



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공학석사학위논문

텍스트 마이닝 분석을 통한 산업안전
분야의 탐색적 분석

An Exploratory Analysis of Industrial Safety
Field Using Text Mining Analysis

울산대학교 대학원
안전보건전문학과
박선규

텍스트 마이닝 분석을 통한 산업안전
분야의 탐색적 분석

지도교수 장길상

이 논문을 공학석사학위 논문으로 제출함

2021년 12월

울산대학교 대학원
안전보건전문학과
박선규

박선규의 공학석사학위 논문을 인준함

심사위원 박 창 권



심사위원 정 기 효



심사위원 장 길 상



울 산 대 학 교 대 학 원
2021년 12월

국문 요약

텍스트 마이닝 분석을 통한 산업안전 분야의 탐색적 분석

울산대학교 대학원

안전보건전문학과

박 선 규

산업재해와 안전사고가 사회적인 문제로 대두되고 있으며, 국민들의 안전에 대한 관심이 높아지고 있다. 이에 문재인 정부에서 국민생명 지키기 3대 프로젝트를 추진하고 있으며, 2022년까지 산업재해 사망사고만인을 50% 감축을 목표로 산업안전보건법을 전부 개정하고, 중대재해처벌법을 제정하게 되었다.

최근 6년(2016년부터 2021년 6월)간 산업안전 관련한 언론보도 기사 3,647건과 소셜 네트워크 서비스(인스타그램, 트위터, 블로그)의 콘텐츠 83,512건을 분석하였다. 언론보도 기사의 분석은 2016년 443건, 2017년 579건, 2018년 649건, 2019년 647건, 2020년 847건, 2021년 6월까지 482건을 하였으며, 소셜 네트워크 서비스(커뮤니티, 인스타그램, 블로그, 뉴스, 트위터)는 2016년 11,993건, 2017년 12,052건, 2018년 12,927건, 2019년 14,185건, 2020년 17,184건, 2021년 15,171건으로 언론보도 횟수와 소셜 서비스의 언급량이 증가하는 추세이며, 이는 사회적 관심이 높아지고 있음을 나타낸다.

최근 6년의 주요키워드를 분석해보면 매년 특징적인 키워드를 확인할 수 있다. 2016년 서울 지하철 구의역 승강기 사고 관련 키워드, 2017년 삼성전자 반도체 공장 노동자의 산재승인 관련 키워드, 2018년 국민생명 지키기 프로젝트 관련 키워드, 2019년 태안 화력 발전소 사고 관련 키워드, 2020년 코로나와 산업안전보건법 전부 개정과 중대재해처벌법 제정 관련 키워드, 2021년 광주 건물철거 붕괴사고와 중대재해처벌법 공포 관련 키워드를 확인하였다. 사고 발생 이슈는 약 1개월간 언급량이 증가하였으며, 정책이슈는 약 3개월간 언급량이 증가하였다. 2020년 산업안전보건법 전부개정시 언급량이 증가하였고, 벌칙 규정이 강화된 2021년 중대재해처벌법은 제정 전부터 관련 키워드의 언급량이 증가하다 감소하였으며, 공포 시점에 언급량이 급증하였다. 또한 중대재해처벌법의 감정어 분석결과를 살펴보면 법안 제정에 따라 법안 적용시 예상되는 문제점, 산업계에 미치는 영향, 산업재해 예방의 효과성에 대한 우려의 부정적인 반응과 관련한 키워드가 다수 확인되어 추후 법령개정시 실효성 있는 정책 개선과 산업재해 예방을 위해 다양한 계층의 목소리를 반영하여 법안을 개정해야 할 것이다.

산업안전의 사회적 관심도가 증가함에 따라 언론보도 기사와 소셜 네트워크 서비스의 비정형 데이터의 텍스트 마이닝 분석기법을 활용하여 연도별 키워드 분석으로 국민들의 관심사를 파악하고 활용방안을 제시했다는 점에서 의미가 있다. 산업안전보건법 개정과 중대 재해처벌법 제정으로 국민들과 산업계에서 관심사와 우려사항을 분석하여 산업재해 감소를 위한 산업안전 정책을 개선·보완하는 방향으로 발전되어야 할 것이다.

목 차

1. 서론.....	1
1.1 연구의 배경 및 필요성	2
1.2 연구의 목적	2
1.3 연구의 범위 및 내용	2
2. 문헌 연구.....	3
2.1 빅데이터.....	3
2.2 텍스트 마이닝.....	6
2.3 감정분석.....	7
3. 연구방법.....	9
3.1 연구방법 5단계.....	9
3.2 연구방법 적용사례.....	10
3.2.1 데이터 수집 단계.....	11
3.2.2 데이터 전처리 단계.....	11
3.2.3 데이터 후처리 단계.....	13
3.2.4 데이터 분석 단계.....	13
3.2.5 데이터 시각화 단계.....	14
3.2.6 분석 결과.....	15
3.3 연구 분석환경.....	18
4. 산업안전 분야의 텍스트 마이닝 분석.....	19
4.1 연구대상.....	19
4.2 산업안전 분야 언론보도 분석.....	20
4.3 산업안전 분야 소셜 네트워크 서비스 분석.....	24
4.4 산업안전 분야 텍스트 마이닝 시계열 분석.....	26
4.4.1 2016년 언론보도 분석.....	26
4.4.2 2017년 언론보도 분석.....	29
4.4.3 2018년 언론보도 분석.....	32

4.4.4 2019년 언론보도 분석.....	35
4.4.5 2020년 언론보도 분석.....	38
4.4.6 2021년 언론보도 분석.....	41
4.4.7 연도별 주요 키워드 분석.....	45
4.4.8 통찰.....	48
4.5 중대재해처벌법 분석.....	51
4.5.1 언론보도 분석.....	51
4.5.2 소셜 네트워크 서비스 분석.....	53
4.5.3 통찰.....	56
5. 연구결과 및 맺음말.....	59
5.1 연구결과 및 의의.....	59
5.2 연구의 한계 및 향후 연구방향.....	61
6. 참고문헌.....	62

표 목 차

<표 3-1> 산업안전보건법 다빈도 키워드 100	16
<표 3-2> 중대재해처벌법 다빈도 키워드 100.....	17
<표 4-1> 산업안전 분야 텍스트 마이닝 분석조건	19
<표 4-2> 산업안전 언론보도 횟수.....	20
<표 4-3> 2016년 ~ 2021년 6월 언론사별 기사수와 비율.....	21
<표 4-4> 2016년 ~ 2021년 6월 소셜 네트워크 서비스 언급량 추이.....	24
<표 4-5> 소셜 네트워크 서비스 감정키워드의 빈도와 비율.....	25
<표 4-6> 2016년 언론보도 주요키워드.....	26
<표 4-7> 2016년 소셜 네트워크 서비스 주요키워드.....	27
<표 4-8> 2017년 언론보도 주요키워드.....	29
<표 4-9> 2017년 소셜 네트워크 서비스 주요키워드.....	30
<표 4-10> 2018년 언론보도 주요키워드.....	32
<표 4-11> 2018년 소셜 네트워크 서비스 주요키워드.....	34
<표 4-12> 2019년 언론보도 주요키워드	36
<표 4-13> 2019년 소셜 네트워크 서비스 주요키워드.....	37
<표 4-14> 2020년 언론보도 주요키워드.....	39
<표 4-15> 2020년 소셜 네트워크 서비스 주요키워드.....	40
<표 4-16> 2021년 언론보도 주요키워드.....	42
<표 4-17> 2021년 소셜 네트워크 서비스 주요키워드.....	43
<표 4-18> 언론보도 연도별 주요키워드 10.....	45
<표 4-19> 소셜 네트워크 서비스 연도별 주요키워드 10.....	46
<표 4-20> 언론보도와 소셜 네트워크 서비스의 연도별 특징 키워드.....	47
<표 4-21> 소셜 네트워크 서비스의 반대의미 키워드 연도별 빈도와 비율..	49
<표 4-22> 산재신청 및 인정현황, 근로복지공단.....	49
<표 4-23> 산업안전보건법, 중대재해처벌법의 언론보도 횟수.....	51
<표 4-24> 중대재해처벌법 언론보도 주요키워드 연도별 빈도.....	52
<표 4-25> 중대재해처벌법 감정어 연도별 빈도.....	55
<표 4-26> 공공기관 안전관리자 채용공고현황.....	58

그림 목 차

〈그림 2-1〉 빅데이터 특징, 네이버 지식백과	4
〈그림 3-1〉 연구방법 5단계	9
〈그림 3-2〉 연구방법 적용사례 세부과정	10
〈그림 3-3〉 국가법령정보센터의 산업안전보건법	11
〈그림 3-4〉 산업안전보건법 엑셀파일 확인과정	11
〈그림 3-5〉 법령파일 파이썬에서 불러오기	12
〈그림 3-6〉 파이썬에서 법령파일(엑셀파일) 불러오기 확인	12
〈그림 3-7〉 데이터 후처리 소스코드 일부	13
〈그림 3-8〉 다빈도 키워드 계산과정 소스코드	14
〈그림 3-9〉 데이터 시각화 과정 소스코드	14
〈그림 4-1〉 산업안전 언론보도 횟수의 연도별 추이	21
〈그림 4-2〉 2016년 ~ 2021년 언론사별 보도횟수	23
〈그림 4-3〉 소셜네트워크 서비스 언급량 추이	24
〈그림 4-4〉 소셜 네트워크 서비스의 감정어 비율	25
〈그림 4-5〉 2016년 주요키워드 시각화	29
〈그림 4-6〉 2017년 주요키워드 시각화	32
〈그림 4-7〉 2018년 주요키워드 시각화	35
〈그림 4-8〉 2019년 주요키워드 시각화	38
〈그림 4-9〉 2020년 주요키워드 시각화	41
〈그림 4-10〉 2021년 주요키워드 시각화	44
〈그림 4-11〉 산재보험 사업장 및 징수현황, 근로복지공단	49
〈그림 4-12〉 산업안전보건법, 중대재해처벌법 언론보도 횟수	51
〈그림 4-13〉 산업안전보건법, 중대재해처벌법 소셜 네트워크 언급량	53
〈그림 4-14〉 중대재해처벌법 감정어 비율	54
〈그림 4-15〉 중대재해처벌법 주요키워드 시각화	56
〈그림 4-16〉 안전관리자 검색결과 추이	57

1. 서론

1.1 연구의 배경 및 필요성

산업재해와 각종 안전사고가 빈번하게 발생하여 사회적 문제로 대두되고 있으며, 국민들의 안전에 대한 관심도가 높아지고 있다. 우리나라 노동자 만 명당 사고사망자 수는 OECD 국가 평균의 2~3배 수준으로 높다. 정부에서 2022년까지 자살예방, 교통안전, 산업안전 '3대 분야 사망 절반 줄이기'를 목표로 '국민생명 지키기 3대 프로젝트'를 추진하고 있으며, 근로자의 생명과 안전을 최우선으로 하는 산업안전보건 정책의 패러다임 전환을 위해 제50회 산업안전보건의 날(2017.7.3.) 공표하였다. 2022년까지 산업재해 사망사고 감축목표를 OECD 국가 평균수준인 사고사망만인율 50% 감축하여 0.27‰ 달성을 목표로 설정하였다. 이에 산업안전보건법과 시행령, 시행규칙을 전부 개정(2020.1.16.)하고, 중대한 인명 피해를 주는 산업재해가 발생하였을 경우 사업주에 대한 형사처벌을 강화하는 법안으로 중대재해 처벌법을 제정(2021.1.26.)하게 되었다.

현재 우리는 데이터의 세상에서 살고 있다. IT기술의 발달로 방대한 양의 텍스트데이터가 빠른 속도로 유통되는 시대의 흐름 속에서 다양한 매체를 통해 특정 주제에 대한 동향을 살펴보는 것은 가치로운 정보를 물색하고, 현재 주목해야 할 과제를 파악해 볼 수 있는 연구방법이 될 수 있다(McCombs & Shaw, 1974). 대부분의 언론매체가 온라인 형태로 정책에 대한 비평, 현실과 문제점, 사회의 관점을 반영하여 사회적 현상을 뉴스형태로 전달한다. 특히, 온라인 뉴스기사 매체가 포함된 비정형화된 텍스트 데이터 속에서 산업안전에 대한 내용을 분석해 보는 것은 이에 대한 우리 사회의 인식 정도를 파악하고, 이를 바탕으로 향후 이루어져야 할 연구 과제와 방향을 제시해 줄 것이라 기대할 수 있다. 뿐만 아니라 신문, 홈페이지, 포털, 블로그와 소셜 네트워크 서비스(커뮤니티, 페이스북, 인스타그램, 트위터, 뉴스) 등 다양한 서비스를 통해서 기사를 표현할 수 있다. 다양한 소셜 네트워크 서비스의 등장으로 언론 뿐만 아니라 개인들도 자유롭게 생각과 의견을 표현할 수 있다.

산업안전 관련 분야에서는 정형 데이터만을 분석 대상으로 분석하고 통계처리를 하고 있으며, 비정형 데이터를 분석에 적용한 연구가 거의 이루어지지 않고 있다. 텍스트 키워드 분석을 통한 산업안전 관련 기사의 분석으로 주요 키워드 출현빈도의 중요도를 분석하고 특징적인 패턴을 지닌 키워드를 구조적 분석을 파악할 수 있다. 언론보도 기사와 소셜 네트워크 서비스를 통해 각기 다른 관점으로 이슈나 정책을 바라보고 다양한 방법으로 해석할 수 있을 것이다. 이는 2016년 서울 지하철 2호선 승강기 사고, 2017년 삼성중공업 크레인 사고, 2018년 문재인 정부의 국민생명 지키기 프로젝트, 2019년 태안 화력 발전소 사고, 2020년 코로나와 중대재해처벌법 제정, 2021년 광주 건물철거 붕괴사고와 중대재해처벌법 등 산업안전 관련 이

슈가 발생하면 국민들은 언론보도를 통해 소식을 접하고, 소셜 네트워크 서비스를 통해 의견과 감정을 표현한다. 그로 인해 국민의 관심이 높아지면 사회적 관심사가 되고 제도개선으로 이어지기도 한다. 산업안전에 대한 시각은 시계열적인 변화를 가져왔을 것으로 동향을 살펴볼 필요가 있다.

1.2 연구의 목적

본 연구에서는 언론 보도 기사와 소셜 네트워크 서비스 데이터를 텍스트 마이닝(Text-Mining) 기법을 활용하여 분석하고자 하였다. 특히 소셜 서비스의 경우 사회변화를 능동적으로 모니터링하고 빠른 속도로 데이터 처리하여 정량화, 객관화, 시각화를 통해 효율적 가치판단을 할 수 있는 데이터로 주목받고 있으며, 무엇보다 정책이나 사회적 이슈에 대한 국민들의 인식 및 영향관계를 파악할 수 있다는 점에서 적절한 방법론으로 판단하였다. 따라서 본 연구에서는 언론 보도 기사를 수집하고 텍스트 마이닝 기법을 통해 산업안전과 관련된 주요 키워드를 분석하고 소셜 네트워크 서비스 데이터 분석을 통한 산업안전 분야의 사고와 정책 변화에 따른 국민들의 인식을 살펴보고 텍스트 마이닝 기법의 활용 방안을 제시하고자 한다.

1.3 연구의 범위 및 내용

본 연구에서는 산업안전 관련 키워드로 한정하여 언론 보도 기사와 소셜 네트워크 서비스 데이터를 분석하였다. 제 1장은 서론으로서 연구의 배경 및 필요성, 목적, 범위와 내용으로 구성하였고, 제2장은 문헌 연구로 이론적 배경을 바탕으로 선행연구 사례를 분석하였다. 제3장은 연구방법으로 연구대상, 연구방법 5단계(데이터 수집, 데이터 정제, 데이터 분석, 데이터 시각화, 결과분석)로 각 단계별 절차를 설명하였다. 제 4장은 산업안전 분야의 최근 6년간(2016년 ~ 2021년 6월) 텍스트 마이닝 분석과 소셜 네트워크 서비스 데이터의 감정키워드 분석을 실시하였고, 데이터를 시각화하여 표현하였다. 제5장은 분석결과 및 맺음말로 구성하였다.

2. 문헌연구

2.1 빅 데이터(Big data)

최근 SNS, 클라우드 컴퓨팅 IT 기술의 발전에 따라 다양한 분야에서 정보의 양이 급속히 증가되고 있으며 지난 몇 년간 정보저장 및 통신 기술의 비약적인 발전에 힘입어 누구나 손쉽게 정보를 생성하고 공유할 수 있는 기반이 마련되었다. 이에 따라 많은 양의 데이터가 빠르게 축적되기 시작하였으며 이전에는 불가능했던 규모의 데이터들을 신속하게 분석할 수 있게 되었다(이주용, 2013).

빅데이터(Big Data)란 디지털 환경에서 생성되는 데이터로 그 규모가 방대하고, 생성 주기도 짧고, 형태도 수치 데이터뿐 아니라 문자와 영상 데이터를 포함하는 대규모 데이터를 말한다. 빅데이터 환경은 과거에 비해 데이터의 양이 폭증했다는 점화 함께 데이터의 종류도 다양해져 사람들의 행동은 물론 위치정보와 SNS(Social Network Services)를 통해 생각과 의견까지 분석하고 예측할 수 있다(정용찬, 2013)

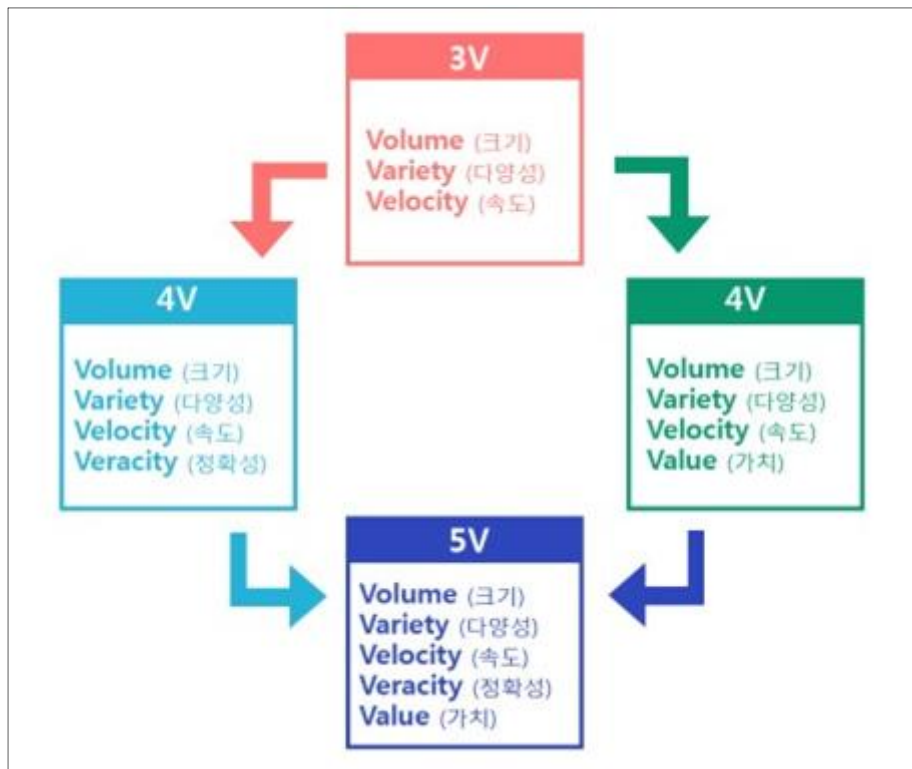
스마트폰, 스마트 TV 등 전자제품의 스마트화와 트위터(Twitter), 페이스북(Facebook), 인스타그램(Instagram), 블로그(Blog)와 같은 소셜 네트워크 서비스(Social Networking Service, SNS)의 확산은 빅데이터와 높은 관련성을 가지고 있으며, 이용자들이 생성한 데이터 형식과 무관하게 데이터의 양적인 증가로 인해 빅데이터의 생성, 가공, 저장, 분석결과 및 시각화에 관심이 증가되는 추세이다. 또한, 컨설팅 기관인 매킨지(Mckinsey, 2011)에서 “빅데이터(Big data)를 기존 방식으로 저장, 관리, 분석을 할 수 있는 범위를 넘어서는 규모의 데이터”라고 정의하였으며, 다른 컨설팅 기관인 가트너(Gartner, 2011)에서는 “21세기 원유로 다양한 종류의 데이터가 기업이 감당할 수 없을 정도로 빠르게 생성되는 현상”으로 정의하였다. IDC(Industrial Development Corporation)은 “다양한 데이터로 구성된 방대한 볼륨의 데이터로부터 고속 캡처, 데이터 탐색 및 분석을 통해 경제적으로 필요한 가치를 추출할 수 있도록 디자인된 차세대 기술과 아키텍처”라고 정의하고 있다.

빅데이터의 정의는 선행연구들에서 다양하게 제시되고 있으며, 위키백과(Wikipedia)는 “기존 데이터베이스 관리도구의 능력을 넘어서는 대량(수십 테라바이트)의 정형 또는 심지어 데이터베이스 형태가 아닌 비정형의 데이터 집합조차 포함한 데이터로부터 가치를 추출하고 결과를 분석하는 기술”로 정의하였다. 즉, 데이터베이스 등 기존의 데이터 처리 응용 소프트웨어(data-processing application software)로는 수집 · 저장 · 분석 · 처리하기 어려울 정도로 방대한 양의 데이터를 의미한다.

Gartner(2011)에 빅데이터의 데이터의 규모가 크고 (high-volume), 종류가 다양하고 (high-variety), 실시간으로 처리해야 하는(high-velocity)의 3가지의 특징을 지닌 향상된

통찰력, 의사 결정 및 프로세스 자동화를 가능하게 하는 비용 효율적이고 혁신적인 형태의 정보처리를 요구하는 자산으로 정의하였으며, 2012년에 추가로 데이터 관리의 복잡성 (Complexity)속성을 더하여, 빅 데이터의 특징을 3VC로 정의하였다.

