물, 공중이용시설, 공중교통수단의 설계, 제조, 설치, 관리상 결함으로 발생한 재난

### 기관별 재난조사보고서 현황

- 산림청: 산불피해대장과 산불원인조사보고서
- 농림축산검역본부: 전염병발생보고서
- 한국산업안전보건공단: 재난분야별 보고서, 재해사례집, 중대사고보고서
- 중앙해양안전심판원: 재결서/재결요약서, 사고사례
- 질병관리본부: ○○역학조사서, ○○질병 사례조사서
- 한국가스안전공사: 사고정보, 사고조사서
- 국토안전관리원: 사고요약
- 한국원자력안전기술원: ○○사고조사보고서
- 한국전기안전공사: 사고조사 지침서, 사고통계/모니터링
- 항공철도조사위원회: 철도사고 사례집, ○○철도사고조사보고서, ○○항공사고조사보고서
- 화학물질안전원: 중대사고 사례집, 사고 사례 정보
- 도로교통공단: 교통사고분석시스템, ○교통사고 분석
- 국립과학수사연구원: 각종사고 분석 보고서

## 사회재난 조사 예측 시스템 분석

### 국도교통부 사고조사 시스템

국도교통부는 건설사고신고 매뉴얼에 따라 건설안전정보시스템(COSMIS)에 입력하여 자체사고조사 및 정밀 현장 조사 결과를 관리하고 있다. 제공되는 정보는 공사종류별, 발생공종별, 사고원인별, 사고유형별로 분류 제공하고 있다.



Fig. 3. Construction safety management information system

### 국립재난안전연구원의 실시간 재난안전뉴스 모니터링 시스템(<http://rscanner.ndmi.go.kr/scanning/>)

국립재난안전연구원의 시스템(NDMI., 2014)은 R-Scanner에 기반하여 재난 예측을 구현하였다. Fig. 4는 실시간 재난안전뉴스 모니터링 시스템 화면으로 뉴스 정보를 분석하여 예상되는 사고 관련 정보를 보여주고 있다.

### IBM의 Blue CRUSH

영화 마이너리티 리포트처럼 범죄 예측을 위한 기술(Predictive Policing)로 미국 테네시주 멤피스는 IBM의 Blue CRUSH(Criminal Reduction Utilizing Statistical History)를 도입하고 운영하고 있다(Friend, 2013). Blue CRUSH는 빅 데이터 분석을 통하여 범죄 발생을 예측하여 해당 지점에 경찰을 미리 배치하여 범죄를 예방하는 기술로 범죄 발생이 줄어드는 효과를 가져왔다.

### 복합재난확산예측시스템

복합재난확산예측 시스템(Park et al., 2020; Jung et al., 2019)은 머신 러닝 기술 보다는 전통적 인덱스(지표) 기반으로 재난 확산을 예측하는 연구를 수행하였다.



Fig. 4. Real-time disaster news monitoring system of NDMI

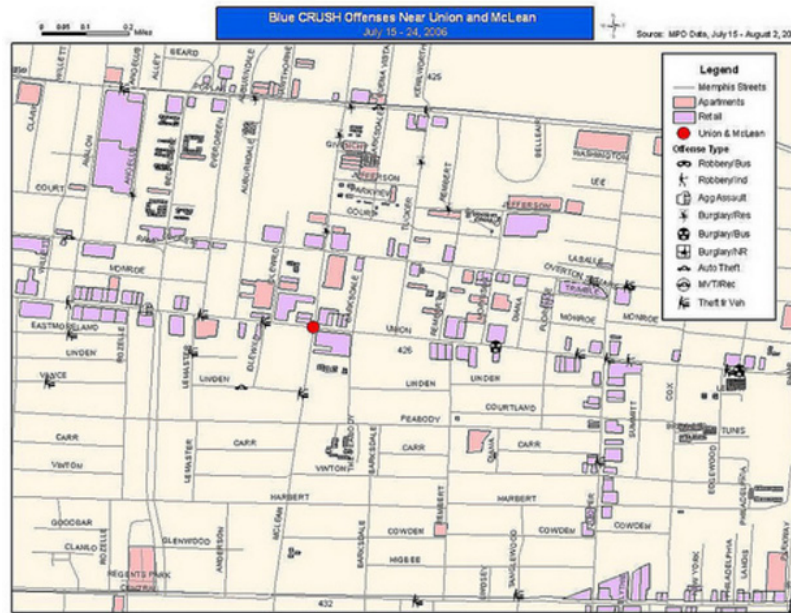


Fig. 5. Blue CRUSH of IBM

## 사회재난 및 안전사고 데이터 설계

### 데이터 구조 설계

기존의 각 부처별 사고조사 보고서 및 국립재난안전연구원의 실시간 재난뉴스 모니터링시스템 등을 참고하여 Fig. 6과 같은 절차에 따라 사고 데이터를 설계하였다. 본 데이터 설계는 기존의 부처별 사고 조사보고서 등을 참고하였으면, 데이터 구조 정보는