빅데이터의 특징을 살펴보면 가트너에서 정의한 기본 3가지 특징(크기, 다양성, 속도)에 다양한 특징들이 추가된다. 추가 될 수 있는 특징 중 하나가 정확성이다. 방대한 양의 데이터를 분석하여 일정한 패턴을 추출하고, 일정 패턴을 설명할 수 있는 신뢰성을 의미하는 정확성(veracity)와 소셜 네트워크 서비스의 확산으로 자기 의견을 웹을 통해 자유롭게 게시하는데 자신의 의도와 다르게 다른사람은 해석될 수 있는 가변성(Variability), 정형 및 비정형 데이터를 수집하여 분석을 실행하여 정보를 가공을 하게 된다. 가공한 정보를 정보의 사용 대상자가 쉽게 이해할 수 있는 시각화(Visualization)의 특성을 가진다. 따라서, 빅데이터의 특징은 특정 한 분야를 나타내거나 한 가지 특징을 개념으로 사용하기 어려운 부분이 있다. 사용하는 분야마다 중요하게 생각하는 부분이 다를 것이고, 필요 없는 부분도 있을 수 있기 때문이다. 이러한 부분이 빅데이터의 정의를 어렵게 만들고 있다. 앞으로도 빅데이터의 특징은 새롭게 추가될 수 있을 것이다.



<그림 2-1> 빅데이터 특징, 네이버 지식백과 재구성

안효선(2017)은 문헌고찰과 실증적 고증을 통해 텍스트 마이닝과 연결망 의미연결망 분석 방법을 활용하여 네이버와 다음 블로그의 사용자들이 게시한 문서, 댓글 등의 데이터를 분석하였다. 일반 사용자들이 주요하게 언급하고 있는 현실언어 기반의 어휘를 도출하고, 시대적, 사회적 상황을 반영하는 속성과 감정 어휘간의 관계를 파악하여 사용자 중심으로 활용하고자 하였다.

홍석은(2018)은 디지털 민원 텍스트 자료로부터 민원의 유형과 이슈를 추출하고 민원 간 연관관계를 정의하여 의미 있는 토픽을 제시할 수 있는 빅 데이터 분석방안을 제시하고자 하였다. 연구에서 환경민원 빅 데이터를 활용하여 비정형 텍스트 마이닝 기법과 민원인의 감정 척도를 산정하기 위해 토픽 모델링과 감성분석을 실시하였다. 민원 데이터에 대한 단어 사전을 구축하였으며 LDA(Latent Dirichlet Allocation) 알고리즘을 이용하여 민원 문장과 단어와의 관계 속에 주제문장을 분류하는 토픽 모델링을 수행하고, 민원 유형별 대표단어에 대한 감정 점수를 산정하여 공간분포를 제시하였다.

강현빈(2019)은 한국어 비정형 텍스트 데이터를 문서 내부의 정보를 학습하여 특정 범주로 분류하는 모델을 구현하고 한국어로 축적된 건설재해사례 데이터를 재해 유형별로 분류하는 실험을 통하여 분류 성능을 개선할 수 있는 방안을 데이터 관리 및 분석의 측면에서 제시하고자 하였다.

이희영(2019)의 연구에서는 미세먼지 문제가 장기화됨에 따라 대중들의 불안이 증폭되고 있는 시점에서 미디어가 미세먼지의 원인에 대해 어떤 정보를 전달하고 있으며, 이러한 정보를 바탕으로 구성된 대중들의 인식을 파악하고자 하였다. 기사와 댓글의 연도별 중심의제를 파악하여 비교하였고, 상관관계 분석을 통해 기사 속성 네트워크와 댓글 속성 네트워크의 유사성을 검증하여 온라인 뉴스 기사와 댓글 간 네트워크 의제 설정 효과가 존재함을 확인하였다.

문진(2019)의 연구에서는 안전사고 사례 분석에 있어서 기존의 방법과 구분이 되는 텍스트 마이닝기법과 언어 네트워크 분석 방법을 적용한 방법으로 실시하였다. 630건의 자료를 기존의 방법으로 분류 분석하여 비교해 본 결과, 재해유형에 따른 분류에 따른 주요한 재해발생 원인, 기인물에 따른 분류에서의 주요 기인물 등 분석 결과가 일치하는 것으로 나타났다. 주요 재해에 대한 정보를 쉽게 얻을 수 있고, 관련된 인자를 바로 확인 가능하며, 또한 마이닝을 통하기 때문에 직접적인 빈도를 추출하므로 정보가 매몰되지 않아 핵심 키워드를 직관적으로 파악할 수 있었다. 감정분석이나 주제분석과 연결하여 다양한 분석방법으로 다양하게 적용할 수 있고, 위험성평가의 경우 위험도의 분석과 내용의 적절성 분석, 판별에 도움이 될 것이라고 하였다.

2.2 텍스트 마이닝

많은 정보들이 온라인 뉴스 기사, 기술문서, 도서, 전자 우편(이메일) 메시지, 블로그(blog), 소셜 네트워크 서비스(SNS) 및 웹사이트의 다양한 형태의 텍스트 형식으로 저장된다. 이렇게 공개된 다양하고 풍부한 텍스트 정보에서 특정 주제와 관련한 부분을 뽑아 의미를 분석하고 사회 현상이나 여론의 경향 등 고품질의 정보를 도출하기 위한 방법으로 텍스트 마이닝 기법을 활용한다. 데이터 마이닝이 구조화되고 사실적인 방대한 데이터베이스에서 관심 있는 패턴을 찾아내는 기술 분야라면 텍스트 마이닝은 텍스트를 분석하고 구조화하여 의미를 찾아내는 기술 분야다.

텍스트 마이닝(text mining)의 기반 기술로는 대용량 텍스트 데이터를 저장하고 처리하는 빅데이터 기술과 텍스트 데이터 구조를 분석하고 포함된 정보를 통계 처리가 가능한 형태로 변환하는 자연어 처리(NLP: Natural Language Processing) 기술이 있다. 이를 바탕으로 데이터 안에서 단어의 출현 빈도를 파악하는 단어 빈도 분석(frequency analysis), 유사한 단어 또는 비슷한 성격의 단어들을 묶어주는 군집 분석(cluster analysis), 단어에 나타난 긍정 혹은 부정 등의 감정적 요소를 추출하여 그 정도를 판별하는 감성 분석(sentiment analysis) 그리고 서로 다른 단어가 동시에 나타날 확률에 기초하여 단어 간 연관성을 추출하는 연관 분석(association analysis) 등의 통계적 방법들이 사용된다.

김준수(2015)는 건설현장에서 발생한 안전사고와 관련된 인터넷 기사 7,384건의 분석을 통해 현장에서 빈번히 발생하는 사고유형의 경중 도출과 주성분 분석을 통한 계절별 사고 유형의 요인 도출을 하였다. 기존 연구방식으로 분석한 결과와 비슷한 결과를 도출하고, 본 연구의 방법론인 인터넷 안전사고 기사 검색을 통한 텍스트 및 데이터 마이닝 기법의 적용이 유효하다는 것을 입증하였다.

현진우(2015)는 도로교통공단 교통사고분석시스템에서 제공하는 2014년 교통사고데이터를 기반으로 서울특별시에서 발생한 버스 사고 중 버스 차내 안전사고를 분류하여 분석하였다. 분류한 버스 차내 안전사고 데이터 중 텍스트 기반의 사고개요 데이터를 통계분석 패키지 R Project를 활용하여 텍스트 마이닝 분석을 통해 버스 차내 안전사고와 관련된 단어들의 빈도수를 파악한 후 각 단어들이 갖는 의미들을 분석하여 사고 유형과 원인을 구분하여 재정립하였으며 사고 유형별 원인에 따른 해결방안을 추가적으로 제시하였다.

김연철(2016)은 건설현장에서 발생 가능한 건설현장 안전사고를 사전에 예측하는데 있어 데이터마이닝 기법을 적용하고 그 활용 가능성을 검토하고자 하였다. 데이터마이닝 기법을 활용하여 사전에 발생 가능한 안전사고를 보다 정확하게 예측하고, 예방적 활동을 지원하기 위한 의사결정지원 도구로서 유효하다는 것을 확인하였다.

김홍관(2020)은 위험성평가와 관련 연구논문과 언론보도 기사에서 산업재해 원인 키워드들을 추출하여 키워드간 네트워크 분석방법으로 산업재해의 위험 요인에 대한 구조적인 분석을 시도하였으며 TF-IDF(Term Frequency-Inverse Document Frequency) 분석, 키워드 간의 중심성 분석, 네트워크 간의 유사성 분석을 위한 QAP(Quadratic Assignment Procedure) 상관분석, MR-QAP(Multiple Regression-Quadratic Assignment Procedure) 회귀분석, 키워드간 상관관계의 CONCOR(CONvergence of iterated CORrelation) 분석 등을 수행하였다.

문진혁(2020)은 전자산업 분야의 산업재해 현황분석을 실시하였다. 심층 분석을 위해 R 패키지를 활용하여 워드클라우드, 토픽 모델링 등 텍스트 마이닝 분석기법을 통해 중대재해조사 보고서를 분석하였다. 최근 10년간 발생한 전자산업 분야의 산업재해 분석을 통해 질병사망자의 비율과 직업성 암의 증가를 확인하였으며, 전자산업의 안전보건 증진을 위해 산업생태계의 정의·분류가 필요하며, 전체적인 재해를 줄이기 위해서 사고재해 예방 및 감소의 집중이 필요하다고 연구하였다.

2.3 감정분석

감정분석은 '텍스트 문서에서 감성이 어떻게 표현되고 있는지, 특정한 주제에 대해 긍정 혹은 부정적 표현들이 포함되어 있는지를 확인하는 방법론'이다. SNS가 사람들의 감정과 의견을 표현하는 도구가 되어가면서 그들이 남긴 텍스트들은 특정대상에 대한 감성을 추출하여 마케팅 또는 정치적 캠페인에 많이 이용되고 있다. 감정분석은 크게 두 가지 수준에서 분석이 이루어진다. 첫째로 문서 수준의 감정분석을 통해 문서의 긍정 또는 부정적 성향을 판단한다. 분석은 문서 내에 있는 특정 키워드에 대한 긍정 또는 부정을 분석하는 것이 아닌 감성표현의 단어, 어휘들의 출현 빈도를 통해 그 문서의 긍정/부정의 정도를 판별해낸다. 둘째, 구나 문장 수준의 분석방법으로 감정분석을 이용해 리뷰에서 감성을 포함한 문장을 추출하고 그 문장을 대상으로 키워드 측면에서 키워드들에 대한 감성을 분석하는 방법이다. 한국어의 경우 감성사전에 대한 구축이 되어있지 않은 상태이기 때문에 감성사전부터 작성하는 연구들이 많다. 이러한 연구들은 감성사전을 작성하기 위해 회귀분석과 같은 기계적 학습 기반 예측기반을 이용하며 사람들의 평가를 통해 부여받은 문장들로 구성된 트레이닝 데이터를 생성하여 이 문장을 구성하는 단어들과 감성의 적재 값으로 구성된 감성사전을 작성한다. 이러한 회귀식은 독립변수가 수 만개의 단어로 구성되기 때문에 차원축소를 위해 라쏘(lasso), 릿지(ridge), 엘라스틱(elastic) 등의 축소기법을 적용하기도 한다. 영어권의 경우 공인된 감성사전인 WordNet을 이용하여 어휘의 의미 극성을 판단하여 이를 바탕으로 긍정 또는 부정의 단순 구분이 아닌 감성의 정도 값을

정의한 SentiWordNet 관련 연구가 있다(김동성, 2015).

이수아(2018)은 안전과 관련된 규제들은 국민의 생명과 건강과 관련되어 있어 합리적인 생각보다는 감정이 지배하기(emotion-laden) 쉬운 분야로, 재해·재난 사고가 많았던 최근 안전과 관련된 규제를 강화해야 한다는 주장 자체가 당위성을 가지며 규제에 대한 제대로 된 평가가 잘 이루어지고 있지 않다. 산업안전보건규제의 효과를 '산업재해의 감소'의 차원에서 실증적으로 평가하고자 하였고, 안전과 보건의 차원으로 규제효과를 구분하여 규제별로 효과를 평가하고 비교하였다.

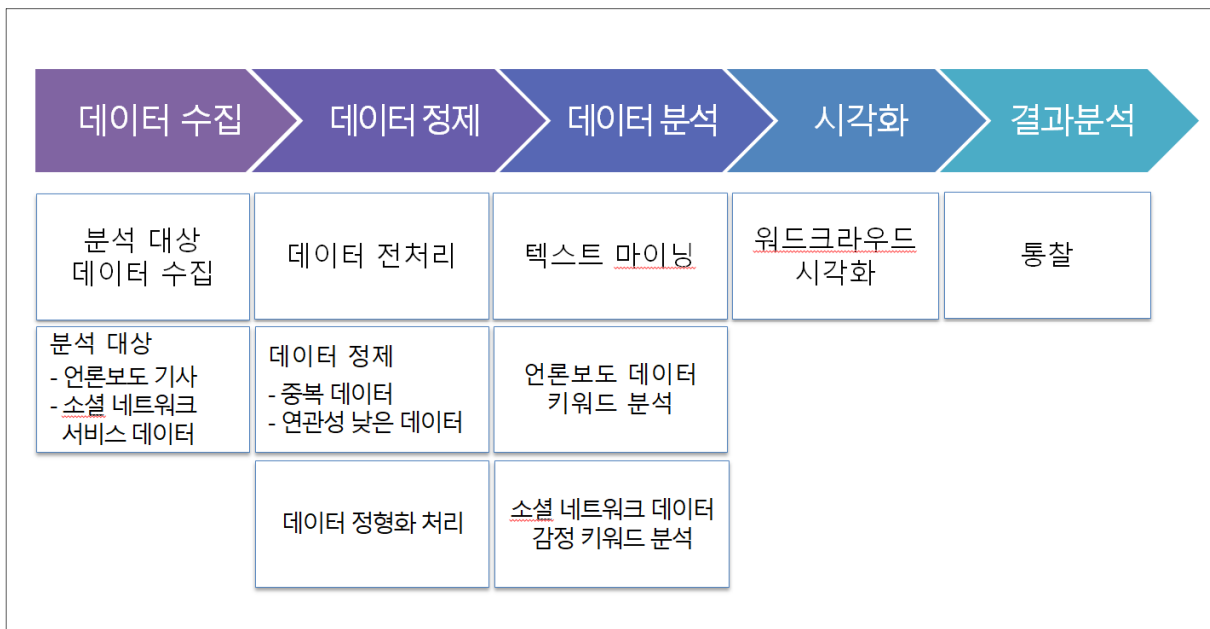
최정용(2021)은 소셜 네트워크 서비스 중 하나인 트위터에 생성되는 비정형 자료를 이용하여 이용자의 감정 상태를 네 가지 감정상태(기쁨, 분노, 슬픔, 즐거움)로 분류하는 분석을 진행하였다. 검색 키워드로 고빈도 감정 단어를 추출하여 회귀분석, 라소(lasso) 회귀분석, 능형(ridge) 회귀분석 그리고 엘라스틱 넷(elastic net) 회귀분석을 이용하여 감정 단어 사전을 생성하고, 네 종류의 감정 단어 사전을 이용하여 네 가지 분석모델로 분류 연구하였다.

김동준(2021)은 텍스트 마이닝을 통해 소셜 커머스 사용자 경험과 서비스 품질 개선에 관한 내용을 살펴보기 위하여 웹데이터를 크롤링하여 감성어 분석 후 분류분석, 회귀분석, 토픽 모델링과 워드 클라우드 분석하였다. 서비스 재사용의도 분석한 결과 소셜 네트워크 서비스(트위터, 블로그, 인스타그램)별 차이를 확인하였다. 텍스트 데이터에 토픽 모델링 결과를 바탕으로 마케팅 부분과 제품 전략 개선안을 도출할 수 있다고 하였다.

3. 연구방법

3.1 연구방법 5단계

본 연구는 5단계로 진행하였다. 1단계는 연구대상의 데이터를 수집하는 단계이다. 연구 주제인 산업안전 관련 데이터 수집을 위해 언론보도 기사와 소셜 네트워크 서비스 데이터를 수집한다. 2단계는 데이터를 정제하는 과정으로 수집된 데이터 중 중복 데이터와 연관성이 낮은 데이터를 제거하는 과정이다. 불필요한 데이터를 제거하고 언론보도 기사와 소셜 네트워크 데이터의 비정형 데이터를 정형화 데이터로 변환한다. 3단계는 데이터 분석과정으로 데이터 정제를 수행한 데이터를 대상으로 파이썬 프로그램을 활용하여 텍스트마이닝 분석한다. 언론보도 데이터를 키워드와 소셜 네트워크 데이터의 감정 키워드를 분석한다. 데이터 수집 기간 중 시계열 분석을 위해 연도를 구분하여 분석한다. 4단계는 분석한 텍스트마이닝 결과를 바탕으로 파이썬의 시각화 솔루션(워드클라우드)를 활용하여 시각화를 수행한다. 5단계는 결과를 분석하는 통찰의 과정으로 산업안전의 텍스트 마이닝 분석의 키워드를 다른 관점에서 분석한다. 아래의 <그림 3-1>은 연구분석 단계를 도표로 표현한 것이다.

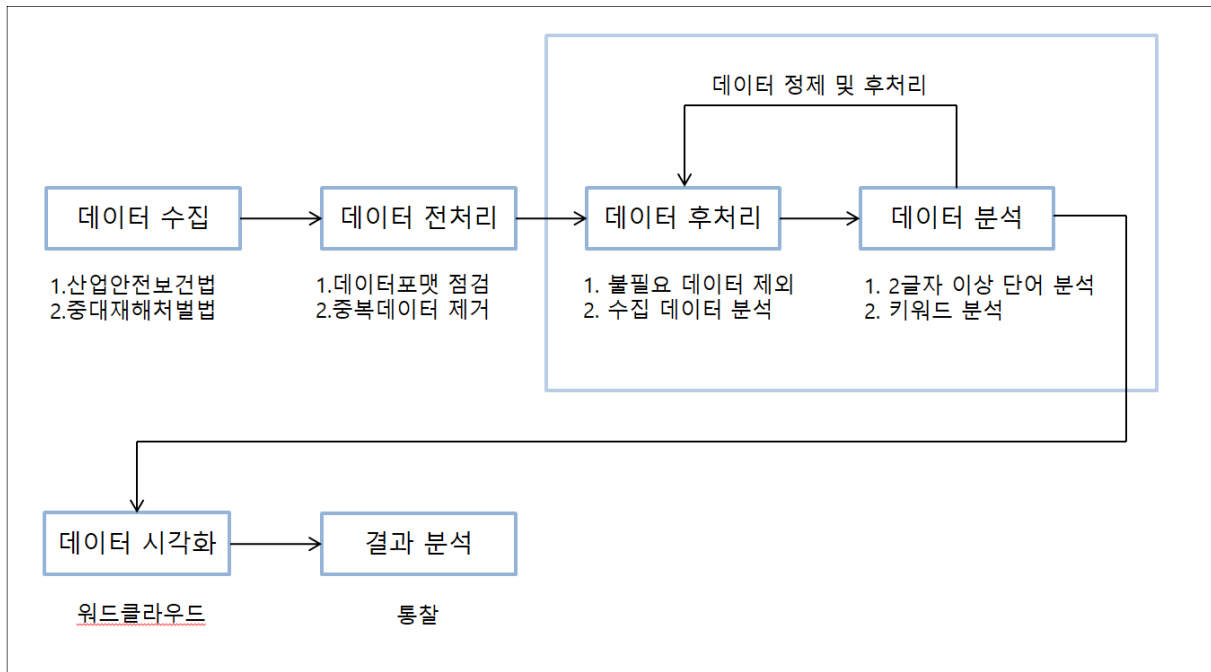


<그림 3-1> 연구방법 5단계

3.2 연구방법 적용사례

연구방법을 산업안전과 가장 관련이 있는 법령(산업안전보건법, 중대재해처벌법)을 대상으로 분석하였다. 산업안전보건법, 중대재해처벌법을 대상으로 법령 데이터를 수집하고, 데이터의 정제과정과 분석과정을 수행하고, 수행한 결과 데이터를 파이썬 프로그램으로 시각화를 수행하고, 결과 분석을 통찰하는 과정으로 진행한다.

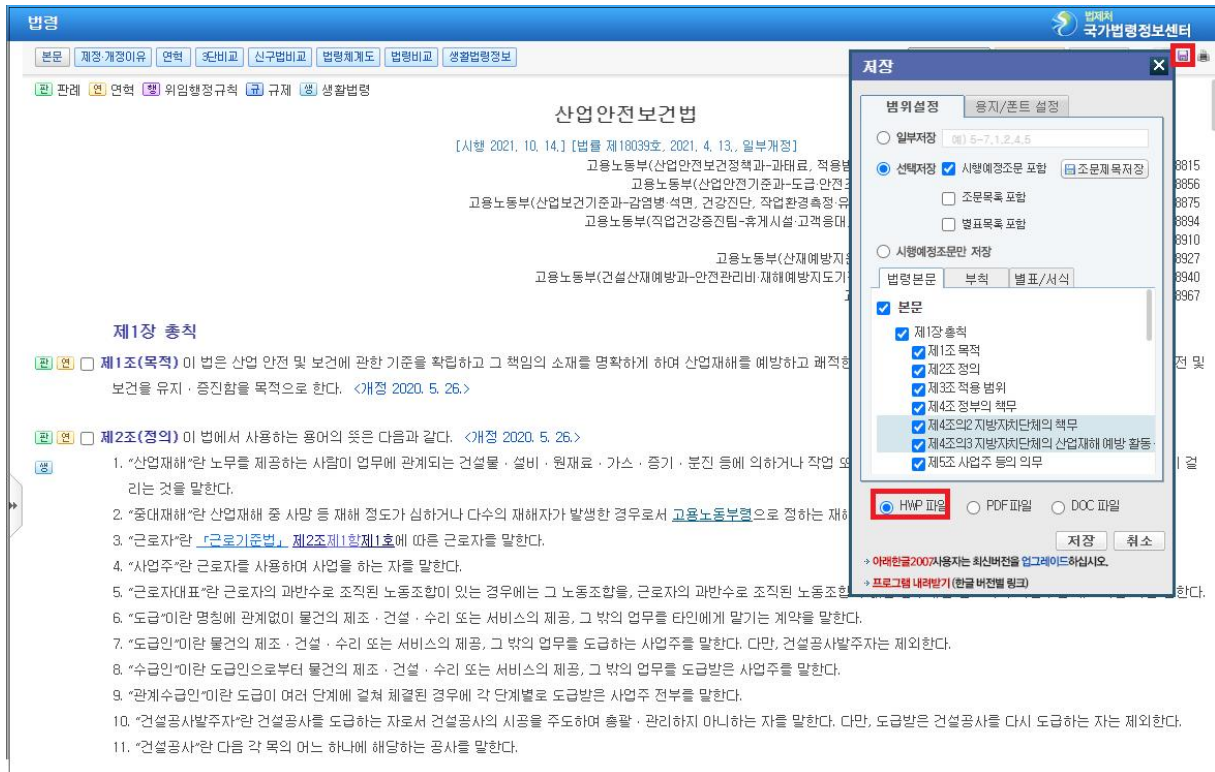
연구방법 적용사례의 세부 과정은 6단계로 구성된다. 1단계는 데이터 수집 단계로 법제처의 산업안전보건법, 중대재해처벌법 법령 데이터를 다운받는다. 2단계는 데이터 전처리 단계로 법령 데이터의 파이썬 프로그램을 활용하여 분석하기 위해 데이터 포맷을 엑셀파일 형태로 변환한다. 3단계는 데이터 후처리 단계로 법령 데이터의 불필요한 키워드를 제외하고, 새로운 단어를 사전에 등록한다. 4단계는 데이터 분석 단계로 키워드 분석한 결과 중 중복되거나 의미없는 단어를 찾고, 신규 단어가 누락되었는지 검증하는 과정으로, 분석결과의 품질을 높이기 위한 정제과정으로 전 단계의 후처리 과정을 반복수행한다. 5단계는 데이터 시각화 단계로 파이썬 프로그램(워드클라우드)를 활용하여 분석된 데이터를 시각화 하는 과정이다. 6단계는 결과 분석 단계로 두 법령을 비교하여 데이터를 해석하는 과정이다. 전체적인 데이터 분석 세부과정을 <그림 3-2>로 표현하였다.



<그림 3-2> 연구방법 적용사례 세부과정

3.2.1 데이터 수집 단계

데이터 수집 단계는 법제처의 국가법령정보센터(<https://www.law.go.kr>)에서 “산업안전보건법”을 검색한다. 산업안전보건법을 상단의 아이콘에서 다양한 파일의 형태로 저장할 수 있다. 아래의 <그림 3-3>은 국가법령정보센터의 산업안전보건법 검색 화면이다.



<그림 3-3> 국가법령정보센터의 산업안전보건법

3.2.2 데이터 전처리 단계

데이터 전처리 과정은 본문 내 줄바꿈, 공백 등 불필요한 데이터를 삭제하고, 전체 데이터를 파이썬 프로그램에서 분석할 수 있는 엑셀파일 형태로 저장한다. 산업안전 보건법 [시행 2021. 10. 14.]파일의 전처리 과정을 끝낸 파일은 1137행으로 구성되어 있다. 엑셀파일을 'law.xlsx'로 저장한다. 아래의 <그림 3-4>는 산업안전보건법의 엑셀파일 확인과정이다.

1128	13. 제122조제3항을 위반하여 고용노동부장관에게 신고하지 아니한 자								
1129	14. 제124조제1항에 따른 증명자료를 제출하지 아니한 자								
1130	15. 제125조제5항, 제132조제5항 또는 제134조제1항·제2항에 따른 보고, 제출 또는 통보를 하지 아니하거나 거짓으로 보고, 제출 또는 통보한 자								
1131	16. 제155조제1항(제166조의2에서 준용하는 경우를 포함한다)에 따른 질문에 대하여 답변을 거부·방해 또는 기피하거나 거짓으로 답변한 자								
1132	17. 제156조제1항(제166조의2에서 준용하는 경우를 포함한다)에 따른 검사·지도 등을 거부·방해 또는 기피한 자								
1133	18. 제164조제1항부터 제6항까지의 규정을 위반하여 서류를 보존하지 아니한 자								
1134	㉠ 제1항부터 제6항까지의 규정에 따른 과태료는 대통령령으로 정하는 바에 따라 고용노동부장관이 부과·징수한다.								
1135	[시행일: 2021. 11. 19.] 제175조								
1136	부칙 <제18039호,2021. 4. 13.>								
1137	이 법은 공포 후 6개월이 경과한 날부터 시행한다.								

<그림 3-4> 산업안전보건법의 엑셀파일 확인과정

법령내용을 저장한 엑셀파일 “law.xlsx”을 파이썬으로 불러온다. 파이썬 프로그램의 엑셀 파일을 읽는 load_workbook 함수로 저장한 엑셀파일을 불러온다. 아래의 <그림 3-5>는 법령파일을 파이썬 프로그램에서 불러오는 소스코드의 부분이다.

```
#엑셀파일 불러오기
read_workbook = load_workbook(filename='law.xlsx')
read_cell = read_workbook.active
```

<그림 3-5> 법령파일 파이썬에서 불러오기

불러온 파일의 끝까지 읽었는지 확인한다. 엑셀파일은 1137행으로 구성되어 있으며, 파이썬 프로그램은 0부터 1136까지 1137개의 행을 불러온다. 아래의 <그림 3-6>은 파이썬 프로그램에서 법령파일(엑셀파일)로 불러오기를 실행한 화면이다.

```
In [3]: keyword_list = []
        for i in range(1,1137):
            keyword_list.append(read_cell.cell(i+1,1).value)
        print(keyword_list)
```

으로 정하는 작업환경측정의 방법을 준수하지 아니한 사업주(같은 조 제3항에 따라 작업환경측정기관에 위탁한 경우는 제외한다)', '14. 제125조 제4항 또는 제132조제1항을 위반하여 근로자대표가 요구하였는데도 근로자대표를 참석시키지 아니한 자', '15. 제125조제6항을 위반하여 작업환경 측정 결과 해당 작업장 근로자에게 알리지 아니한 자', '16. 제155조제3항(제166조의2에서 준용하는 경우를 포함한다)에 따른 명령을 위반하여 보고 또는 출석을 하지 아니하거나 거짓으로 보고한 자', '㉠ 다음 각 호의 어느 하나에 해당하는 자에게는 300만원 이하의 과태료를 부과한다. <개정 2020. 3. 31.>', '1. 제32조제1항(제5호의 경우만 해당한다)을 위반하여 소속 근로자로 하여금 같은 항 각 호 외의 부분 본문에 따른 안전보건교육을 이수하도록 하지 아니한 자', '2. 제36조를 위반하여 근로자대표에게 통지하지 아니한 자', '3. 제40조(제166조의2에서 준용하는 경우를 포함한다), 제108조제5항, 제123조제2항, 제132조제3항, 제133조 또는 제149조를 위반한 자', '4. 제42조제2항을 위반하여 자격이 있는 자의 의견을 듣지 아니하고 유해위험방지계획서를 작성·제출한 자', '5. 제43조제1항 또는 제46조제2항을 위반하여 확인을 받지 아니한 자', '6. 제73조제1항을 위반하여 지도를 받지 아니한 자', '7. 제84조제6항에 따른 자료 제출 명령을 따르지 아니한 자', '8. 삭제<2021. 5. 18.>', '9. 제111조제2항 또는 제3항을 위반하여 물질안전보건자료의 변경 내용을 반영하여 제공하지 아니한 자', '10. 제114조제3항(제166조의2에서 준용하는 경우를 포함한다)을 위반하여 해당 근로자를 교육하는 등 적절한 조치를 하지 아니한 자', '11. 제115조제1항 또는 같은 조 제2항 본문을 위반하여 경고표시를 하지 아니한 자', '12. 제119조제1항에 따라 일반선면조사를 하지 아니하고 건축물이나 설비를 절거하거나 해체한 자', '13. 제122조제3항을 위반하여 고용노동부장관에게 신고하지 아니한 자', '14. 제124조제1항에 따른 증명자료를 제출하지 아니한 자', '15. 제125조제5항, 제132조제5항 또는 제134조제1항·제2항에 따른 보고, 제출 또는 통보를 하지 아니하거나 거짓으로 보고, 제출 또는 통보한 자', '16. 제155조제1항(제166조의2에서 준용하는 경우를 포함한다)에 따른 질문에 대하여 답변을 거부·방해 또는 기피하거나 거짓으로 답변한 자', '17. 제156조제1항(제166조의2에서 준용하는 경우를 포함한다)에 따른 검사·지도 등을 거부·방해 또는 기피한 자', '18. 제164조제1항부터 제6항까지의 규정을 위반하여 서류를 보존하지 아니한 자', '㉠ 제1항부터 제6항까지의 규정에 따른 과태료는 대통령령으로 정하는 바에 따라 고용노동부장관이 부과·징수한다.', '[시행일: 2021. 11. 19.] 제175조', '부칙 <제18039호,2021. 4. 13.>', '이 법은 공포 후 6개월이 경과한 날부터 시행한다.'

<그림 3-6> 파이썬에서 법령파일(엑셀파일) 불러오기 확인

3.2.3 데이터 후처리 단계

데이터 후처리 단계는 비정형 데이터를 분석이 가능한 형태인 정형 데이터로 변환한다. 단어사전의 자연어 처리를 위해 서울대학교 IDS(Intelligent Data Systems) 연구실의 꼬꼬마(kkma) 라이브러리를 활용하여 분석하였다. 데이터 분석 과정은 추출된 키워드 중 두 글자 이상의 단어를 분석하였다. 한글자 단어는 정확한 의미를 확인하기 어려울 것으로 판단하여, 두 글자 이상의 단어를 대상으로 분석 하였다. 법령집 문장을 명사사전을 이용하여 키워드의 빈도를 추출한다. 후처리 과정에서 숫자, 접속사, 의미 없는 단어를 불용 키워드를 제거하였다. 불용 키워드 목록은 stopwords에 정의한다. 아래의 <그림 3-7>은 데이터 후처리 과정과정으로 2글자 이상의 단어의 빈도를 계산하고 불용 키워드를 제거 하는 소스코드 부분이다.

```
# 기사 읽어서 명사 추출
for i in keyword_list :
    list_temp = list_temp + kkma.nouns(i)

#추출된 데이터 중 2글자 이상의 단어로 정제
for j in list_temp :
    if len(j) > 1 :
        list_temp2.append(j)

#불용 키워드 제외처리
list_temp3 = [ k for k in list_temp2 if k not in stopwords ]
```

<그림 3-7> 데이터 후처리 소스코드 일부

3.2.4 데이터 분석 단계

산업안전보건법령을 2글자 이상의 단어로 다빈도 추출하였다. 추출한 결과 안전 373회, 고용 364회, 보건 264회, 노동부 223회, 장관 216회, 고용노동부장관 215회, 노동 211회, 고용노동부령 209회, 필요 183회, 사항 181회 순으로 나타났다. 분석된 결과를 확인하고 불필요한 데이터가 확인되면 데이터 후처리 단계의 불용 키워드에 목록을 추가한다. 분석 과 후처리 과정을 반복하여 수행한다. 아래의 <그림 3-8>은 단어의 빈도수를 계산하고 다빈도 100개를 출력하는 화면이다.

```
#단어의 빈도수 계산
tags = collections.Counter(list_temp3).most_common(100)

#최다빈도 순으로 화면출력
print(tags)

[('안전', 373), ('고용', 364), ('보건', 264), ('노동부', 223), ('장관', 216), ('고용노동부장관', 215), ('노동', 211), ('고용노동부령', 209), ('필요', 183), ('사항', 181), ('사업주', 172), ('근로자', 163), ('산업', 152), ('기관', 143), ('재해', 124), ('작업', 113), ('위반', 111), ('대통령령', 107), ('업무', 105), ('예방법', 96), ('산업재해', 94), ('관리', 91), ('조치', 89), ('방법', 86), ('건강', 83), ('물질', 80), ('사업장', 79), ('포함', 75), ('지도', 74), ('기준', 72), ('자료', 71), ('결과', 70), ('대상', 68), ('유해', 64), ('개정', 62), ('검사', 62), ('진단', 62), ('사용', 60), ('조사', 60), ('준용', 60), ('제출', 59), ('위험', 58), ('지정', 57), ('발생', 56), ('교육', 56), ('건설', 56), ('확인', 56), ('명령', 55), ('규정', 55), ('평가', 54), ('사람', 53), ('사업', 53), ('인증', 51), ('수행', 51), ('제공', 49), ('전문', 49), ('제조', 44), ('공사', 44), ('절차', 44), ('근로', 43), ('물질안전보건자료', 43), ('기계', 43), ('환경', 42), ('자율', 42), ('변경', 42), ('도급', 41), ('설비', 41), ('관련', 41), ('등록', 41), ('시설', 40), ('승인', 40), ('취소', 40), ('지원', 38), ('수급인', 38), ('내용', 37), ('관계', 36), ('실시', 36), ('자가', 36), ('거짓', 36), ('본문', 36), ('166조', 36), ('석면', 35), ('화학', 35), ('사유', 35), ('건설공사', 34), ('관리자', 32), ('측정', 32), ('부분', 32), ('수입', 31), ('인절', 30), ('기간', 30), ('설치', 29), ('장소', 29), ('건강진단', 28), ('정보', 28), ('안전인증', 28), ('서류', 28), ('종류', 27), ('건강장해', 27), ('장해', 27)]
```

<그림 3-8> 다빈도 키워드 계산과정 소스코드

3.2.5 데이터 시각화 단계

데이터 시각화 단계는 데이터 분석과정의 결과로 다빈도 키워드를 추출한다. 키워드별 빈도정보를 입력하여, 워드클라우드 솔루션으로 활용하여 시각화하여 표현한다. 아래의 <그림 3-9>는 데이터를 시각화하는 소스코드이다.

```
import matplotlib.pyplot as plt
from wordcloud import WordCloud

#워드클라우드 패키지를 이용한 시각화 표현
wordcloud = WordCloud(font_path="C:\Windows\Fonts\HYKANM.TTF", width=1000, height=1000,
                      background_color="white", relative_scaling=0.3, max_words=100).generate_from_frequencies(dict(tags))
wordcloud.words_

#차트 설정
plt.figure(figsize=(12,12))
plt.imshow(wordcloud, interpolation = "bilinear")
plt.axis("off")
plt.show()
```

<그림 3-9> 데이터 시각화 과정 소스코드

산업안전보건법은 ‘안전’, ‘고용’, ‘보건’, ‘노동부’, ‘장관’ 키워드순으로 확인되었으며, 중대재해처벌법은 ‘재해’, ‘중대’, ‘안전’, ‘조치’, ‘사업주’ 키워드순으로 확인되었다. 아래의 <그림 3-9>는 산업안전보건법(좌)과 중대재해처벌법(우)를 다빈도 키워드 100개로 시각화하여 표현하였다.



<그림 3-10> 산업안전보건법(좌)과 중대재해처벌법(우) 키워드 시각화

3.2.6 분석결과

산업안전보건법은 1981년 12월 31일 최초 제정되었으며 2021년 8월 17일 51번째로 개정되었으며, 법령의 본문은 1137라인으로 구성되어 있다. 중대재해처벌법은 2021년 1월 26일 최초 제정되어 2022년 1월 27일 시행되는 법령으로 본문은 99라인으로 구성되어 있다. 법령의 본문량의 차이에 따라 빈도수가 10배 이상 차이나 났다.

산업안전보건법, 중대재해처벌법의 다빈도 키워드 100개를 분석하면 32개의 키워드가 공통적으로 등장하였다. 산업안전보건법은 법의 적용대상이 되는 근로자의 키워드가 등장하였지만 중대재해처벌법에는 등장하지 않았다. 산업안전보건법의 의무영역과 관련된 키워드(건강진단, 건설공사, 물질안전보건자료, 안전인증, 화학물질)가 특징적으로 나타났으며, 중대재해처벌법의 새로운 개념의 키워드(공중교통수단, 공중이용시설, 안전보건관리체계, 중대산업재해, 중대시민재해, 중대재해)가 특징적으로 나타났다. 산업안전보건법은 의무사항을 수행하는 주체가 어떤 역할을 해야 하는지에 관련한 키워드가 많이 나타났으며, 중대재해처벌법은 의무대상과 역할에 대한 설명, 그리고 처벌 관련한 키워드가 많이 나타났다. 산업안전보건법의 목적과 관련한 키워드(‘안전’, ‘보건’, ‘산업’, ‘필요’, ‘사항’, ‘재해’, ‘예방’, ‘산업재해’)가 다빈도로 확인되었으며, 법령과 관련된 키워드(‘고용’, ‘노동부’, ‘장관’, ‘고용노동부장관’, ‘고용노동부령’, ‘대통령령’)를 확인할 수 있다. 법령의 주요 키워드(‘사업주’, ‘근로자’)를 확인할 수 있으며, 의무 사항 관련한 다수 키워드(‘필요’, ‘사항’, ‘작업’, ‘업무’, ‘예방’, ‘조치’, ‘방법’, ‘검사’, ‘진단’, ‘준용’, ‘제출’, ‘지정’, ‘교육’, ‘확인’, ‘규정’,

'평가', '절차', '자율', '등록', '승인', '취소', '실시', '자가', '관리자', '건강진단', '안전인증', '서류'를 확인할 수 있다. 또한 166조(수수료 등) 내는 조건과 관련한 조항으로 별칙조항에서 다수 언급되었다. 화학물질 관련 키워드('물질', '유해', '물질안전보건자료', '화학')를 확인할 수 있다. 아래의 <표 3-1>은 산업안전보건법의 다빈도 주요키워드 100위를 도표로 표현하였다.

<표 3-1> 산업안전보건법 다빈도 키워드 100

단어	빈도수	단어	빈도수	단어	빈도수	단어	빈도수
안전	373	물질	80	사업	53	자가	36
고용	364	사업장	79	인증	51	거짓	36
보건	264	지도	74	수행	51	본문	36
노동부	223	기준	72	제공	49	166조	36
장관	216	자료	71	전문	49	석면	35
고용노동부 장관	215	결과	70	제조	44	화학	35
노동	211	대상	68	공사	44	사유	35
고용노동부령	209	유해	64	절차	44	건설공사	34
필요	183	개정	62	근로	43	관리자	32
사항	181	검사	62	물질안전보건자료	43	측정	32
사업주	172	진단	62	기계	43	부분	32
근로자	163	사용	60	환경	42	수입	31
산업	152	조사	60	자율	42	인정	30
기관	143	준용	60	변경	42	기간	30
재해	124	제출	59	도급	41	설치	29
작업	113	위험	58	설비	41	장소	29
위반	111	지정	57	등록	41	건강진단	28
대통령령	107	발생	56	시설	40	정보	28
업무	105	교육	56	승인	40	안전인증	28
예방	96	건설	56	취소	40	서류	28
산업재해	94	확인	56	지원	38	종류	27
관리	91	명령	55	수급인	38	건강장해	27
조치	89	규정	55	내용	37	장해	27
방법	86	평가	54	관계	36	화학물질	27
건강	83	사람	53	실시	36	개선	26

중대재해처벌법의 목적과 관련키워드('중대', '재해', '안전', '조치', '의무', '이행')와 적용 대상 키워드('사업주', '경영', '책임자', '법인', '경영책임자')를 확인할 수 있으며, 중대재해 구분 키워드('중대', '재해', '중대산업재해', '중대시민재해', '공중이용시설', '공중교통수단')를 확인할 수 있다. 처벌 관련키워드('벌금', '책임', '징역', '10억원', '1년', '제6조', '10조')를 다수 확인하였다. 아래의 <표 3-2>은 중대재해처벌법의 다빈도 키워드 100위를 도표로 표현하였다.

<표 3-2> 중대재해처벌법 다빈도 키워드 100

단어	빈도수	단어	빈도수	단어	빈도수	단어	빈도수
중대	23	중대재해	8	실질적	5	구축	4
재해	23	시민	8	자배·운영·관리	5	재발방지	4
안전	22	이용	7	10조	5	대책	4
조치	21	대통령령	7	감독	4	개선	4
사업주	20	방지	7	안전·보건	4	징역	4
기관	20	사항	7	피해	4	10억원	4
발생	18	사업장	6	생명	4	의무위반행위	4
경영	16	공중	6	신체	4	시행	3
책임자	16	종사자	6	동일	4	법률	3
법인	16	중대시민재해	6	원인	4	고용	3
위반	15	관리상의	6	교육	4	노동부	3
경영책임자	15	지원	6	3조	4	목적	3
이행	15	6조	6	장소	4	원료	3
보건	14	법령	6	도급	4	조물	3
관리	14	행위	6	책임	4	처벌	3
산업	13	공중이용시설	5	업무	4	개월	3
사업	13	공중교통수단	5	중앙	4	치료	3
의무	13	교통	5	지방	4	유해	3
2조	13	수단	5	지방자치단체	4	자가	3
시설	12	규정	5	자치	4	1년	3
중대산업재해	10	사용	5	단체	4	설계	3
벌금	10	제외	5	확보	4	제조	3
예방	9	사람	5	예산	4	결함	3
필요	9	행정	5	안전보건 관리체계	4	규모	3
관계	9	4조	5	체계	4	고려	3

3.3 연구 분석환경

분석프로그램은 다음의 프로그램과 패키지를 사용하였다. 개발환경 도구들(프로젝트, 패키지, 파이썬 등)을 관리하기 위한 그래픽 사용자 인터페이스인 아나콘다 네비게이터와 컴퓨터 언어의 한 종류로 간단한 문법으로 구성되어 있는 파이썬, 대화형 개발 프로그램 노트북 프로그램을 사용하여 연구하였다. 엑셀을 불러오는 패키지(load_workbook)과 한글 단어사전을 분석하는 패키지(Kkma), 단어의 빈도를 추출하는 패키지(collections), 단어의 중요도를 시각화하여 표현하는 패키지(Wordcloud, pyplot, plt)를 사용하였다.