URN(Universal Resource Names) 체계에 기반한 ID, XML 기반 및 기 연구 분석된 재난 분류 체계를 기반으로 데이터 표준 체계를 설계하였다. 설계된 표준안은 데이터 활용 공유 기관의 검토를 거쳐서 표준안으로 확정하며, 확정된 표준안을 기반으로 사회재난 통합 플랫폼, 빅데이터 플랫폼 등에서 활용할 수 있을 것이다. 설계되는 데이터는 다음과 같은 세부사항을 포함한다.

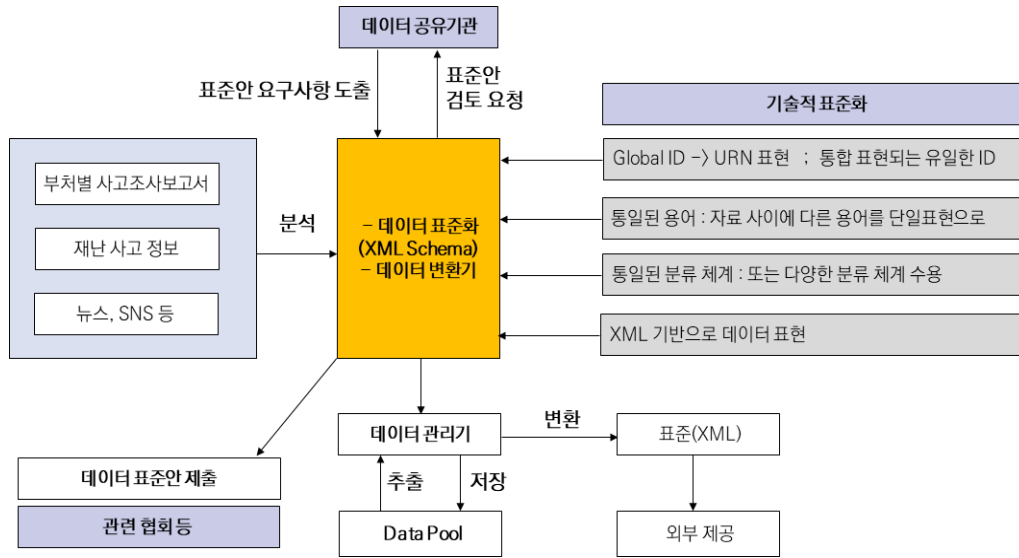


Fig. 6. Data standardization processing

### 데이터 정보

- 기관 코드: 재난정보공동이용센터에서 사용하는 기관코드
- 재난 분류: 행정안전부 현장조치행동매뉴얼의 재난 분류, NDMS(National Disaster Management System; 국가재난관리시스템)의 재난사고보고서를 제공하는 기관의 기관별/유형별 재난 상세 분류, NDMI 재난 분류
- 재난 원인 분류: 각 기관별로 정의한 재난 원인분류 사용
- 사고 일반 정보: 6하 원칙과 재난 수습 과정 정보 포함
- 환경 정보: 기상 등 정보
- 검색 분석 정보: 키워드 등 정보
- 머신 러닝을 위한 특성 정보: 사고 원인, 재난 키워드, 피해 정보, 피해 등급 정보 등

### 데이터 운영

표준의 상위 구조는 사고 데이터와 관련된 기관 정보와 사고 자체에 대한 정보로 다음과 같이 2가지 형태로 구분하였다.

- 전송 관련 정보: Header
- 사고 관련 정보: 여러 사고를 묶음으로 표현할 수 있기 때문에 Events로 정의

### 데이터셋 서비스

사고 데이터 표준으로 다음 Table 1.과 같이 개념적 구조를 제시하였으며 일부만 설명하면 다음과 같다.