연구분석에 사용한 프로그램 및 패키지 목록

- Anaconda Navigator 1.10.0
- Python 3.9.1
- Notebook 6.3.0
- 파이썬 패키지 : load_workbook, Kkma, collections, wordcloud, pyplot, plt

4. 산업안전 분야의 텍스트 마이닝 분석

4.1 연구대상

산업안전과 관련된 데이터 수집을 위해 한국언론진흥재단의 빅카인즈와 주식회사 바이브컴퍼니의 썸트렌드 플랫폼을 활용하여 데이터를 수집하였다. 언론보도 키워드 분석을 위해 빅카인즈를 활용하고, 감정 키워드 분석을 위해 썸트렌드를 활용하였다. 빅카인즈에 제공한 54개 언론사의 언론보도 내용을 대상으로 산업안전과 관련된 키워드(산업안전, 산업재해, 산재예방)의 키워드를 포함하고 카테고리 분류상 사건사고(사고, 재해)로 분류된 기사의 제목 또는 본문을 대상으로 선정하였다. 감정 키워드 분석을 위해 썸트렌드에서 동일한 키워드로 검색하고 4종의 소셜 네트워크 서비스(블로그, 트위터, 뉴스, 인스타그램)을 대상으로 분석하였다.

연구의 분석기간 2016년부터 2021년 6월까지로 설정하였으며, 2016년부터 2020년까지는 1년 단위로 분석하고 2021년은 1월부터 6월까지를 분석기간으로 하였다. 언론보도 데이터의 연구분석은 빅카인즈에서 제공하는 데이터를 수집하고 정제와 분석 과정을 거쳐 시각화의 과정으로 진행하였으며, 감정어 데이터 분석은 썸트렌드에서 제공하는 분석결과에 추가적으로 데이터 정제와 분석과정을 수행하고 시각화순으로 진행하였다.

산업안전 분야의 텍스트 마이닝 분석 조건은 아래의 <표 4-1>와 같다.

<표 4-1> 산업안전 분야 텍스트 마이닝 분석조건

구분	언론보도 분석 (주요키워드 분석)	소셜 네트워크 서비스 (감정어 분석)
검색어	산업안전 키워드(산업안전, 산업재해, 산재예방)	
기간	2016년 ~ 2021년 6월	
분석대상	언론사 54개의 사건사고(사고,재해) 카테고리	커뮤니티, 블로그, 트위터, 뉴스, 인스타그램
데이터수집	빅카인즈 (https://www.bigkinds.or.kr/)	썸트렌드 (https://www.some.co.kr)

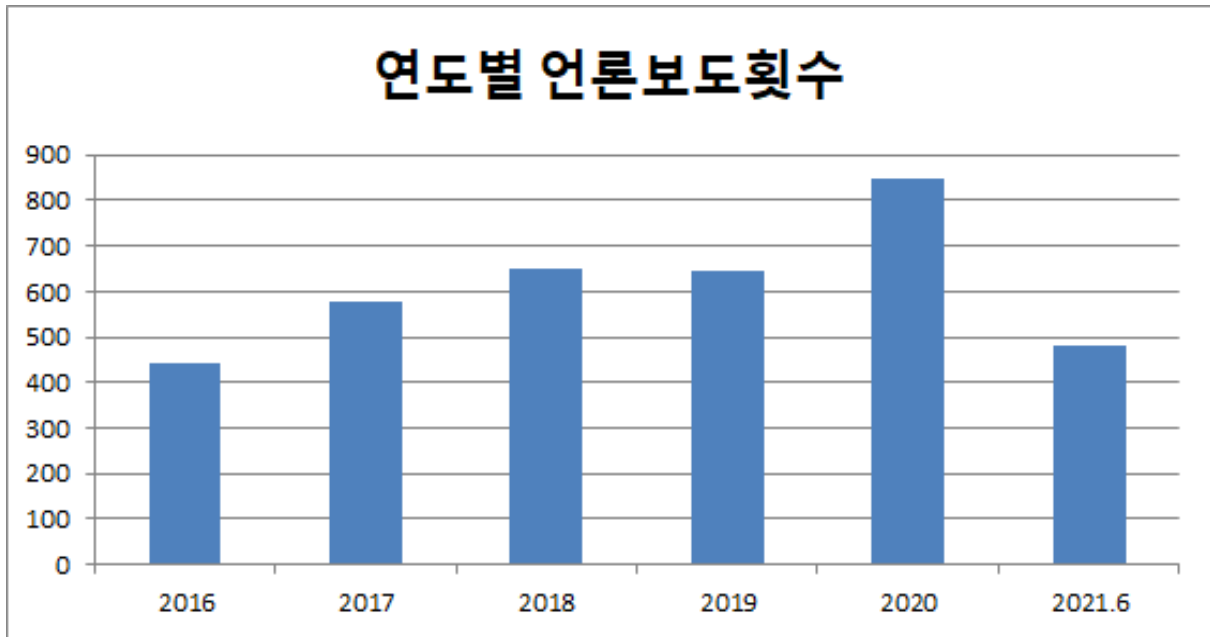
4.2 산업안전 분야 언론보도 분석

빅카인즈에 등록된 54개 언론사(중앙지 : 11, 경제지 : 8, 지역종합지 : 28, 방송사 : 5, 전문지 : 2)의 언론보도 기사 수집자료를 대상으로 연도별 언론보도 횟수 변화추이는 다음과 같다. 2016년부터 2021년 6월까지의 기간동안 검색한 결과 4,020건이며, 중복데이터와 산업안전과 관련이 낮은 데이터를 제외한 3,647건을 분석대상으로 하였다. 2016년부터 2021년까지 언론보도 기사의 수가 전반적으로 증가하고 있다. 2016년부터 2018년까지 매년 언론보도 횟수가 증가하였으며, 2019년 소폭으로 감소하다 하였으며 2020년부터 다시 언론보도 횟수가 증가하였다. 아래의 <표 4-2>는 산업안전 관련 키워드 언론보도 횟수를 필터링(중복제거, 관련성 낮은 데이터 제거) 작업의 전과 후를 나누어 표현하였다.

<표 4-2> 산업안전 언론보도 횟수

연도	필터링 전		필터링 후	
	언론보도 횟수	비중	언론보도 횟수	비중
2016.1 ~ 2016.12	478	11.89	443	12.15
2017.1 ~ 2017.12	612	15.22	579	15.88
2018.1 ~ 2018.12	689	17.14	649	17.80
2019.1 ~ 2019.12	682	16.97	647	17.74
2020.1 ~ 2020.12	896	22.29	847	23.22
2021.1 ~ 2021.6	663	16.49	482	13.22
합계	4,020	100	3,647	100

언론보도 횟수를 살펴보면, 2016년 443회(12.15%), 2017년 579회(15.88%), 2018년 649회(17.80%), 2019년 647회(17.74%), 2020년 847회(23.22%), 2021년 6월까지 482회(13.22%)로 산업안전에 대한 언론의 관심도가 증가하고 있음을 알 수 있다. 아래의 <그림 4-1>은 연도별 언론보도횟수를 시각화하여 표현한 것이다.



<그림 4-1> 산업안전 언론보도 횟수의 연도별 추이

54개 언론사 중 '머니투데이'가 220회(6.03%)으로 가장 많은 기사를 보도하였으며, 경향신문 206회(5.65%), 아시아경제 180회(4.94%), 한국경제 148회(4.06%), 한겨레 140회(3.84%) 등의 순으로 나타났다.

지역종합지 28개 언론사가 1149회(31.51%), 중앙지 11개 언론사가 1106회(30.33%), 경제지 8개 언론사가 1077회(29.53%), 방송사 5개 언론사가 239회(6.55%), 전문지 2개 언론사가 76회(2.08%) 순으로 나타났다. 2016년부터 2021년 6월까지 언론사별 기사수와 비율은 <표 4-3>과 같다.

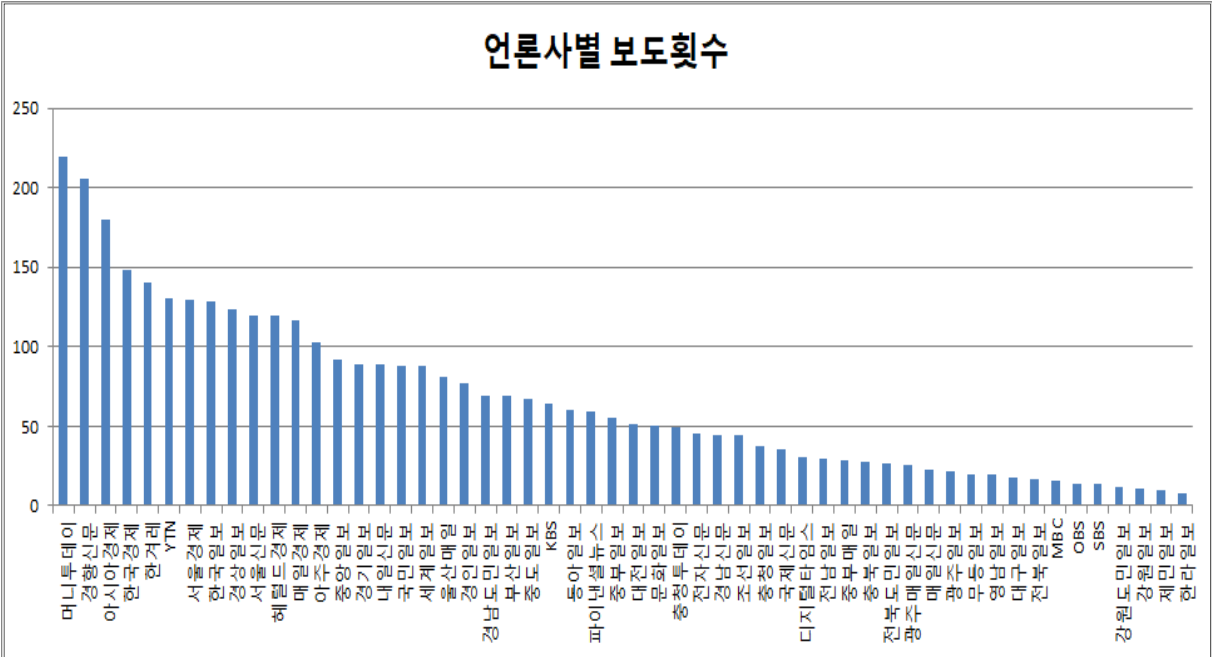
<표 4-3> 2016년 ~ 2021년 6월 언론사별 기사수와 비율

언론사	연도별 기사수		년도					
	연도	비율	2016	2017	2018	2019	2020	2021.6
합계	3647	100%	443	579	649	647	847	482
KBS	64	1.75	1	3	8	4	30	18
MBC	16	0.44	3		3	8		2
OBS	14	0.38		5	1	1	3	4
SBS	14	0.38			2	4	6	2
YTN	131	3.59	12	17	17	28	41	16
강원도민일보	12	0.33	5	1	1	4	1	

강원일보	11	0.30	3	2	3	1	1	1
경기일보	89	2.44	9	18	19	15	20	8
경남도민일보	69	1.89	3	22	13	10	15	6
경남신문	44	1.21	4	16	12	4	2	6
경상일보	124	3.40	33	27	23	21	16	4
경인일보	77	2.11	4	12	17	8	31	5
경향신문	206	5.65	22	36	29	36	60	23
광주매일신문	26	0.71	5	4	2	4	6	5
광주일보	22	0.60		1	1	3	5	12
국민일보	88	2.41	11	16	8	19	18	16
국제신문	36	0.99	2	12	4	9	7	2
내일신문	89	2.44	12	15	13	11	25	13
대구일보	18	0.49	3	2	4	1	5	3
대전일보	51	1.40	2	4	19	16	6	4
동아일보	60	1.65	6	5	10	10	19	10
디지털타임스	31	0.85	2	4	6	5	9	5
매일경제	117	3.21	24	15	22	18	22	16
매일신문	23	0.63	1	1	3	5	8	5
머니투데이	220	6.03	38	59	39	28	31	25
무등일보	20	0.55	5	1	3	3	6	2
문화일보	50	1.37	10	8	9	8	13	2
부산일보	69	1.89	2	12	14	14	14	13
서울경제	130	3.56	11	20	27	29	23	20
서울신문	120	3.29	15	17	15	21	40	12
세계일보	88	2.41	1	19	12	19	25	12
아시아경제	180	4.94	21	20	46	24	49	20
아주경제	103	2.82	11	10	21	25	20	16
영남일보	20	0.55	1	2	6	6	4	1
울산매일	81	2.22	20	17	15	17	3	9
전남일보	30	0.82	1		7	5	5	12
전북도민일보	27	0.74	4	3	9	5	4	2
전북일보	17	0.47	1	3	2	5	3	3
전자신문	45	1.23	4	5	7	15	6	8

제민일보	10	0.27	1	1	2	4	1	1
조선일보	44	1.21			7	9	22	6
중도일보	67	1.84	7	9	11	24	10	6
중부매일	29	0.80	4	4	8	9	4	
중부일보	55	1.51	8	6	13	5	15	8
중앙일보	92	2.52	10	13	11	11	26	21
충북일보	28	0.77	5	7	3	4	7	2
충청일보	37	1.01	7	5	5	8	8	4
충청투데이	49	1.34	9	5	6	14	14	1
파이낸셜뉴스	59	1.62	16	12	9	10	5	7
한겨레	140	3.84	18	26	21	20	37	18
한국경제	148	4.06	15	19	35	21	33	25
한국일보	129	3.54	11	19	27	23	33	16
한라일보	8	0.22	1		3	1	1	2
헤럴드경제	120	3.29	19	19	16	15	29	22

아래의 <그림 4-2>는 2016년부터 2021년 6월까지의 언론사별 보도횟수를 시각화하여
으로 표현하였다.



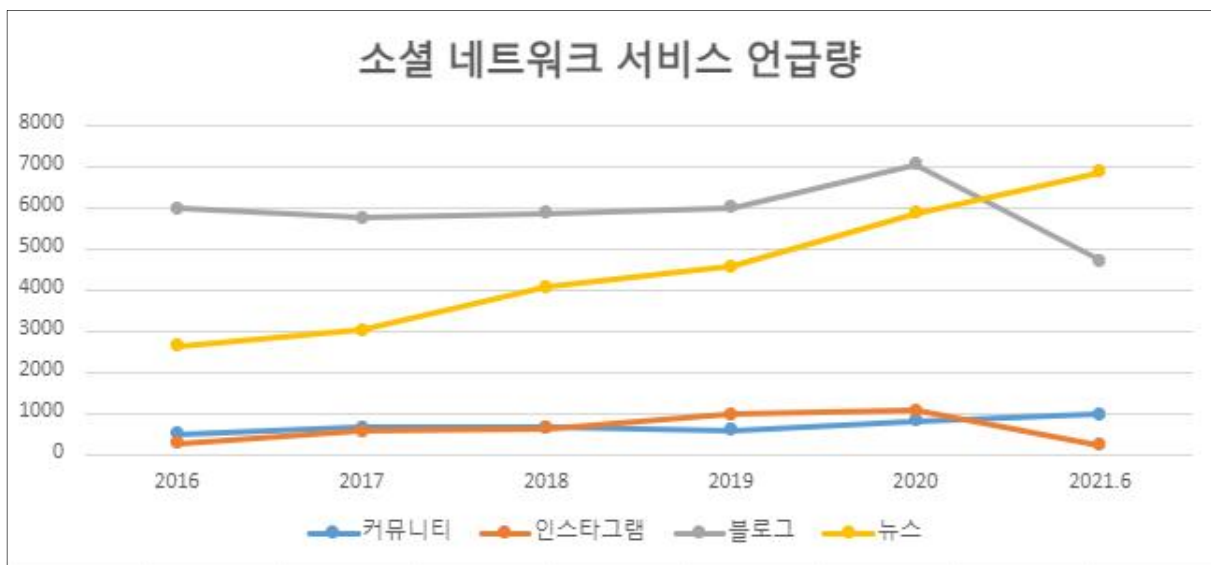
<그림 4-2> 2016년 ~ 2021년 6월 언론사별 보도횟수

4.3 산업안전 분야 소셜 네트워크 서비스 언급량 분석

감정 키워드 분석을 위해 분석플랫폼 썬트렌드를 사용하였으며, 소셜 네트워크 서비스(커뮤니티, 인스타그램, 블로그, 뉴스, 트위터)를 대상으로 산업안전 관련 키워드(산업안전, 산업재해, 산재예방)를 분석하였다. 소셜 네트워크 서비스 전체 언급량은 매년 증가하는 추세이며, 블로그와 뉴스가 매년 증가하였으며, 특히 뉴스에서 2021년의 언급량이 급격하게 증가하였다. 인스타그램은 2020년까지 증가하다 2021년 감소하였다. 트위터는 2018년까지 언급량이 감소하다 2019년부터 증가하였다. 아래의 <표 4-4>와 <그림 4-3>은 2016년부터 2021년 6월까지 소셜 네트워크 서비스의 언급량을 도표와 그래프로 시각화하여 표현한 것이다.

<표 4-4> 2016년 ~ 2021년 6월 소셜 네트워크 서비스 언급량

년도	커뮤니티	인스타그램	블로그	뉴스	트위터	합계
2016	494	268	5,973	2,638	2,620	11,993
2017	657	561	5,736	3,009	2,089	12,052
2018	658	634	5,851	4,053	1,731	12,927
2019	591	958	5,990	4,561	2,085	14,185
2020	811	1,054	7,026	5,859	2,434	17,184
2021.6	964	221	4,691	6,863	2,432	15,171

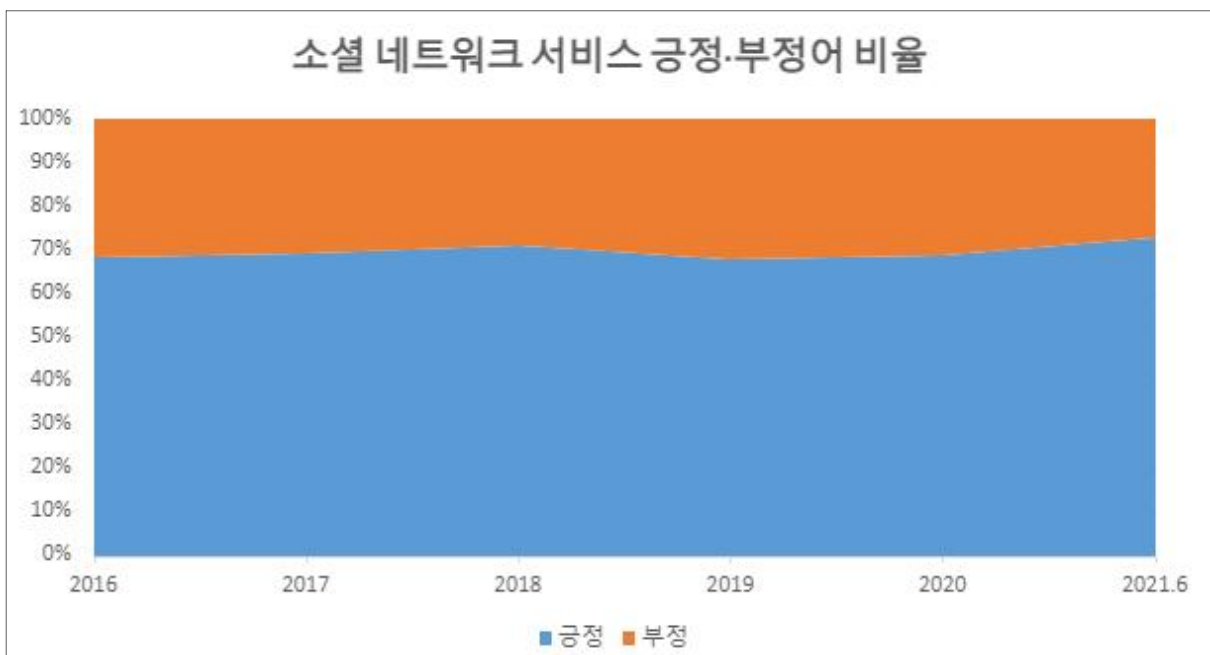


<그림 4-3> 소셜네트워크 서비스 언급량 추이

연도별 감정키워드를 추출하고 후처리 작업으로 산업안전과 관련이 없거나 감정표현과 관련없는 키워드를 제거하였다. 연도별 다빈도 키워드 100개의 감정(긍정, 부정)과 키워드 빈도수를 합하여 계산하였다. 매년 부정어의 수가 많았으며, 긍정 키워드는 41~46개로 나타났으며, 부정 키워드는 54~59개로 나타났다. 연도별 긍정어의 비율은 63~68%로 나타났으며, 부정 키워드의 비율은 32~37%로 나타났다. 매년 다빈도 순위 1위 키워드(안전), 2위 키워드(보상)의 영향으로 긍정어의 비율이 더 높게 나타났다. 아래의 <표 4-5>와 <그림 4-4>은 2016년부터 2021년 6월까지 소셜 네트워크 서비스의 감정 키워드의 수와 빈도의 비율을 도표와 그래프로 시각화하여 표현한 것이다.

<표 4-5> 소셜 네트워크 서비스 감정키워드의 빈도와 비율

감정어		2016	2017	2018	2019	2020	2021.6
긍정	키워드수	42	46	41	46	41	41
	빈도	5,753	6,326	7,231	8,521	11,262	9,254
	비율	64.16	66.14	62.95	68.19	65.74	63.90
부정	키워드수	58	54	59	54	59	59
	빈도	3,214	3,238	4,256	3,975	5,869	5,228
	비율	35.84	33.86	37.05	31.81	34.26	36.10
-	빈도합계	8,967	9,564	11,487	12,496	17,131	14,482



<그림 4-4> 소셜 네트워크 서비스의 긍정·부정어 비율

4.4 산업안전 분야 텍스트마이닝 시계열 분석

주요 키워드 빈도분석은 꼬꼬마사전(kkma)에 정의된 단어를 기반으로 텍스트를 분석하는 것이다. 언론보도 기사는 빅카인즈 플랫폼의 54개 언론사의 언론보도의 사고사고 카테고리들을 대상으로 분석하였고, 소셜 네트워크 서비스 데이터는 썸카인즈 플랫폼을 활용하여 분석하였다. 산업안전 관련 키워드(산업안전 또는 산업재해 또는 산재예방)을 포함한 데이터를 분석 대상으로 하였으며, 중복데이터와 관련성이 낮은 데이터와 불필요한 데이터를 제거하고 분석하였다. 분석기간(2016년 ~ 2021년 6월) 중 2016년부터 2020년까지는 1년 단위로 분석하고, 2021년은 6개월을 분석 기간으로 하였다. 연도별 키워드 빈도분석과 감정 키워드의 분석을 하고, 분석된 결과를 파이썬 프로그램(워드클라우드)로 시각화하였다.

4.4.1 2016년 시계열 분석

언론보도의 특징 키워드로 2016년 5월 28일 서울 지하철 2호선 구의역 승강장에서 스크린 도어를 정비하던 직원의 산업재해와 관련된 장소 키워드(서울, 경기도, 지하철, 구의역)와 사고발생의 원인 관련 키워드(발생, 사고, 위험, 관리, 원청, 하청, 설치, 재해, 현장, 시설, 조사)를 확인할 수 있으며, 남양주 지하철 공사장 폭발사고 관련 키워드(사고, 발생, 재해, 현장, 폭발, 가스, 물질, 책임, 점검, 원인, 하청, 관계자, 안전사고, 공사장, 사업주), 2016년 지진관련 경주 월성원전 방사능 누출 키워드(산업, 안전, 사고, 예방, 관리, 작업, 산업재해, 위험, 누출, 매뉴얼, 피해)이 있다. 아래의 <표 4-6>은 2016년 언론보도 주요키워드를 표현한 것이다.

<표 4-6> 2016년 언론보도 주요키워드

단어	빈도	단어	빈도	단어	빈도	단어	빈도
산업	434	책임	150	정책	114	개정	92
안전	408	안전사고	146	전국	112	노동	92
사고	395	붕괴	139	원청	111	확인	92
발생	336	대상	138	기업	111	가능	91
재해	334	사업장	137	설치	110	시작	91
현장	310	조사	137	공단	109	직원	91
예방	243	관계자	133	사망자	108	집중	91
관리	242	점검	133	서울	108	피해	91

작업	241	결과	131	상황	106	활동	91
위험	235	계획	130	감독	105	수준	88
고용	228	대책	130	업무	105	사건	88
근로자	223	개선	129	국민	105	강조	87
업체	213	원인	129	시설	104	예정	87
폭발	210	한국	128	환경	103	사례	87
공사	206	교육	127	기준	101	사업주	86
보건	202	하청	122	처벌	100	대응	86
사망	199	사회	122	방지	100	사용	85
건설	184	추진	122	진행	98	국가	85
강화	180	기관	121	울산	97	평가	85
산업재해	179	보건법	120	공장	97	운영	84
화재	179	가스	120	지하철	96	제도	83
노동부	176	화학	116	경기도	96	노동자	83
산재	165	지적	116	협력	95	문화	83
정부	159	고용노동부	115	물질	94	정보	82
조치	154	마련	114	확대	93	대비	82

2016년 소셜 네트워크 서비스의 특징적인 키워드로 경제협력개발기구(OECD) 국가중 산업재해 1위라는 언론보도를 접하고, 다양한 소셜 네트워크 서비스에서 부정적인 반응들이 나왔다. 관련된 키워드(산재1위, 위험, 스트레스, 불안 등)가 나타났으며, 세월호 사고 2년 경과 후 구조업무 중 사망 또는 장애를 입은 민간잠수부에 대한 보상지연과 관련 키워드(보상, 위험, 기여하다, 최선다하다, 보상받지 못하다, 고통받다, 어려운상황)로 나타났다. 지하철 2호선 구의역 승강장에서 안전문 정비 중 발생한 사고와 반응과 관련한 키워드(위험, 참사, 사고, 사고당하다, 명복빌다)를 확인할 수 있다. 아래의 <표 4-7>은 2016년 소셜 네트워크 서비스 주요키워드를 표현한 것이다.

<표 4-7> 2016년 소셜 네트워크 서비스 주요키워드

단어	빈도	감정	단어	빈도	감정	단어	빈도	감정
안전	2183	긍정	기여하다	51	긍정	참사	25	부정
보상	1379	긍정	부담	48	부정	최악	24	부정
산재1위	425	긍정	위반하다	46	부정	체계적	24	긍정
위험	357	부정	최선다하다	45	긍정	임금체납	24	부정

강화하다	196	긍정	후유증	45	부정	고통받다	24	부정
스트레스	149	부정	추락하다	42	부정	충격	23	부정
불안	145	부정	희망	41	긍정	어려운상황	23	부정
피해	143	부정	인정받지 못하다	41	부정	위기	22	부정
노력하다	141	긍정	도움되다	39	긍정	꿈꾸다	22	긍정
가능하다	138	긍정	어렵다	38	부정	조치취하다	22	긍정
인정받다	137	긍정	경제적	37	긍정	도움받다	22	긍정
불법	135	부정	어려움겪다	36	부정	우려	21	부정
효율적	123	긍정	손실	36	부정	폭행	21	부정
보상하다	115	긍정	비극	34	부정	보상받지 못하다	21	부정
보상받다	108	긍정	비판	33	부정	높은수준	21	긍정
과실	101	부정	문제	32	부정	서비스 제공하다	21	긍정
적극적	87	긍정	피해입다	32	부정	비용절감	21	긍정
손해	79	부정	아픔	31	부정	문제있다	20	부정
인정하지 않다	78	부정	부상당하다	31	부정	정상적	20	긍정
장애	77	부정	차별화되다	31	긍정	행복	20	긍정
관리하다	76	긍정	승인받다	30	긍정	호소하다	19	부정
바라다	74	긍정	논란	30	부정	독성물질	18	부정
해결하다	73	긍정	위험한일	29	부정	개선되다	18	긍정
안전하다	69	긍정	의혹	28	부정	반발	18	부정
최선	69	긍정	금품	28	부정	고의적	18	부정
고통	63	부정	개탄	28	부정	유용한정보	18	긍정
폭발사고	63	부정	거부하다	27	부정	허위	17	부정
혐의	61	부정	힘들다	27	부정	열악하다	17	부정
사고당하다	59	부정	일방적	26	부정	인증받다	17	긍정
범죄	57	부정	추락	26	부정	치명적	17	긍정
개선하다	56	긍정	신속하다	26	긍정	기대하다	17	긍정
도움	55	긍정	취약하다	25	부정	위반	17	부정
합리적	52	긍정	최적	25	긍정	건강하다	17	긍정
차별	51	부정						

아래의 <그림 4-5>는 언론보도 기사(좌)와 소셜 네트워크 서비스(우)의 주요 키워드를 시각화하여 비교한 것이다.



<그림 4-5> 2016년 주요키워드 시각화

4.4.2 2017년 시계열 분석

특징 키워드로 2017년 5월 1일 삼성중공업 타워크레인 사고의 업종관련 키워드(크레인, 타워크레인, 타워, 조선, 삼성, 해양, 조선해양, 경남, 창원)와 사고의 기인물, 원인 관련 키워드(크레인, 타워크레인, 붕괴, 추락, 관계자, 원청, 하청, 원인, 설치, 피해, 감독, 대책, 산업재해)와 2017년 남양주 타워크레인 전도 사고와 관련된 키워드(산업, 안전, 사고, 발생, 현장, 안전사고, 추락, 재해)와 삼성전자 반도체 공장에서 근무했던 황유미씨의 산업재해 승인 요청과 관련한 키워드(작업, 위험, 조사, 고용노동부, 원인, 책임, 사망자, 화학)로 나타났다. 아래의 <표 4-8>은 2017년 언론보도 주요키워드를 표현한 것이다.

<표 4-8> 2017년 언론보도 주요키워드

단어	빈도수	단어	빈도수	단어	빈도수	단어	빈도수
산업	562	건설	201	협력	145	기업	121
안전	538	감독	195	기준	145	처리	120
사고	506	대상	195	조선	142	사실	119
발생	451	사업장	193	해양	142	전문가	117

현장	404	강화	191	업무	140	전국	116
재해	391	대책	189	환경	140	참여	116
작업	367	계획	185	한국	140	처벌	115
예방	350	교육	181	상황	139	창원	114
위험	335	원인	179	사항	139	지원	114
고용	334	확인	178	시설	136	중지	114
관리	323	붕괴	177	마련	135	여부	114
보건	315	점검	169	지적	135	피해	112
노동부	294	산업재해	169	원청	134	가능	112
업체	278	공단	165	사망자	133	사업	111
근로자	270	위반	163	예정	132	확대	111
화재	265	관계자	163	화학	131	가스	110
사망	250	크레인	162	기관	131	구조	110
폭발	234	하청	161	방지	130	조선해양	109
조치	230	설치	158	사회	127	국민	109
조사	226	노동자	157	노동	126	방침	107
산재	219	보건법	157	특별	125	부과	107
고용노동부	211	진행	157	정책	123	집중	106
결과	206	개선	155	운영	123	대형	105
공사	206	사용	154	추락	122	물질	104
정부	201	책임	149	안전사고	121	강조	104

2017년의 특징적인 키워드로 삼성전자 반도체 공장에서 근무하다 백혈병에 걸려 사망한 황유미씨 산업재해 인정과 관련한 키워드(아프다, 산업재해, 보상받지 못하다, 발암물질)로 나타났으며, 서울신문에서 연재한 특별기획 '2017 대한민국 과로 리포트' 와 관련하여 2018년부터 과로로 발생한 산재에 대해 인정받는 범위의 확대와 관련한 키워드(스트레스, 인정받다, 보상받다, 인정받지 않다, 보상받지 못하다)로 나타났다. 아래의 <표 4-9>은 2017년 소셜 네트워크 서비스 주요키워드를 표현한 것이다.

<표 4-9> 2017년 소셜 네트워크 서비스 주요키워드

단어	빈도	감정	단어	빈도	감정	단어	빈도	감정
안전	2330	긍정	폭행	59	부정	위험하다	29	부정
보상	1380	긍정	개선하다	58	긍정	충격	29	부정

위험	325	부정	금품	55	부정	위반	28	부정
강화하다	249	긍정	획기적	52	긍정	추락	27	부정
스트레스	180	부정	고통	52	부정	편하다	27	긍정
불법	170	부정	부담	52	부정	불량	25	부정
인정받다	167	긍정	체계적	49	긍정	어려움겪다	25	부정
피해	157	부정	우려	47	부정	힘들다	25	부정
최선	156	긍정	위반하다	45	부정	처우개선	24	긍정
가능하다	152	긍정	승인받다	45	긍정	경제적	24	긍정
보상받다	147	긍정	악화되다	44	부정	행복	24	긍정
범죄	136	부정	안전하다	44	긍정	안정적	24	긍정
과실	126	부정	인명피해	42	부정	안타깝다	23	부정
보상하다	119	긍정	최선다하다	41	긍정	조치취하다	23	긍정
산재1위	117	긍정	불복하다	39	부정	높은수준	23	긍정
사고당하다	109	부정	인정받지 못하다	39	부정	좋다	23	긍정
기여하다	102	긍정	문제	38	부정	쾌적한환경	22	긍정
차별	100	부정	후유증	38	부정	혜택받다	22	긍정
해결하다	95	긍정	고통받다	38	부정	기대하다	22	긍정
도움	92	긍정	비정상	38	부정	기부	22	긍정
폭발사고	87	부정	최악	37	부정	꼼수	22	부정
노력하다	86	긍정	부상당하다	37	부정	비판	21	부정
적극적	85	긍정	합리적	36	긍정	거부하다	21	부정
불안	85	부정	피해입다	36	부정	건강지키다	21	긍정
혐의	80	부정	발암물질	36	부정	스트레스받다	21	부정
관리하다	78	긍정	많은관심	35	긍정	효율적	21	긍정
인정하지않다	77	부정	압력	34	부정	처벌받다	21	부정
손해	74	부정	보상받지 못하다	34	부정	완화하다	21	긍정
바라다	73	긍정	논란	33	부정	개선되다	20	긍정
신속하다	67	긍정	고민	31	부정	다양하다	20	긍정
아프다	63	부정	괴롭히다	31	부정	의혹	20	부정
정상적	63	긍정	효과적	31	긍정	웃음	19	긍정
장애	62	부정	희망	30	긍정	추락하다	18	부정
서비스 제공하다	62	긍정						

아래의 <그림 4-6>는 언론보도 기사(좌)와 소셜 네트워크 서비스(우)의 주요 키워드를 시각화하여 비교한 것이다.



<그림 4-6> 2017년 주요키워드 시각화

4.4.3 2018년 시계열 분석

특징 키워드로 문재인 정부의 국민생명 지키기 3대 프로젝트 관련 키워드(한국, 대통령, 정부, 교통, 산업안전, 산재, 자살, 국가, 사망자, 생명, 프로젝트, 대한민국, 정책, 경제) 등이 있다. 사망자 절반 줄이기와 관련된 프로젝트 관련 키워드의 특징키워드로 확인할 수 있다. 산업안전보건법 30년만의 전면개정과 관련한 키워드(산업, 안전, 사고, 발생, 재해, 예방, 위험, 원청, 하청, 노동자, 사업장, 관계자, 방지, 마련)가 나타났고, 2018년 8월 21일 인천 남동산단에서 발생한 세일전자 화재 사고 관련 키워드(사고, 재해, 현장, 화재, 위험, 사망, 시설, 산업재해, 사망자)로 나타났다. 아래의 <표 4-8>은 2018년 언론보도 주요키워드를 표현한 것이다.

<표 4-10> 2018년 언론보도 주요키워드

단어	빈도수	단어	빈도수	단어	빈도수	단어	빈도수
산업	633	한국	217	추진	173	재난	148
안전	591	조사	214	원인	172	생명	147
사고	534	사회	213	가능	172	지적	147

발생	466	폭발	213	피해	171	시간	146
재해	423	교육	210	정책	171	요구	146
현장	419	노동부	210	업무	170	국회	146
관리	402	업체	206	대응	169	경제	143
예방	378	공사	205	진행	167	강조	143
화재	360	시설	204	확인	165	대비	143
위험	332	관계자	200	방지	163	화학	141
고용	322	기업	199	마련	163	고용노동부	140
정부	304	사업장	197	제도	161	사망자	138
작업	278	산재	196	구조	161	체계	138
사망	261	대상	195	붕괴	161	노력	136
점검	252	노동자	190	공장	161	협력	135
결과	247	건설	189	위반	160	처리	134
계획	246	기관	189	지방	160	대한	134
보건	244	환경	183	국가	155	수준	133
조치	238	운영	181	사업	155	집중	132
강화	238	설치	180	사실	153	활동	130
대책	231	산업재해	179	경기도	153	나라	128
개선	230	근로자	178	위원회	152	전문가	128
국민	228	보건법	178	사람	151	특별	128
책임	225	노동	176	시작	150	공단	127
상황	223	지원	173	확대	148	교통	126

2018년의 특징적인 키워드로 태안 화력발전소에서 근무하던 하청업체 근무하던 김용균 씨의 죽음으로 정규직과 비정규직의 차별을 없애고, 비정규직의 정규직화를 추진하던 정부정책과 관련된 키워드(안전, 위험, 우려, 강화하다, 희망, 노력하다, 간사, 피해, 문제, 비용절감, 아픔, 위반, 의혹, 위험한일)로 나타났다. 국민생명지키기 3대 프로젝트와 위험의 외주화를 없애기 위한 산업안전보건법의 전부 개정안과 관련한 키워드(안전, 위험, 획기적, 문제, 인명피해, 음주운전)가 나타났다. 아래의 <표 4-11>은 2018년 소셜 네트워크 서비스 주요키워드를 표현한 것이다.

<표 4-11> 2018년 소셜 네트워크 서비스 주요키워드

단어	빈도	감정	단어	빈도	감정	단어	빈도	감정
안전	2687	긍정	불안	68	부정	비용절감	36	긍정
보상	1404	긍정	최선다하다	66	긍정	힘들다	35	부정
위험	815	부정	체계적	65	긍정	부상당하다	35	부정
강화하다	392	긍정	보상하다	61	긍정	안심	34	긍정
인정받다	326	긍정	혼란	61	부정	비판	34	부정
피해	263	부정	괴롭히다	58	부정	사고발생	33	부정
우려	223	부정	고통	57	부정	취약하다	33	부정
최선	214	긍정	아픔	57	부정	위기	32	부정
부담	155	부정	폭행	55	부정	경제적	32	긍정
과실	147	부정	충격	51	부정	후유증	32	부정
보상받다	136	긍정	효과적	50	긍정	최적	32	긍정
적극적	125	긍정	합리적	50	긍정	살인기업	32	부정
개선하다	121	긍정	희망	50	긍정	거부하다	32	부정
불법	116	부정	긍정적	50	긍정	걱정	32	부정
바라다	115	긍정	진심	50	긍정	다행	31	긍정
스트레스	112	부정	높은수준	48	긍정	심각한문제	31	부정
장애	107	부정	위반하다	47	부정	부작용	31	부정
범죄	106	부정	보상받지 못하다	47	부정	기여하다	31	긍정
논란	104	부정	인명피해	44	부정	위반	30	부정
획기적	99	긍정	갈등	44	부정	효율적	29	긍정
가능하다	98	긍정	기부	43	긍정	의혹	29	부정
차별	93	부정	문제	43	부정	손실	29	부정
손해	91	부정	최악	43	부정	정상적	29	긍정
노력하다	90	긍정	만족	43	긍정	공식사과	29	부정
해결하다	83	긍정	음주운전	41	부정	간사	29	부정
관리하다	80	긍정	안타깝다	41	부정	과학적	28	긍정
안전하다	79	긍정	불가능	40	부정	피해 입다	28	부정
불량	75	부정	인정받지 못하다	39	부정	도움되지 않는다	28	부정
어렵다	74	부정	높은수준	38	부정	인정하지 않는다	27	부정
도움	73	긍정	기대하다	37	긍정	일방적	27	부정

승인받다	73	긍정	불가피하다	37	부정	위험한일	27	부정
안정적	69	긍정	도움받다	36	긍정	폭발사고	26	부정
반발하다	69	부정	도움되다	36	긍정	잘못되다	26	부정
사고 당하다	68	부정						

아래의 <그림 4-7>는 언론보도 기사(좌)와 소셜 네트워크 서비스(우)의 주요 키워드를 시각화하여 비교한 것이다.



<그림 4-7> 2018년 주요키워드 시각화

4.4.4 2019년 시계열 분석

특징 키워드로 2018년 12월 11일 발생한 한국서부발전의 태안화력발전소에서 발생한 김용균씨 사망사고와 관련한 후속조치로 산업안전보건법 개정과 인권위의 노동부 제도개선과 관련 키워드(태안, 발전소, 김용균, 화력, 먼지, 위험, 도급, 하청, 강화, 노동자, 산업재해, 보건법, 노동부, 고용노동부), 로켓추진체 연료 생산하는 대전 한화공장 폭발사고 관련 키워드(대전, 취급, 제조, 폭발, 재해, 현장, 위험, 작업, 조치, 위반, 방지, 질식, 노출, 사고, 안전사고)가 나타났다. 아래의 <표 4-12>은 2019년 언론보도 주요키워드를 표현한 것이다.

<표 4-12> 2019년 언론보도 주요키워드

단어	빈도수	단어	빈도수	단어	빈도수	단어	빈도수
산업	637	산재	211	기관	162	업무	141
안전	590	노동부	210	개정	160	집중	137
사고	524	결과	209	책임	159	지원	136
발생	443	한국	208	위반	159	국가	134
재해	414	보건법	207	설비	159	사항	131
현장	387	계획	205	지적	158	특별	130
위험	363	공장	205	확대	158	회사	129
관리	356	개선	204	추락	157	사망자	128
작업	351	대책	202	방지	156	합동	128
화재	332	시설	202	교육	156	대응	127
예방	314	관계자	201	기준	155	안전사고	125
폭발	299	업체	199	사업	154	대전	123
고용	299	진행	193	운영	153	사람	123
조치	275	확인	193	사용	152	요구	122
보건	272	설치	188	시행	151	피해	120
사망	270	건설	186	가능	151	하청	119
정부	247	마련	185	전국	150	서울	119
사업장	243	추진	184	물질	147	화학	119
강화	231	환경	181	예정	146	제도	117
대상	226	공사	179	사회	146	지방	116
근로자	225	감독	173	공단	145	노력	116
노동자	219	원인	170	중지	143	노동	116
조사	216	붕괴	169	고용노동부	143	활동	116
산업재해	214	상황	169	직원	142	센터	115
점검	213	기업	165	대표	142	대비	115

2019년의 특징적인 키워드로 2018년 발생한 김용균씨 사망사건의 개선 조치로 위험의 외주화 금지 관련키워드(안전, 위험, 개선하다, 보상, 인정받다, 차별, 불법)가 나타났으며, 사회 부조리 문제를 해결해나가는 수목드라마 '닥터탐정'의 관련키워드(안전, 아프다, 해결하다, 활약)가 나타났다. 아래의 <표 4-13>은 2019년 소셜 네트워크 서비스 주요키워드를 표현한 것이다.