- 레벨 1: Header(데이터 전송 관련 정보), Events(1개 이상의 사고 정보)
- 레벨 2: Events 하위 정보만 설명함. 1개의 사고 정보
- 레벨 3: Type(사고 유형), ID(사고 번호), Title(사고 제목), SubTitle(부제목), Classification(분류), Contents(사고 내용), Keywords(관련 키워드), SenderData(전송기관 정보), Place(사고 발생 장소), Weather(기상정보), Institute(재난관리기관), Cause(사고 원인), PesonalDamage(인명피해), FacilityDamage(시설피해), FinancialDamage(경제적피해), Actions(조사 대응 조치), Regulations(관련 법령), Materials(관련 자료)

### 도출된 데이터 항목

**Table 1.** Common data elements for the incidents data

레벨1	레벨2	레벨3	레벨4	레벨5	반복	비고
Header					0.1	헤더
	DataID				1.1	전송ID
	SendingType				1.1	전송 유형
	SendingDate				1.1	전송일
	Sender				1.1	전송자
		Institute			1.1	기관코드
		System			1.1	전송시스템
		Contacts			1.1	담당자
Events					1.1	이벤트
	Event				1.N	이벤트-반복구조
		Type			1.1	보고서, 상황정보, 사고정보
		ID			0.1	통합사고번호
		Title			1.1	제목
		SubTitle			0.1	부제목
		Classification			1	분류
			NDMS		1.1	NDMS분류 코드
			Institute		1.1	해당기관분류 코드
		Contents			1.1	사고내용
		Keywords			1.1	사고 키워드 리스트
		SenderData			0.1	해당기관 데이터
			IncidentCode		1.1	사고 코드
			ReceivedNo		0.1	접수번호
			ReceivedTime		0.1	접수시간
			EventTime		1.1	사고 시간
		Place			1.1	장소
			GPS		0.1	위성



**Table 1.** Common data elements for the incidents data (Continue)

레벨1	레벨2	레벨3	레벨4	레벨5	반복	비고
				Latitude	1.1	위도
				Longitude	1.1	경도
			Address		0.1	행정주소
			Name		0.1	명칭/지역/설명
		Weather			0.1	기상
			Warning		0.1	XX 주의보 상태
			Temperature		0.1	온도
			Humidity		0.1	습도
			WindSpeed		0.1	풍속
			Status		0.1	맑음/흐림/눈/비
		Institute			1.1	관련기관
			Responsibility		1.1	재난관리책임기관
			Primary		0.1	담당기관
			Investigation		0.1	원인조사기관
		Cause			0.1	사고 원인
			CauseCode		1.2	원인 코드-반복
			Description		0.1	원인설명
		PersonalDamage			0.1	인명피해
			Casualty		0.1	사망
			Lost		0.1	실종
			Injured		0.1	부상
			Other		0.1	기타
		FacilityDamage			0.1	시설피해
		FinancialDamage			0.1	재정피해
		Actions			1.1	대응
			Contents		1.1	조치 내용
			Problems		0.1	문제점
			History		1.N	조치 이력
			Date		1.1	날짜
			Status		1.1	재난-안전 상태
			Response		1.1	조치
		Regulation			0.1	관련법령
		Materials			0.1	관련자료
			Reports		1.1	메인자료
			Others		0.1	기타자료

본 연구에서는 위와 같이 제시한 표준을 SDML(Safety Data Markup Language)로 가칭하였다. SDML에 대한 세부 코드는 분량이 많은 관계로 아래 그림과 같이 일부만 제시하였다.

```

<?xml version="1.0" encoding="UTF-8"?>
<xs:schema xmlns:xs="http://www.w3.org/2001/XMLSchema" elementFormDefault="qualified" attributeFormDefault="unqualified">
  <xs:element name="SDML">
    <xs:complexType>
      <xs:sequence>
        <xs:element name="Header" minOccurs="0">
          <xs:complexType>
            <xs:sequence>
              <xs:element name="DataID"/>
              <xs:element name="SendingType"/>
              <xs:element name="SendingDate"/>
              <xs:element name="Sender">
                <xs:complexType>
                  <xs:sequence>
                    <xs:element name="Institute"/>
                    <xs:element name="System"/>
                    <xs:element name="Contacts"/>
                  </xs:sequence>
                </xs:complexType>
              </xs:element>
            </xs:sequence>
          </xs:complexType>
        </xs:element>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
</xs:schema>

```

Fig. 7. XML schema representation of SDML

### 데이터 구조 설계

앞에서 정의한 SDML에 따라 본 데이터를 관리하는 지능형 재난안전정보 통합플랫폼을 설계하였다. Fig. 8은 지능형 재난안전정보 통합플랫폼의 일부 모듈(사고 데이터 입력기)를 보여주고 있다. Fig. 8의 내용에 사고 정보가 기록되어 있으며 해당 정보 일부를 살펴보면 다음과 같다.