<표 4-13> 2019년 소셜 네트워크 서비스 주요키워드

단어	빈도	감정	단어	빈도	감정	단어	빈도	감정
안전	3629	긍정	획기적	66	긍정	효과적	39	긍정
보상	1293	긍정	관리하다	65	긍정	위험한 일	38	부정
위험	844	부정	장애	63	부정	인위적	38	부정
강화하다	262	긍정	의혹	63	부정	보상하다	37	긍정
인정받다	246	긍정	다양한유형	61	긍정	쉽다	37	긍정
피해	240	부정	기대	59	긍정	현명한선택	36	긍정
스트레스	235	부정	충격	58	부정	논란	36	부정
최선	220	긍정	사고발생	57	부정	경제적	35	긍정
해결하다	187	긍정	바라다	56	긍정	억울하다	35	부정
가능하다	178	긍정	강화되다	55	긍정	정상적	34	긍정
도움	176	긍정	갑질	52	부정	아픔	34	부정
체계적	168	긍정	아프다	51	부정	공포	32	부정
적극적	140	긍정	폭발사고	51	부정	안심	32	긍정
보상받다	139	긍정	공감	50	긍정	기대하다	32	긍정
추락	137	부정	기여하다	50	긍정	인증하다	31	긍정
활약	136	긍정	도움받다	49	긍정	고통받다	30	부정
불법	132	부정	분노	48	부정	악화되다	29	부정
과실	125	부정	부담	47	부정	좋다	29	긍정
희망	111	긍정	최선다하다	46	긍정	독성물질	29	부정
개선하다	108	긍정	호소하다	45	부정	어려운상황	29	부정
우려	99	부정	피해 입다	44	부정	불편	29	부정
노력하다	98	긍정	성공하다	44	긍정	보상받지 못하다	28	부정
손해	90	부정	무단복제	43	부정	불안	28	부정
괴롭히다	89	부정	폭행	43	부정	인정하지 않는다	28	부정
후유증	88	부정	높은수준	43	긍정	비용절감	28	긍정
차별	87	부정	최고	43	긍정	위험한상황	28	부정
다양하다	77	긍정	서비스 제공하다	42	긍정	인명피해	27	부정
고통	76	부정	불량	42	부정	혐의	27	부정
안전하다	74	긍정	많은 기대	42	긍정	어려움겪다	27	부정
승인받다	72	긍정	참사	41	부정	다양한 정보	27	긍정

범죄	70	부정	문제	41	부정	비난	27	부정
어렵다	69	부정	부상당하다	40	부정	손상	26	부정
비극	67	부정	도움되다	39	긍정	비판	26	부정
사고당하다	67	부정						

아래의 <그림 4-8>는 언론보도 기사(좌)와 소셜 네트워크 서비스(우)의 주요 키워드를 시각화하여 비교한 것이다.



<그림 4-8> 2019년 주요키워드 시각화

4.4.5 2020년 시계열 분석

2020년의 특징적인 키워드로 경기도 이천 물류창소 화재사고로 38명의 사상자 발생한 사건의 반응들과 관련된 키워드(물류, 창고, 이천, 제정, 처벌법, 경기도, 민주당, 용접, 벌금, 38명, 시공사, 냉동, 과실, 우레탄 등), 산업재해의 발생을 줄이기 위해 법의 보호대상 확대, 안전보건 확보 의무 강화, 책임 범위 확대와 도급, 용역, 위탁 등의 관계에서 안전보건 확보 의무가 규정된 중대재해처벌법 입법 관련 키워드(제정, 처벌법, 중대재해, 민주당, 정의당, 입법, 통과, 검토, 촉구 등), 코로나19 관련 키워드(코로나, 코로나19, 통과, 긴급, 법안 등)을 확인하였다. 아래의 <표 4-14>은 2020년 언론보도 주요키워드를 표현한 것이다.

<표 4-14> 2020년 언론보도 주요키워드

단어	빈도수	단어	빈도수	단어	빈도수	단어	빈도수
산업	840	폭발	308	마련	243	의원	198
안전	791	노동부	306	참사	243	추진	197
사고	716	처벌	303	사회	241	사람	196
화재	646	노동	302	사업주	241	코로나	195
발생	629	보건법	299	감독	239	교육	194
현장	606	건설	293	관계자	237	처벌법	194
재해	601	상황	285	국회	233	설치	194
관리	476	방지	283	개정	233	지원	192
위험	423	대책	277	결과	231	산업안전 보건법	189
예방	419	산업재해	275	개선	230	경제	186
작업	411	한국	271	환경	230	요구	185
사망	401	근로자	267	공단	229	가능	185
물류	383	이천	267	확인	227	시작	184
노동자	377	대상	266	사망자	224	확보	184
고용	373	원인	265	강조	222	업무	182
기업	367	계획	264	대표	219	위원회	182
강화	364	업체	257	책임자	216	정책	182
참고	360	사업장	257	의무	214	목숨	181
경기	355	피해	255	기관	214	인명	181
보건	353	진행	252	제도	213	경영	180
정부	345	위반	251	센터	212	국민	178
산재	344	지적	248	규정	209	기간	178
공사	343	조사	245	제정	207	서울	176
책임	331	점검	244	시행	205	고용노동부	170
조치	328	기준	244	시설	200	예정	169

2020년의 특징적인 키워드로 경기도 이천 물류창소 화재사고로 38명의 사상자 발생한 사건의 반응들과 관련된 키워드(안전, 참사, 사고당하다, 인명피해, 안타깝다, 사고발생, 위반하다, 관리하다)와 대통령의 11월 17일 국무회의의 산재 사망사고 지적과 관련된 키워드(안전, 위험, 강화하다, 우려, 참사, 체계적, 추락, 안타깝다, 최선 다하다, 높은수준)가

나타났고, 중대재해처벌법 입법과 관련하여 법을 강화한다는 내용의 보도와 실제 사업체에서의 적용과 관련하여 다양한 반응들과 관련한 키워드(안전, 강화하다, 위험, 과실, 우려, 위반하다, 범죄, 적극적, 참사, 비판, 반발, 논란, 효과적, 보상, 부담)를 확인할 수 있다. 코로나19 바이러스의 중대재해처벌법 적용과 관련하여 산업현장에서 코로나 확진자 10명 이상 발생하면 중대재해에 속할 수 있다는 고용노동부의 해석과 관련한 키워드(감염되다, 가능하다, 위험, 감염위험, 우려, 부담, 경제위기, 혼란)가 나타났다. 아래의 <표 4-15>은 2020년 소셜 네트워크 서비스 주요키워드를 표현한 것이다.

<표 4-15> 2020년 소셜 네트워크 서비스 주요키워드

단어	빈도	감정	단어	빈도	감정	단어	빈도	감정
안전	4815	긍정	획기적	89	긍정	손실	49	부정
보상	1668	긍정	보상하다	88	긍정	반하다	48	긍정
위험	816	부정	공감	88	긍정	경제위기	48	부정
강화하다	612	긍정	사고당하다	87	부정	충격	47	부정
가능하다	501	긍정	감염위험	84	부정	기여하다	46	긍정
피해	376	부정	차별	83	부정	기대	46	긍정
적극적	355	긍정	괴롭히다	82	부정	고통받다	45	부정
과실	293	부정	부족하다	81	부정	위기	45	부정
인정받다	275	긍정	비판	81	부정	일방적	45	부정
바라다	242	긍정	어렵다	80	부정	간사	45	부정
최선	232	긍정	인명피해	79	부정	어려움겪다	44	부정
우려	222	부정	강화되다	77	긍정	좋다	44	긍정
참사	207	부정	손해	76	부정	불행	44	부정
스트레스	192	부정	폭행	75	부정	비극	44	부정
보상받다	180	긍정	안타깝다	74	부정	성공적	43	긍정
감염되다	177	부정	최선다하다	72	긍정	갈등	43	부정
부끄럽다	170	부정	폐업	72	부정	만족스러운결과	43	긍정
범죄	159	부정	논란	71	부정	혐의	41	부정
후유증	157	부정	위반	70	부정	영업정지	41	부정
체계적	156	긍정	해결하다	68	긍정	합리적	40	긍정
노력하다	151	긍정	공감하다	67	긍정	불복하다	40	부정
추락	143	부정	사고발생	66	부정	폭발사고	40	부정
승인받다	129	긍정	높은수준	65	긍정	경제적	39	긍정

희망	123	긍정	불안	64	부정	피해 입다	39	부정
부담	122	부정	관리하다	63	긍정	최고	39	긍정
불법	121	부정	반발	62	부정	혼란	38	부정
고통	114	부정	손상	61	부정	효율적	38	긍정
도움	113	긍정	호소하다	58	부정	치료 가능하다	38	긍정
개선하다	110	긍정	위반하다	56	부정	배려	37	긍정
장애	101	부정	부작용	53	부정	갑질	37	부정
안전하다	97	긍정	부상당하다	51	부정	인정받지 못하다	36	부정
특별	93	긍정	정상적	51	긍정	폭언	36	부정
기대하다	92	긍정	의혹	51	부정	걱정되다	35	부정
효과적	89	긍정						

아래의 <그림 4-9>는 언론보도 기사(좌)와 소셜 네트워크 서비스(우)의 주요 키워드를 시각화하여 비교한 것이다.



<그림 4-9> 2020년 주요키워드 시각화

4.4.6 2021년 시계열 분석

특징 키워드로는 광주 학동4구역 건물철거 붕괴사고 관련 키워드(광주, 철거, 비판, 중대재해, 재해법, 부실, 하청 등) 중대재해처벌법 관련 키워드(처벌법, 법안, 중대재해처벌

법, 제정, 비판, 통과, 민주당, 중대재해, 재해법, 법률, 처벌, 보완 등)을 확인할 수 있다. 산업재해 예방활동 강화를 위한 조직확대와 관련하여 산업안전보건본부 출범 관련 키워드(예방, 산재, 기업, 사업장, 조치, 노동부, 산업재해, 강화, 정부, 확대)를 확인할 수 있다. 아래의 <표 4-16>은 2021년 언론보도 주요키워드를 표현한 것이다.

<표 4-16> 2021년 언론보도 주요키워드

단어	빈도수	단어	빈도수	단어	빈도수	단어	빈도수
산업	475	산업재해	173	폭발	128	교육	106
안전	414	책임	172	사망자	128	요구	106
사고	413	대책	171	원인	128	위원회	106
재해	408	강화	169	관계자	128	설치	106
발생	378	정부	169	경영	127	참사	105
현장	333	처벌법	167	광주	123	국민	105
사망	288	상황	165	경기도	123	사업주	104
관리	273	처벌	165	책임자	121	고용노동부	103
화재	248	업체	164	추락	121	구조	102
예방	246	대표	160	물류	120	센터	102
산재	230	지적	159	진행	118	규모	102
기업	229	점검	158	기관	118	예정	99
작업	222	개선	155	건물	116	전문	98
건설	214	근로자	153	결과	116	하청	97
고용	207	붕괴	151	시설	116	확대	95
위험	203	환경	149	시민	115	중대재해 처벌법	95
노동자	197	조사	146	사회	111	전국	95
사업장	196	확인	145	추진	111	사건	93
보건	193	노동	144	기준	110	지방	92
조치	189	마련	143	위반	109	정책	92
공사	181	감독	142	의무	108	여부	92
계획	178	국회	138	지원	108	전문가	91
시행	175	사업	135	강조	107	회사	91
대상	174	방지	134	내년	107	철거	89
노동부	174	한국	131	보건법	107	코로나	89

2021년의 산업재해 예방활동 강화와 중대재해처벌법 시행준비를 위한 고용노동부 내 조직의 확대와 관련한 산업안전보건본부 신설과 관련된 키워드(안전, 보상, 기여하다, 노력하다, 피해, 획기적, 최선, 적극적, 논란)와 2020년에 이어 산업안전보건법과 중대재해처벌법의 적용과 관련한 업계의 우려와 도입을 부담스러워 하는 부정적인 반응과 관련한 키워드(우려, 비판, 과실, 반발, 유감, 부작용, 위반하다, 보완하다, 차별, 부담, 분노, 피해, 혼란)으로 전년도보다 부정적인 반응들이 더 강하게 다양한 매체를 통해 나타났다. 아래의 <표 4-17>은 2021년 소셜 네트워크 서비스 주요키워드를 표현한 것이다.

<표 4-17> 2021년 소셜 네트워크 서비스 주요키워드

단어	빈도	감정	단어	빈도	감정	단어	빈도	감정
안전	4374	긍정	분노	89	부정	안타깝다	46	부정
보상	1033	긍정	합리적	88	긍정	효율적	45	긍정
위험	463	부정	위반	87	부정	지나치다	44	부정
강화하다	455	긍정	후유증	84	부정	인정하지않다	44	부정
최선	331	긍정	최선다하다	76	긍정	충격	43	부정
과실	280	부정	손해	76	부정	문제	43	부정
우려	275	부정	강화되다	75	긍정	괴롭히다	42	부정
피해	247	부정	혼란	73	부정	심각한문제	42	부정
가능하다	247	긍정	의혹	73	부정	비판하다	40	부정
비판	229	부정	고통	72	부정	바라다	40	긍정
적극적	217	긍정	개선하다	71	긍정	포용	40	긍정
혐의	176	부정	간사	68	부정	최적	40	긍정
해결하다	172	긍정	승인받다	66	긍정	어렵다	39	부정
노력하다	165	긍정	장애	66	부정	심려	39	부정
유감	151	부정	위기	64	부정	보건범죄	38	부정
인정받다	149	긍정	참사	63	부정	긍정적	38	긍정
범죄	148	부정	비극	63	부정	활약	38	긍정
부담	147	부정	아프다	59	부정	효과적	37	긍정
추락	146	부정	조치 취하다	58	긍정	호소하다	36	부정
보상받다	139	긍정	성공적	57	긍정	부정적	36	부정
기여하다	134	긍정	최고	56	긍정	불량	35	부정
안전하다	134	긍정	진정성	56	긍정	감동	35	긍정
위반하다	130	부정	사고발생	55	부정	처벌받다	34	부정

체계적	124	긍정	보완하다	54	긍정	관심가지다	34	긍정
스트레스	121	부정	개선되다	54	긍정	손실	34	부정
부작용	118	부정	사고당하다	52	부정	정전	34	부정
차별	113	부정	공감하다	51	긍정	공포	33	부정
반발	111	부정	부족하다	51	부정	갑질	33	부정
진심	109	긍정	전세계적	50	긍정	취약하다	33	부정
논란	104	부정	관리하다	50	긍정	고통받다	33	부정
불법	103	부정	죄송하다	48	부정	기대하다	33	긍정
획기적	100	긍정	폭행	47	부정	반발하다	32	부정
걱정	97	부정	최악	46	부정	기대	32	긍정
도움	97	긍정						

아래의 <그림 4-10>는 언론보도 기사(좌)와 소셜 네트워크 서비스(우)의 주요 키워드를 시각화하여 비교한 것이다.



<그림 4-10> 2021년 주요키워드 시각화

4.4.7 연도별 주요 키워드 분석

2016년부터 2021년 6월까지 3,647건의 언론보도 기사의 주요키워드 순위 10위까지를 분석하면 산업(6회), 안전(6회), 재해(6회), 사고(6회), 발생(6회), 현장(6회), 관리(5회), 예방(5회), 위험(5회), 화재(4회), 작업(3회), 사망(1회), 고용(1회)의 횟수로 나타났다. 주요 사고 원인(위험,현장,위험,관리,작업)과 결과(재해,사고,발생,화재,사망)와 대책(산업,안전,관리)의 키워드가 나타났다. 2016년부터 2021년 6월까지 소셜 네트워크 서비스 주요키워드 순위 10위까지를 분석하면 안전(6회), 보상(6회), 위험(6회), 피해(6회), 강화하다(6회), 가능하다(5회), 인정받다(4회), 최선(4회), 스트레스(3회), 과실(3회), 우려(2회), 노력하다(1회), 바라다(1회), 불법(1회), 불안(1회), 산재1위(1회), 부담(1회), 해결하다(1회), 적극적(1회), 비판(1회)로 나타났다. 산업재해의 원인(위험, 과실, 불법)과 결과(피해,산재1위)에 따른 우려의 반응(스트레스,우려,불안,부담,비판)과 대책방안(안전,보상,강화하다,가능하다,최선,노력하다,바라다,해결하다,적극적)의 키워드가 등장하였으며 산업재해 해결과 보상을 요구하는 의견이 있었다. 언론보도에서 13개의 키워드로 나타났으며, 소셜 네트워크 서비스에서 20개의 키워드가 나타났으며, 이는 소셜 네트워크 서비스로 다양한 의견이 표출되고 있음을 의미한다. 아래의 <표 4-18>은 2016년부터 2021년까지의 언론보도 연도별 주요키워드를 표현하였고, <표 4-19>는 2016년부터 2021년까지 소셜 네트워크 서비스 연도별 주요키워드를 표현한 것이다.

<표 4-18> 언론보도 연도별 주요키워드 10

순위	2016년	2017년	2018년	2019년	2020년	2021년
1	산업	산업	산업	산업	산업	산업
2	안전	안전	안전	안전	안전	안전
3	사고	사고	사고	사고	사고	사고
4	발생	발생	발생	발생	화재	재해
5	재해	현장	재해	재해	발생	발생
6	현장	재해	현장	현장	현장	현장
7	예방	작업	관리	위험	재해	사망
8	관리	예방	예방	관리	관리	관리
9	작업	위험	화재	작업	위험	화재
10	위험	고용	위험	화재	예방	예방

<표 4-19> 소셜 네트워크 서비스 연도별 주요키워드 10

순위	2016년	2017년	2018년	2019년	2020년	2021년
1	안전	안전	안전	안전	안전	안전
2	보상	보상	보상	보상	보상	보상
3	산재1위	위험	위험	위험	위험	위험
4	위험	강화하다	강화하다	강화하다	강화하다	강화하다
5	강화하다	스트레스	인정받다	인정받다	가능하다	최선
6	스트레스	불법	피해	피해	피해	과실
7	불안	인정받다	우려	스트레스	적극적	우려
8	피해	피해	최선	최선	과실	피해
9	노력하다	최선	부담	해결하다	인정받다	가능하다
10	가능하다	가능하다	과실	가능하다	바라다	비판

언론보도 기사의 특징적인 키워드는 사고 발생과 직·간접적으로 연관된 키워드(산업, 안전, 사고, 현장, 관리, 작업, 건설, 위험, 공사, 지적, 점검 등)와 사고의 결과 관련 키워드(재해, 발생, 사망, 화재, 붕괴, 산업재해, 처벌, 확인 등)와 예방대책의 주체와 대상 관련 키워드(예방, 산재, 노동자, 사업장, 보건, 조치, 계획, 시행, 대상, 노동부, 책임, 대책, 강화, 정부, 업체, 대표, 개선, 근로자, 환경, 조사, 노동, 마련, 감독, 국회, 방지 등)가 주요 키워드로 확인되었다.

소셜 네트워크 서비스의 감정분석 주요키워드는 사건, 사고의 발생에 따른 감정들과 관련된 키워드(스트레스, 피해, 노력하다, 가능하다, 인정받다, 인정받지 못하다, 인정받지 않다, 보상받다, 보상받지 않다, 관리하다, 바라다, 해결하다, 안전하다, 사고당하다, 개선하다, 기여하다, 부담, 최선다하다, 후유증, 비판, 승인받다, 논란, 의혹 등)를 주요 키워드로 확인하였으며 국민들의 사건, 사고, 정책을 접하고 직접적으로 관련 있는 보상, 인정, 노력의 감정들과 개인에게 영향도가 깊은 정책이 주로 나타났다.

아래의 <표 4-20>는 언론보도와 소셜 네트워크 서비스의 연도별 특징 키워드다.

<표 4-20> 언론보도와 소셜 네트워크 서비스의 연도별 특징 키워드

	언론보도 기사	소셜 네트워크 서비스
매년 반복되는 키워드	사업, 안전, 사고, 재해, 발생, 현장, 사망, 관리, 화재, 예방, 산재, 기업, 작업, 건설, 고용, 위험, 노동자, 사업장, 보건, 조치, 공사, 계획, 시행, 대상, 노동부, 산업재해, 책임, 대책, 강화, 정부, 상황, 처벌, 업체, 대표, 지적, 점검, 개선, 근로자, 붕괴, 환경, 조사, 확인, 노동, 마련, 감독, 국회, 사업, 방지	안전, 보상, 위험, 강화하다, 스트레스, 피해, 노력하다, 가능하다, 인정받다, 불법, 보상받다, 과실, 적극적, 손해, 장애, 관리하다, 바라다, 해결하다, 안전하다, 최선, 고통, 사고당하다, 범죄, 개선하다, 도움, 차별, 기여하다, 부담, 최선을다하다, 후유증, 비판, 승인받다, 논란, 의혹, 체계적
2016년 특징키워드	지하철, 남양주, 구의역, 원청업체, 사상자, 스크린, 지칭, 산업현장, 누출, 지하, 부상, 석유, 벌금, 공사장, 하청업체, 시공사, 공사현장, 인정	산재1위, 추락하다, 차별화되다, 개탄, 신속하다, 어려운상황, 문제있다, 행복, 특성물질, 고의적, 유용한, 정보, 허위, 열악하다, 인증받다, 치명적
2017년 특징키워드	크레인, 조선, 해양, 창원, 조선해양, 경남, 타워, 삼성, 과태료, 사법, 중공업, 선박, 타워크레인, 탱크, 석유, 지칭, 위험성, 배치, 적발, 규칙, 수칙, 하청업체, 발견, 밀폐, 삼성중공업, 조선소, 도장	산재1위, 신속하다, 악화되다, 불복하다, 비정상, 발암물질, 많은관심, 압력, 고민, 위험하다, 많편하다, 처우개선, 행복, 안정적, 쾌적한환경, 혜택받다, 흡수, 건강지키다, 스트레스받다, 처벌받다, 완화하다
2018년 특징키워드	교통, 발전, 대통령, 프로젝트, 자살, 서비스, 성장, 산업안전, 정비, 대한민국, 국경, 요양원, 차례, 크레인, 기본, 전방, 교통안전, 반영, 전면, 국민생명, 합의, 일자리, 예산, 보고	안정적, 반발하다, 긍정적, 진심, 갈등, 기부, 만족, 불가능, 불가피하다, 안심, 살인기업, 걱정, 다행, 심각한문제, 공식사과, 과학적, 도움되지않다, 잘못되다
2019년 특징키워드	대전, 발전소, 한화, 화력, 김용균, 발전, 태안, 질서, 추급, 노동청, 제조, 범위, 참석, 산안법, 법령, 적발, 절차, 평균, 규명, 주민, 충남, 승인, 시공	활약, 공감, 분노, 성공하다, 많은기대, 쉽다, 현명한선택, 억울하다, 공포, 안심, 인증하다, 악화되다, 특성물질, 어려운상황, 불편, 위험한상황, 비난, 손상
2020년 특징키워드	물류, 창고, 이천, 제정, 코로나, 처벌법, 경기도, 코로나19, 민주당, 용접, 벌금, 법안, 38명, 중대재해기업처벌법, 시공사, 냉동, 과실, 민주당, 통과, 우레탄, 산안법, 효과, 검토, 긴급, 법인, 정의당	감염되다, 특별, 공감, 감염위험, 부족하다, 폐업, 공감하다, 손상, 반하다, 경제위기, 불행, 성공적, 갈등, 만족스러운결과, 영업정지, 불복하다, 치료가능하다, 배려, 폭언, 걱정되다
2021년 특징키워드	처벌법, 법안, 광주, 물류, 중대재해처벌법, 철거, 코로나, 제정, 창고, 비판, 코로나19, 통과, 민주당, 중대재해, 재해법, 제외, 법률, 보완, 경기도, 이천, 포스코, 경영자, 제조업, 회장, 부실, 유예, 법인, 논란, 해체	유감, 진심, 걱정, 분노, 성공적, 진정성, 보완하다, 공감하다, 부족하다, 죄송하다, 지나치다, 심각한문제, 비판하다, 포용, 심려, 보건범죄, 긍정적, 활약, 부정적, 감동, 처벌받다, 관심가지다, 공포, 반발하다

4.4.8 통찰

연도별 특징적 이슈로 2016년 구의역 사고, 2017년 삼성전자 황유미씨 산재승인, 2018년 국민생명지키기 3대 프로젝트, 2019년 김용균씨 사망사고, 2020년 산업안전보건법 개정과 중대재해처벌법 시행, 2021년 중대재해처벌법 제정이 있었다. 사고에 대한 반응보다 정책 관련 이슈에 더 민감하고 장기간 반응하는 것으로 나타났다. 사고 발생시 1개월간 언론보도나 소셜 네트워크 서비스의 언급량이 증가하다 감소하였으며, 산업안전보건법과 중대재해처벌법과 같은 정책 이슈의 경우 3개월간 언급량이 증가하다 감소하는 추세를 보였다. 산업안전보건법 전부개정시 언급량의 증가보다 벌칙 규정이 강화된 중대재해처벌법 제정과 관련한 언급량이 더 많이 증가하였다. 중대재해처벌법 제정 관련한 소셜 네트워크 서비스 키워드 분석결과를 살펴보면 법안 제정에 따른 산업계의 우려반응과 도입을 부담스러워 하는 부정적인 반응의 키워드가 다수 등장하였다. 법안 적용시의 예상되는 문제점, 법안 시행에 따른 산업계에 미치는 영향, 법안 제정에 따른 산업재해 예방 효과성에 대한 우려의 반응들로 부정적인 키워드가 다수 확인되었다. 중대재해처벌법의 제정이 언급된 2020년 7월보다 2021년 1월에 공표된 시점에 부정적인 반응들이 더 많이, 더 강하게 나타났다.

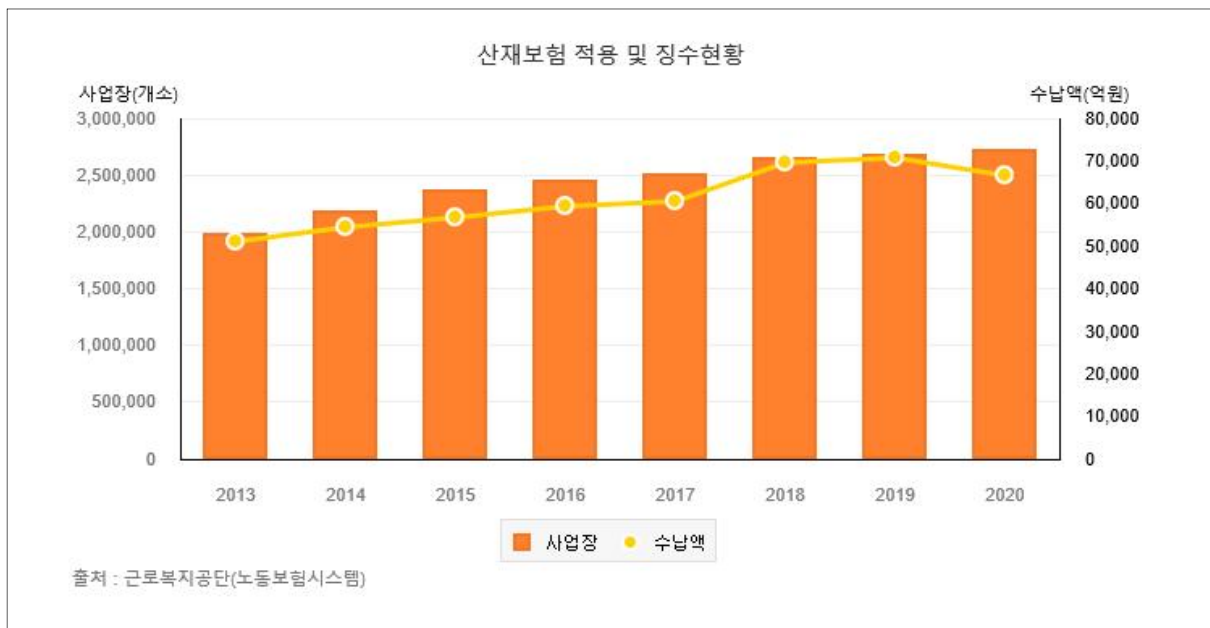
연도별 특징적인 부분으로 보상과 관련한 키워드가 있다. 상반된 의미의 키워드 '인정받다'와 '인정하지 않다', '인정받지 못하다'가 함께 등장했다. 보상 관련한 키워드는 근로복지공단의 산업재해 중 업무상 질병 인정율과 밀접한 관련이 있다. 아래의 <표 4-21> 소셜 네트워크 서비스의 보상관련 키워드 빈도 추이를 살펴보면 2019년까지 긍정키워드 '인정받다'의 비율이 매년 증가하고 있으며 부정키워드 '인정하지 않다'와 '인정받지 못하다'의 비율은 매년 감소하고 있다. 2020년과 2021년의 긍정키워드 '인정받다'의 비율이 감소하고 있다. 이는 근로복지공단의 산재신청 및 인정현황 자료 <표 4-21>에 따르면 산업재해 중 업무상 사고 인정율의 비율은 매년 유사한 94~96% 수준이며, 업무상 질병 인정율 매년 높아지고 있다.(2016년 44.1%, 2017년 52.9%, 2018년 63.0%) 2019년 이후의 산업재해 인정율 자료는 공개되지 않아 확인을 할 수 없었다. 2018년의 산업재해 신청 건수의 급격한 증가는 18년 1월 1일부터 사업주 확인 제도를 폐지와 출퇴근 중 발생한 사고까지 산재보상 대상의 확대로 산재신청이 전년대비 약 2.5만건 증가였다고 볼 수 있다. e-나라지표의 근로복지공단 노동보험시스템 자료 <그림 4-11>을 참고하면 2019년까지 매년 산재보험 수납액이 증가추세이며, 2020년 산재보험 수납액이 소폭 감소한 것을 확인할 수 있다. 2020년 전체 사업장 개소는 증가했지만 보험 수납액 감소에 따라 인정율이 낮아질 것으로 추정해 볼 수 있다. 이를 근거로 2021년 소셜 네트워크 서비스의 인정율이 전년대비 낮아졌음을 추측할 수 있다.

<표 4-21> 소셜 네트워크 서비스의 보상관련 키워드 연도별 빈도와 비율

감정어(언급량)	2016년	2017년	2018년	2019년	2020년	2021년
인정받다	137	167	326	246	275	149
인정하지 않다	78	77	27	28	32	25
인정받지 못하다	41	39	39	19	36	44
합계	256	283	392	293	343	218
긍정비율 (인정받다)	53.52	59.01	83.16	83.96	80.17	68.35
부정비율 (인정하지 않다 인정받지 못하다)	46.48	40.99	16.84	16.04	19.83	31.65

<표 4-22> 산재신청 및 인정현황, 근로복지공단

	2015년	2016년	2017년	2018년
신청 건수	114,167	113,858	113,716	138,576
전체 인정율	89.4	89.7%	89.5%	91.5%
업무상 사고 인정율	94.2	94.9	94.7	95.8
업무상 질병 인정율	44.9	44.1	52.9	63.0



<그림 4-11> 산재보험 사업장 및 징수현황, 근로복지공단

고용노동부에 따르면 특수고용 종사자 중 2019년 1월부터 건설기계 기사 11만명을 산재보험 적용대상에 포함하였고, 2021년까지 방문 서비스 종사자, 화물차주, 돌봄 서비스 종사자 및 정보통신(IT) 업종 자유계약자도 산재보험 대상자에 포함됨에 따라 특수고용형태 근로자의 산재보험 가입이 증가되었다.

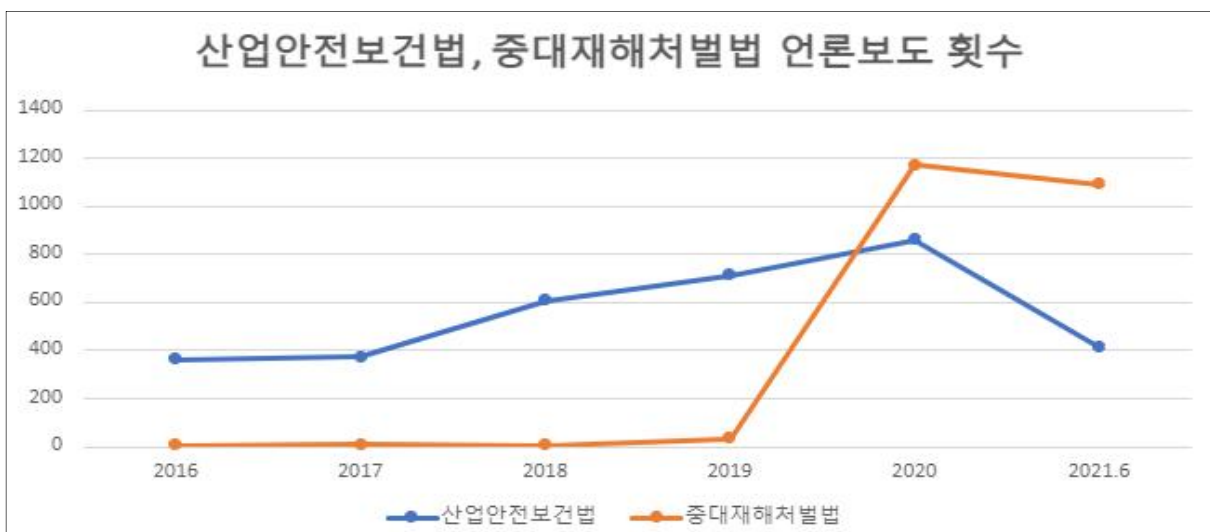
4.5 중대재해처벌법 분석

4.5.1 언론보도 분석

2016년부터 2021년 6월까지 산업안전과 관련된 법률 중 산업안전보건법과 중대재해처벌법을 비교하였다. 산업안전보건법의 언론보도 추이는 매년 증가하고 있으며 2018년 11월 산업안전보건법 전부개정안이 국회 제출되어 언론보도 횟수가 전년대비 대폭 증가하였다. 중대재해처벌법의 언론보도 추이는 2020년부터 급증하였다. 2020년 7월 정부의 중대재해처벌법 입법예고안 발의에 따라 그해 7월부터 보도 횟수가 증가하였다. 2021년 1월에 중대재해처벌법 제정시 더 많은 언론기관에서 보도하였다. 아래의 <표 4-23>과 <그림 4-12>은 산업안전보건법과 중대재해처벌법의 언론보도 횟수를 표와 그래프로 표현하였다.

<표 4-23> 산업안전보건법, 중대재해처벌법의 언론보도 횟수

연도	산업안전보건법	중대재해처벌법
2016	364	5
2017	374	8
2018	606	4
2019	712	35
2020	861	1,172
2021.6	413	1,091



<그림 4-12> 산업안전보건법, 중대재해처벌법 언론보도 횟수 비교

언론보도의 중대재해처벌법 주요 키워드를 살펴보면 법안 제정과 관련된 법안 처리 상황, 다양한 계층, 법안명 등이 관련 키워드 다음과 같다. '중대', '재해', '처벌법', '기업', '국회', '산업', '안전', '제정', '처벌', '사고', '대표', '발생', '노동자' 순으로 나타났다. 아래의 <표 4-24>는 2016년부터 2021년 6월까지 소셜 네트워크 서비스 연도별 주요키워드를 표현한 것이다.

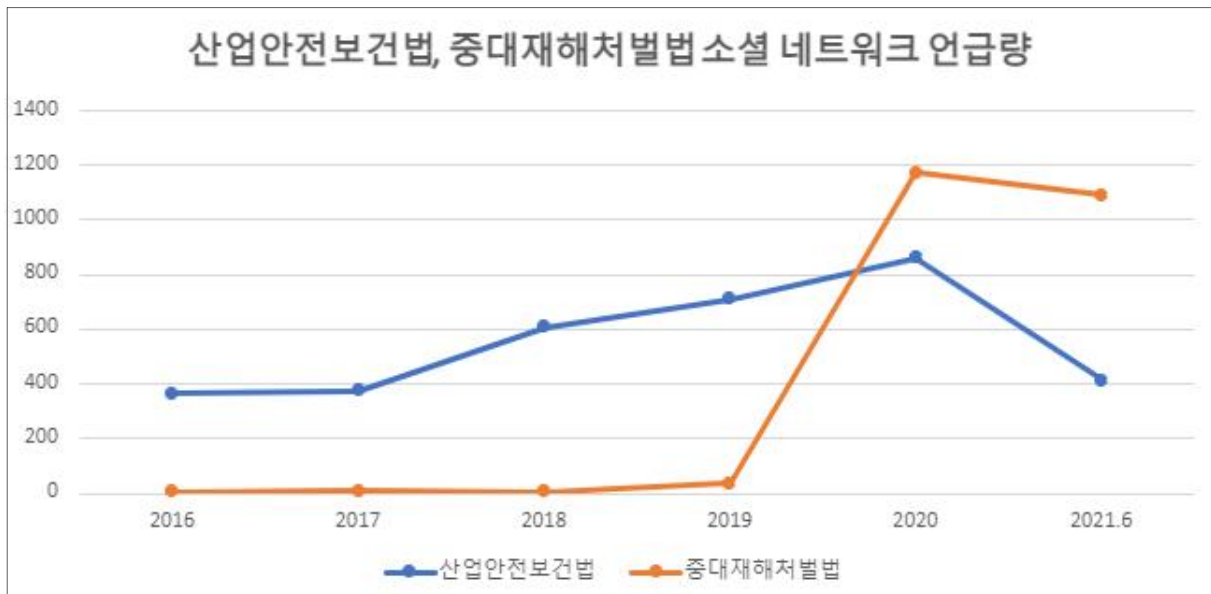
<표 4-24> 중대재해처벌법 언론보도 주요키워드 빈도

키워드	2016	2017	2018	2019	2020	2021.6	총합계
중대	5	8	4	34	1,172	1,092	2,315
재해	5	8	4	34	1,168	1,092	2,311
처벌법	5	8	4	33	1,137	978	2,165
기업	5	8	4	33	1,154	874	2,078
국회	3		4	17	943	646	1,613
산업		6	3	25	668	721	1,423
안전	5		4	30	668	712	1,419
제정	5	7		24	881	487	1,404
처벌	5	6	4	26	678	643	1,362
사고	5	7	3	27	597	702	1,341
대표	4			20	705	545	1,274
발생	5	8	2	22	531	646	1,214
노동자	4	6	4	30	633	503	1,180
책임	5	6	3	23	560	541	1,138
사망	4		4	24	517	552	1,101
정부		6		20	524	551	1,101
법안			2		663	406	1,071
국민					601	429	1,030
중대재해기업처벌법	4		2	27	761		794
민주당					702		702
정의당					637		637
의원					601		601
촉구					597		597
현장		6	3	26		551	586
입법					512		512
시행						510	510
노동			3	18	468		489
경영						488	488
처리					477		477

지적		5				448	453
요구					451		451
농성					450		450
개정					448		448
적용						440	440
사업장						438	438
대상						437	437
통과			2			434	436
중대재해법						422	422
책임자						415	415
중대재해처벌법						413	413
상황						401	401

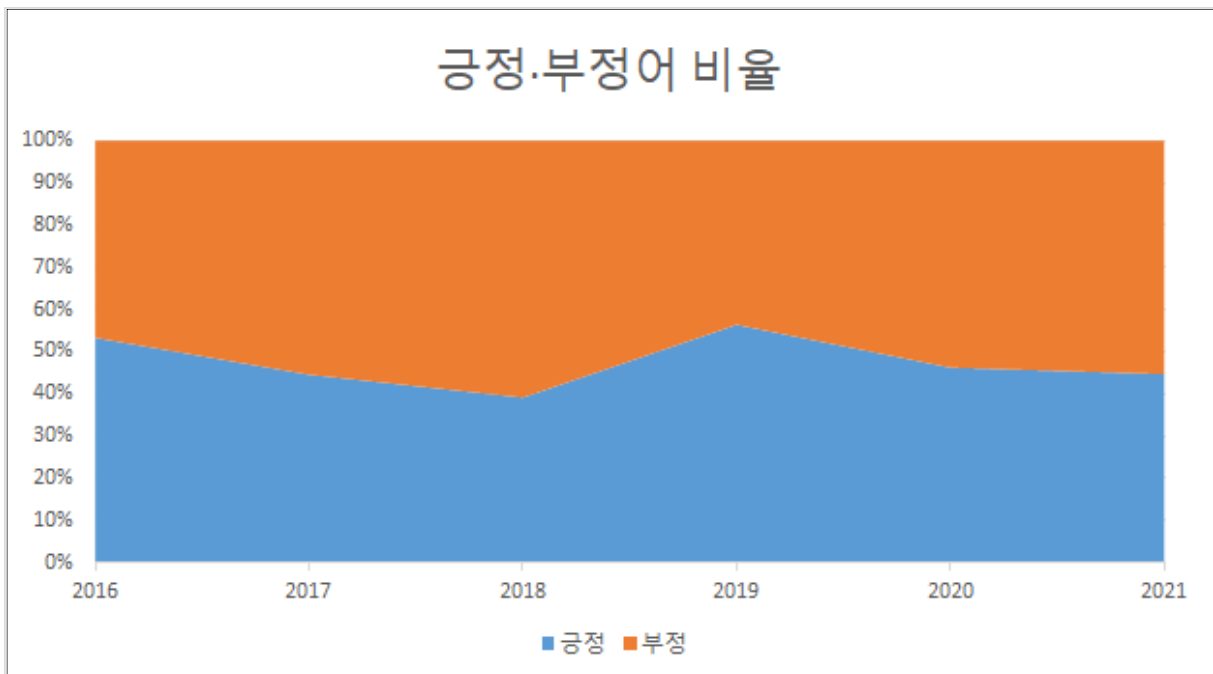
4.5.2 소셜 네트워크 서비스 분석

소셜 네트워크 서비스의 언급량은 2018년 11월 산업안전보건법의 전부개정안이 국회에 제출되어 언급량이 급증하였으며, 2019년부터 점차 감소하였다. 2020년 7월 중대재해처벌법의 입법예고 발의에 따라 언급량이 급증하였으며 2021년 1월 26일 중대재해처벌법의 법안의 제정과 2022년 1월 시행이 언급량의 증가로 나타났다. 아래의 <그림 4-13>은 2016년부터 2021년 6월까지 산업안전보건법과 중대재해처벌법의 소셜 네트워크 서비스 연도별 언급량을 표현한 것이다.



<그림 4-13> 산업안전보건법, 중대재해처벌법 소셜 네트워크 서비스 언급량

중대재해처벌법 관련키워드로 ‘안전’, ‘강화하다’, ‘위험’, ‘우려’, ‘어렵다’, ‘부족하다’, ‘부족한 부분’, ‘범죄’, ‘비판’, ‘해결하다’ 순으로 나타났다. 2016년부터 2019년까지 언급된 데이터 언급량은 5%에 불과하였으며, 2020년 33%, 2021년 62%로 나타났다. 법령 예고(2020.7)한 이후의 시점부터 데이터가 폭발적으로 증가하기 시작하였다. 2016년부터 2011년까지는 긍정어의 비율이 2016년 53.09%, 2017년 44.58%, 2018년 39.14%, 2019년 56.40%, 2020년 46.25%, 2021년 46.60%으로 나타났다. 2016년부터 긍정어 비율이 2018년 낮아졌다가 2019년도 급격하게 올라가는 추세를 보이다 중대재해처벌법의 제정과 시행의 영향으로 2020년과 2021년은 46% 수준으로 낮아졌다. 아래의 <그림 4-14>는 2016년부터 2021년 6월까지 중대재해처벌법의 소셜 네트워크 서비스의 긍정·부정어 비율을 표현한 것이다.