- 제목: 암모니아 누출 피해 사고
- 행정안전부 분류: 사업장 대규모 인적사고
- NDMI 분류: 폭발
- 유형: 보고서
- 사고시간: 2021-10-10T10:20:30
- 발생위치: GPS(34.23, 125.24), 행정주소(부산시 강서구 대저1동 123-10, 명칭(부산스테인레스))
- 담당지자체: 부산시
- 담당기관: 한국가스안전공사
- 중대재해분류: 중대산업재해
- 사고 원인: 타공사
- 키워드: 가스, 폭발
- 기상 정보: 맑음
- 피해: 사망(0), 부상(5), 실종(0), 동산(1억원), 시설(건물균열)
- 피해등급(5)



Fig. 8. Data input module of intelligent disaster & safety data platform

위와 같이 구축된 데이터는 추후 사고 예측을 위하여 활용할 수 있으며, Fig. 9는 이러한 사고 예측에 대한 설계정보를 제공하고 있다.

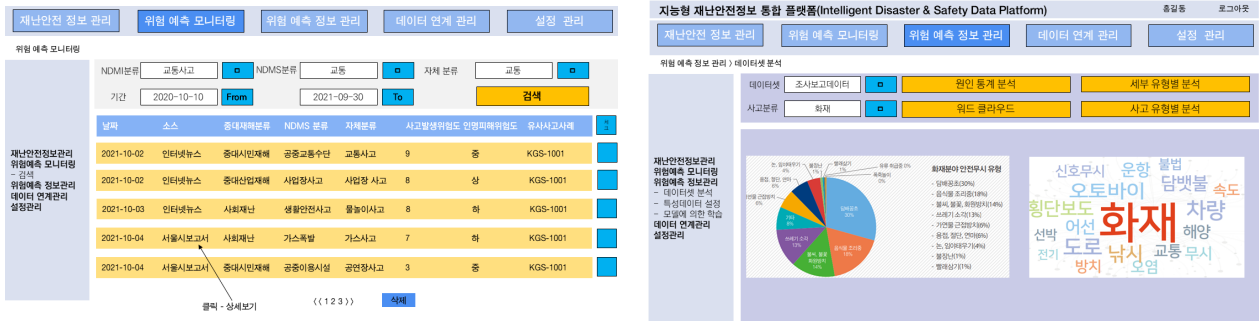


Fig. 9. Disaster prediction module of intelligent disaster & safety data platform

## 결론

부처별, 재난유형별 사고 조사 보고서 양식이나 시스템이 존재하지만, 다양한 재난 유형을 통합할 수 있는 체계는 제시되지 않았다. 이러한 관점에서 본 연구는 사회재난 사고 정보에 대한 통합적 데이터 표준 체계를 제시하고 있다. 이러한 표준 체계는 추후 재난정보를 통합적으로 관리할 수 있는 플랫폼의 기반이 될 것이다.

제시한 표준 정보는 재난 및 안전사고의 원인 정보, 피해 정보, 조사 정보를 포함하고 있으며, 또한 머신러닝 개발을 위한 특성 정보 포함하여 제시하고 있다.

본 표준에 의해 구축되는 사회재난 및 안전사고 데이터는 빅데이터 표준인 DCAT(Data Catalog Vocabulary)으로 쉽게 변경될 수 있으며, C-KAN과 같은 공개 플랫폼에서 운영될 수 있을 것으로 예측된다.

## Acknowledgement

본 논문은 2021년 행정안전부 연구과제(한국산업기술평가원 20015461)의 지원을 받아 수행되었음.

## References

- [1] Friend, Z. (2013). Predictive Policing : Using Technology to Reduce Crime, FBI Law Enforcement Bulletin, April 9, 2013
- [2] Jung, W.-S., Oh, S.-H., Lee, Y.-T. (2019). "A study on the system implementation of mega complex disaster damage evaluation and estimation using disaster bigdata." The Journal of Korean Institute of Communications and Information Sciences, Vol. 44, No. 7, pp. 1433-1438.
- [3] Lee, C.Y. (2018). "A development of the integrated model for the disaster field response and situation information management." Journal of Korea society of disaster information, Vol. 14, No. 1.
- [4] Lee, C.Y., Kim, T.H. (2011). "A study of the disaster sign data analysis technologies based on ontology." Journal of the Korean Society of Disaster Information, Vol. 7, No. 3, pp. 220-228.
- [5] Lee, H.J., Yun, H.S., Han, H. (2021). "A Study on the selection of types of social disaster by region." KOSDI, Vol. 17, No. 2, pp. 206-217.
- [6] MiraeIt Ltd. (2016). A development of the disaster operation management system, Ministry of the Interior and Safety.
- [7] MOIS (2018) A development of the social disaster risk index.
- [8] MOIS (2020) A development of the risk analysis model in social disaster and applying to local government. MOIS, Korea.
- [9] NDMI. (2014) A development of the risk prediction technologies from the real-time disaster risk analysis.
- [10] Park, S.Y., Sanggi, H.H., Lee, K.B. (2020). "Trends in disaster prediction technology development and service delivery, electronic and telecommunications trends." Vol. 35, No. 1 pp. 80-88.