<그림 4-14> 중대재해처벌법 감정어 비율

소셜 네트워크 서비스의 주요 키워드를 살펴보면 중대재해처벌법 법안 제정으로 우려의 목소리들이 많이 등장하였다. 법안 적용시의 문제점과 실제 적용했을 때 산업계에 미치는 영향 부분과 법안으로 산업재해 예방의 효과성에 대한 부분들, 법안 실제 적용의 우려의 목소리들이 키워드로 나타났다. 키워드는 다음과 같다. ‘우려’, ‘어렵다’, ‘부족하다’, ‘부족한 부분’, ‘범죄’, ‘비판’, ‘논란’, ‘차별’, ‘반발’, ‘반대하다’, ‘차별하다’, ‘비판하다’, 등의 부정적인 반응들이 대부분이었다. 반면 긍정적인 키워드는 ‘안전’, ‘강화하다’, ‘최선’, ‘원하다’, ‘원칙을 지키다’으로 나타났다. 아래의 <표 4-25>는 2016년부터 2021년까지 중대재해처벌법의 소셜 네트워크 서비스 연도별 주요키워드 빈도를 표현한 것이다.

<표 4-25> 중대재해처벌법 감정어 연도별 빈도

키워드	2016	2017	2018	2019	2020	2021	총합계
안전(긍정)	99	72	163	297	5,485	10,217	16,333
강화하다(긍정)		3		72	1,275	4,280	5,630
위험(부정)		32	112	121	1,089	1,942	3,296
우려(부정)					1,030	1,440	2,470
어렵다(부정)				13	536	1,872	2,421
부족하다(부정)						2,373	2,373
부족한부분(부정)					2,057	2,057	3,682
범죄(부정)	2			13	1,087	746	1,848
비판(부정)					781	1,027	1,808
해결하다(긍정)		2	4			1,777	1,783
논란(부정)		2			714	663	1,379
차별(부정)		3	2	5	373	966	1,349
반발(부정)					614	734	1,348
최선(긍정)		6		28	422	858	1,314
반대하다(부정)					489	699	1,188
차별하다(부정)						1,178	1,178
비판하다(부정)			3	5	459	649	1,116
최악(부정)	12	4	10	7		1,062	1,095
원하다(긍정)	1					1,006	1,007
원칙지키다(긍정)				964		964	1,197
위반하다(부정)	2	7			396	516	921
거부하다(부정)						890	890
호소하다(부정)				6	457	418	881
과실(부정)			2		427	450	879
부담(부정)					284	500	784
외면하다(부정)					302	450	752
참사(부정)	3	31	15	18	637		704
공감하다(긍정)					685		685
반발하다(부정)						671	671
부담스럽다(부정)				657		657	516
꼼수(부정)						598	598
바라다(부정)				8	557		565
보상(부정)						520	520
불매(부정)						448	448
충격적(부정)						437	437
갑질(부정)					405		405

안전하다(긍정)					372		372
유감(부정)				7	342		349
적극적(긍정)		7			322		329
밀려나다(부정)					301		301
염려하다(부정)					297		297
살인기업(부정)	15	12	19	91			137

아래의 <그림 4-15>은 중대재해 처벌법의 언론보도 기사(좌)와 소셜 네트워크 서비스(우)의 주요 키워드를 시각화하여 비교한 것이다.

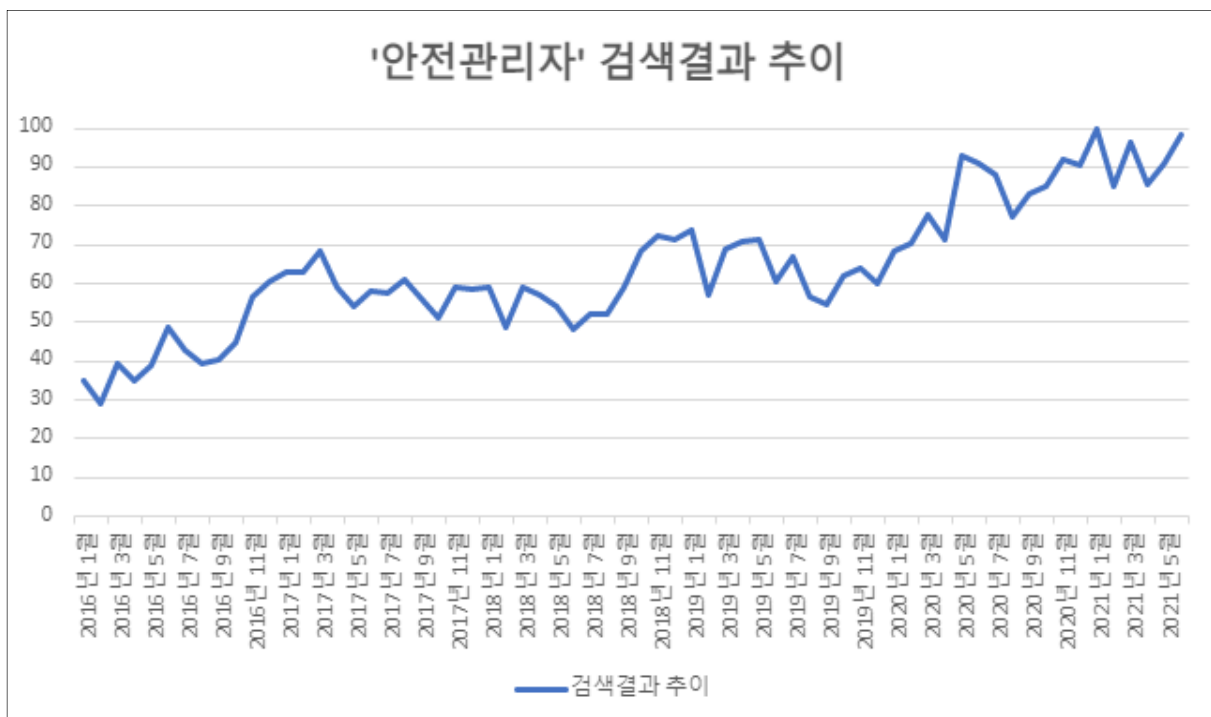


<그림 4-15> 중대재해처벌법 주요키워드 시각화

4.5.3 통찰

28년만에 전면 개정된 산업안전보건법보다 벌칙규정이 강화된 중대재해처벌법에 언론보도, 소셜 네트워크 서비스 언급량 모두 급격하게 증가하였다. 언론보도에서 2021년 전년보다 많이 등장한 키워드는 사고 키워드(사고, 발생, 사망)의 실제 적용(현장, 시행, 적용, 사업장, 대상)의 법적 책임 관련 키워드(책임자, 경영)가 언론보도를 통해 등장하였으며, 소셜 네트워크 서비스에서 법정 제정을 환영하는 목소리(강화하다, 해결하다, 원칙지키다, 공감하다)의 키워드가 등장하지만 반대의 키워드(위험, 우려, 어렵다, 부족하다, 비판, 반발, 차별하다, 거부하다, 부담)가 다양하고 강하게 나타났다. 현장에서 해야 할 의무사항의 세부내용을 담은 가이드를 만들어야 부족한 부분을 채워야 할 것으로 보인다.

중대재해처벌법의 대응으로 기업들은 안전관리자를 채용하는 것으로 예상된다. 네이버 데이터랩을 통해 2016년 1월부터 2021년 6월까지 네이버에서 '안전관리자'를 검색한 횟수를 확인하였다. 월 단위로 구분하였으며, 가장 많이 검색된 월의 값을 기준값(100)으로 표현하고 월별 상대적 수치를 표현했다. 산업안전보건법 전면개정과 중대재해처벌법이 제정된 2020년부터 안전관리자의 검색량이 증가하였다. 2021년 1월 중대재해처벌법이 제정된 시점에 가장 많은 검색량이 확인되었다. 아래의 <그림 4-16>은 2016년부터 2021년 6월까지 네이버의 안전관리자 검색결과 추이를 표현한 것이다.



<그림 4-16> 안전관리자 검색결과 추이

기획재정부가 운영하는 공공기관 채용정보시스템(잡알리오, job.alio.go.kr)에서 안전관리자의 채용공고 현황을 확인하였다. 동일한 건과 채용고는 제거하였으며, 확인결과 공공기관에서 안전관리자의 채용이 증가하는 추세로 확인되었으며 고용은 비정규직의 형태가 가장 많았다. 민간기관의 채용공고현황을 함께 확인하려 하였으나 과거 채용공고에 대한 데이터 접근이 어려워 정보가 공개된 잡알리오의 채용공고를 기준으로 작성하였다. 아래의 <표 4-26>는 잡알리오를 통한 공공기관의 안전관리자 채용공고현황이다.

<표 4-26> 공공기관 안전관리자 채용공고현황

년도	합계	정규직	비정규직	무기계약직
2016	4	1	3	
2017	6	2	4	
2018	10	5	1	4
2019	13	3	6	4
2020	12	2	7	3
2021.6	8	1	7	
합계	53	14	28	11

5. 결론 및 맺음말

5.1. 연구의 결론 및 의의

빅데이터 분석방법 중 텍스트 마이닝 분석방법을 적용하여 언론보도 기사와 소셜 네트워크 서비스 데이터를 분석하였다. 특히 소셜 네트워크 서비스 데이터는 사회변화를 모니터링할 수 있는 데이터로 정책이나 사회적 이슈에 대한 국민들의 관심사나 생각을 파악할 수 있다는 점에서 데이터 활용방안을 제시하고자 한다. 텍스트 마이닝 분석을 위해 빅카인즈, 썬트렌드, 파이썬 프로그램을 사용하였다. 본 연구의 연구결과를 토대로 얻은 결론은 다음과 같다.

한국언론진흥재단의 빅카인즈에서 언론보도 기사를 분석하였다. 분석기간은 2016년부터 2021년 6월까지 3,647건의 기사(2016년 443건, 2017년 579건, 2018년 649건, 2019년 647건, 2020년 847건, 2021년 482건)를 수집·분석하였다. 언론사 중에서 머니투데이가 220회, 경향신문 206회, 아시아경제 180회, 한국경제 148회, 한겨레 140회를 기사로 보도하였다. 소셜 네트워크 서비스(커뮤니티, 인스타그램, 블로그, 뉴스, 트위터)의 언급량은 83,512건(2016년 11,993건, 2017년 12,052건, 2018년 12,927건, 2019년 14,185건, 2020년 17,184건, 2021년 15,171건)으로 긍정키워드 41~46개, 부정키워드 54~59%로 나타났다. 언론보도 횟수와 소셜 네트워크 서비스의 언급량은 전반적으로 증가하였으며, 이는 언론과 국민들의 산업안전에 대한 사회적 관심도가 증가하고 있음을 알 수 있다.

언론보도 기사의 주요키워드와 소셜 네트워크 서비스 감정키워드 분석결과 중 2016년부터 2021년의 연도별 주요키워드를 분석하면 언론보도에서는 사고의 원인, 결과 관련 키워드(현장, 관리, 위험, 화재, 작업, 사망)와 예방 대책 관련 키워드(안전, 관리, 예방, 고용)로 나타났으며 소셜 네트워크 서비스에서는 정책 관련 키워드(산재1위, 강화하다, 적극적, 최선), 보상 관련 키워드(보상, 인정받다, 가능하다), 피해 관련 키워드(스트레스, 불안, 우려, 불법, 피해, 과실)로 나타나 소셜 네트워크 서비스를 통해 다양한 의견이 있음을 알 수 있다.

소셜 네트워크 서비스의 특징적인 부분으로 보상 관련 키워드가 있다. 상반된 의미 키워드(인정받다, 인정받지 않다, 인정받지 못하다)가 동시에 등장했다. 산업재해 승인과 관련된 부분으로 근로복지공단 업무상 질병 인정율, 산재보험 수납액 추이를 따라 긍정어(인정받다)의 비율이 변화함을 확인하였다. 근로복지공단의 2018년까지 공개된 산재 질병 인정률 자료를 바탕으로 산업재해 질병 인정율이 증가하면 긍정키워드(인정율)의 비율이 증

가하였고, 2019년 이후는 산재보험 수납액으로 확인하였으며 2019년까지 수납액이 증가하다 2020년부터 소폭 감소하였다. 이는 긍정어(인정받다)의 비율도 2019년 증가하였고, 2020년부터 감소하였다. 따라서 언론보도와 소셜 네트워크 서비스의 콘텐츠를 분석하여 이슈가 되는 산업 분야와 현황을 조기에 파악하고 선제적 대응방안으로 활용될 수 있을 것이다.

연도별 특징적 이슈로 2016년 구의역 사고, 2017년 삼성전자 황유미씨 산재승인, 2018년 국민생명지키기 3대 프로젝트, 2019년 김용균씨 사망사고, 2020년 산업안전보건법 개정과 중대재해처벌법 시행, 2021년 중대재해처벌법 제정이 있었으며, 사고에 대한 반응보다 정책 관련 이슈에 더 민감하고 장기간 반응하는 것으로 나타났다. 사고 발생시 1개월간 언론보도나 소셜 네트워크 서비스의 언급량이 증가하다 감소하였으며, 정책 관련 이슈의 경우 3개월간 언급량이 증가하다 감소하는 추세를 보였다. 산업안전보건법 전부개정시 언급량의 증가보다 벌칙 규정이 강화된 중대재해처벌법 제정과 관련한 언급량이 더 많았다. 중대재해처벌법 제정 관련한 소셜 네트워크 서비스 키워드 분석결과를 살펴보면 법안 제정에 따른 산업계의 우려반응과 도입을 부담스러워 하는 부정적인 반응의 키워드가 다수 등장하였다. 법안 적용시의 예상되는 문제점, 법안 시행에 따른 산업계에 미치는 영향, 법안 제정에 따른 산업재해 예방 효과성에 대한 우려의 반응들로 부정적인 키워드가 다수 확인되었다. 중대재해처벌법의 제정이 언급된 2020년 7월보다 2021년 1월에 공표된 시점에 부정적인 반응들이 더욱 강하게 나타났다. 다양한 산업계와 국민들의 목소리를 듣고 산재예방을 위한 규제정책과 지원정책의 개선으로 이어져 산업재해가 감소하기를 기대한다.

연구방법 5단계가 텍스트 마이닝 분석기법의 실무적 활용에 의미를 가지며, 산업안전분야의 이슈사항을 조기에 파악할 수 있을 것이며, 연도별 주요키워드 분석을 통해 국민들의 관심분야와 관심도의 양을 객관적으로 파악할 수 있다. 분석된 정보를 기반으로 산업안전 정책수립의 정보로 활용할 수 있을 것이다. 또한 분석된 결과를 해석하는 통찰의 과정으로 특징적인 키워드를 추출하고, 키워드간 연관관계를 파악하여 사회적 변화의 흐름을 파악할 수 있을 것이다.

5.2. 연구의 한계 및 향후 연구방향

본 연구의 한계점 및 향후 연구를 제시하고자 한다. 빅카인즈의 54개 언론의 뉴스기사와 썸카인즈의 소셜분석 서비스를 활용하여 산업안전과 관련된 데이터를 분석하였다. 뉴스기사의 데이터 수집·가공 단계에서 중복데이터와 관련없는 데이터를 제거 등의 노력을 하였으나 연구자가 예측하지 못한 데이터가 포함될 가능성과 관련없는 데이터를 제거하지 못할 가능성이 존재한다. 또한 연구자의 주관을 완전히 배제하지 못할 가능성도 존재한다. 분석에 사용한 단어사전에 최신 단어가 포함되지 않아 별도로 추가하는 작업이 필요하고, 활용 분야별 전문용어도 추가하여 분석해야 한다. 소셜 네트워크 서비스의 데이터의 경우 플랫폼에서 저장하고 있는 데이터를 크롤링하여 저장하고 있다. 과거 데이터 삭제, 계정 삭제 또는 비공개 처리로 내용을 확인하지 못하는 경우도 상당부분 존재했다. 최신의 데이터는 대부분 접근이 가능했지만, 과거 데이터의 경우 제약사항이 있었다. 향후 연구에서는 소셜 네트워크 서비스 범위의 확대와 데이터 분석을 위해 작성자의 나이, 성별 같은 일반적인 정보를 추가한다면 성별과 연령대별로 체감하는 감정들도 분석할 수 있을 것이며, 감정어에 대한 문장의 전체 흐름을 분석할 수 있는 방법의 연구도 필요하다. 최소단어의 분석뿐만 아니라 문장 단위로 자연어 처리의 범위를 넓히는 기술분석 기법이 추가된다면 데이터 분석의 정확도가 증가할 것이다. 산업안전에 대한 국민의 관심도 증가에 따라 언론보도와 소셜 네트워크 서비스의 언급량이 증가하고 있으며 언론보도와 소셜 네트워크 서비스의 반응을 확인하여 정책추진 방향에 대한 피드백으로 활용하여 산업재해 예방정책 개선방향으로 발전되어야 할 것이다.

참고문헌

e-나라지표 <https://www.index.go.kr/>
고용노동부, <http://www.moel.go.kr/>
근로복지공단, <https://www.kcomwel.or.kr/>
꼬꼬마프로젝트, <http://kkma.snu.ac.kr/>
네이버 데이터랩, <https://datalab.naver.com/>
네이버 지식백과, <https://terms.naver.com/>
위키백과, <https://ko.wikipedia.org/>
잡알리오, <https://job.alio.go.kr/>

국내문헌

- 강현빈(2019), Convolutional Neural Networks 기반 건설재해 분류 모델 개발에 관한 연구에 관한 연구, 이화여자대학교, 석사학위논문.
- 김동준(2021), 텍스트 마이닝을 통한 소셜 커머스 사용자 경험과 서비스 품질 개선에 관한 연구: 코로나-19 전/후를 기준으로, 서울시립대학교, 석사학위논문.
- 김연철(2016), 데이터마이닝 기법을 이용한 국내 건설현장 안전사고 예측에 관한 연구, 경기대학교 건설·산업대학원, 석사학위논문.
- 김준수(2015), 빅데이터 분석 방법을 활용한 건설현장 안전사고 요인분석, 영남대학교, 석사학위논문.
- 김홍관(2020), 텍스트마이닝을 이용한 산업재해 원인의 탐색적 연구: 논문과 언론보도 기사를 중심으로, 인하대학교 대학원, 박사학위 논문
- 문진(2019), 빅데이터 분석기법을 활용한 건설현장 안전사고 분석 적용: 부상재해 사례의 발생형태 중심으로, 서울시립대학교, 석사학위논문.
- 문진혁(2020), 텍스트 마이닝 기법을 활용한 전자사업의 재해 특성 분석: 전자 산업생태계 사업장을 중심으로. 울산대학교. 석사학위 논문.
- 안효선(2017), 빅데이터를 활용한 패션디자인 감성분석 연구: 텍스트마이닝과 의미연결망 분석을 중심으로, 이화여자대학교, 박사학위논문.
- 이수아(2018), 이중차분모형을 이용한 산업안전보건규제의 효과 분석, 한양대학교, 박사학위 논문.
- 이주용(2013), 공간빅데이터로서 소셜미디어 자료를 활용한 공간정보 추출 및 표현 시스템 개발” , pp.1.
- 이희영(2019), 온라인 뉴스 기사와 댓글 간 네트워크 의제 설정 효과: 미세먼지 원인에 대한 이슈 속성 네트워크 비교, 연세대학교, 석사학위논문.

정용찬(2013), 빅데이터, 서울:커뮤니케이션북스.

최정용(2021), 빅 데이터에서 나타나는 텍스트 감정 분류 연구, 인제대학교, 석사학위논문.

현진우(2015), 텍스트 마이닝을 이용한 버스의 차내안전 사고분류 재정립, 명지대학교, 석사학위논문.

홍석은(2018), 빅데이터 마이닝에 의한 환경민원의 감성분석 및 공간패턴분석, 인하대학교, 석사학위논문.

Abstract

An Exploratory Analysis of Industrial Safety Field Using Text Mining Analysis

SunKyu PARK, Department of Safety and Health, University of Ulsan

Industrial accidents and workplace accidents are emerging as social issues, and the public's interest in safety is increasing. Accordingly, the Moon Jae In government is pushing for three major projects to protect the lives of the people, completely revising the Occupational Safety and Health Act and enacting the Serious Accident Punishment Act(tentative title) to reduce the number of deaths from industrial accidents by 50 percent by 2022.

In the last six years (from 2016 to June 2021), 3,647 media reports related to occupational safety and 83,512 contents of social network services (Instagram, Twitter, and blogs) were analyzed. The analysis of media reports was 443 in 2016, 579 in 2017, 649 in 2018, 647 in 2019, 847 in 2020, and 482 by June 2021. Social network services (communities, Instagram, blogs, news, and Twitter) are on the rise in the number of media reports and social service mentions to 11,993 in 2016, 12,052 in 2017, 12,927 in 2018, 14,185 in 2019, 17,184 in 2020, and 15,171 in 2021. This indicates that social interest is increasing.

By Analyzing the main keywords of the last six years, you can identify the distinctive keywords of each year. Keywords related to the 2016 elevator accident on Seoul Subway Line 2, keywords related to the 2017 Samsung Heavy Industries crane accident, keywords related to the 2018 Moon Jae In government's national life protection project, keywords related to the 2019 Tae An heat power plant accident, keywords related to the COVID-19 and the completely revising the Occupational Safety and Health Act and enacting the serious accident punishment Act in 2020, keywords related to the enactment of the serious accident punishment law in 2021 were identified. The number of mentions of the accident issue increased over the past month, policy issue increased over three month. When the Occupational Safety and Health Act was completely revised in 2020, the number of references to related keywords increased. When it comes to the Critical Disaster Penalty

Act in 2021, which strengthened punishment regulations, the amount of references went up and down even before the enactment and it surged at the time of promulgation. In addition, looking at the analysis result of the emotional words related to the serious accident punishment Act, a number of keyword related to 'issue expected when applying the enacted act', 'its impact on industries', or 'negative reactions concerning about the effectiveness of industrial accident prevention' were found so that reflecting effective policy improvement and voices from various background is required for the future revision.

As social interest in occupational safety increases, it is meaningful in that it uses text mining analysis techniques of unstructured data from media reports and social network services to identify people's interests and suggest ways to utilize them. With the revision of the Occupational Safety and Health Act and the enactment of the serious accident punishment Act, it should be developed to improve and supplement industrial safety policies to reduce industrial accidents by analyzing interests and concerns in the public and industry.