

# 안전기준 자동검색을 위한 지원도구 시범개발

## Pilot Development of Supporting Tools for Automatic Detection of Safety Standards

임수정, 박덕근  
국립재난안전연구원

Sujung Im(sj7029@korea.kr), Dugkeun Park(drpark@korea.kr)

### 요약

사회가 발달하면서 법령의 규모는 점차 증가하고 있을 뿐만 아니라 내용도 복잡해지고 있다. 법령 내에 존재하고 있는 안전기준의 규모 또한 증대되고 있고 전문화되고 있어 안전기준 상호간의 상충이나 중복 등을 최소화하기 위한 통합관리에 어려움이 있다. 안전기준의 통합관리를 위해서는 기본적으로 법령 내의 안전기준을 검색하여 추출하는 기술이 우선 확보되어야 한다. 본 연구에서는 한정된 시간 및 인력 등의 한계를 고려하여 몇 가지 특정사례를 기반으로 안전기준 자동검색 지원도구를 개발하고자 하였다. 이를 위해 우선 기존 연구에서 분류된 안전기준과 행정안전부에서 고시한 안전기준을 대상으로 안전기준이 가진 특성을 분석하고 2018년에 국립재난안전연구원에서 추출한 안전기준을 포함하는 법령정보를 수집하였다. 수집된 법령을 정제하고 형태소 분석을 실시하여 안전기준 시소러스를 구축한 후 색인 작업을 거쳐 안전기준 검색도구를 개발하였다. 향후 안전기준을 찾는 데 자동검색도구가 일상적으로 적용된다면 방대하고 복잡한 안전기준의 중복 또는 상충 문제해결에 도움이 될 것으로 기대된다.

■ 중심어 : | 안전기준 | 안전규제 | 안전관리 | 자동검색 | 시스템 |

### Abstract

With the development of society, the scale of the statute is not only increasing, but also the content is getting complicated. The scale of safety standards existing in the law is also increasing and specialized, making it difficult to integrate and manage to minimize conflicts or overlaps among safety standards. For the integrated management of safety standards, a technology that searches for and extracts safety standards in laws and regulations must first be secured. In this study, considering the limitations of time and manpower, a tool for automatic detection of safety standards is developed based on several specific cases. The safety standards classified in the previous studies and the safety standards announced by the Ministry of Interior and Safety were analyzed, and also statute information which includes safety standards extracted by the National Disaster Management Institute in 2018 was collected. After the collected laws were refined and morphological analysis was performed, a safety standard thesaurus was constructed and indexed to develop a safety standard search tool. When automatic search tools are routinely applied to find safety standards in the future, it is expected that these tools will help to solve overlapping or conflicting problems of complex safety standards.

■ keyword : | Safety Standards | Safety Regulation | Safety Management | Automatic Search | System |

## I. 서론

4차 산업혁명시대를 맞이하여 사물인터넷, 클라우드 컴퓨팅 등 기술과 산업이 발전하면서 인간은 현재보다 편리하고 풍요로운 생활을 누릴 수 있는 가능성이 높아졌다. 하지만 기술혁명으로 인해 발생할 수 있는 예기치 못한 사고가 발생할 가능성도 함께 높아졌다. 예를 들면, 2015년 독일 카셀 폭스바겐 자동차 공장에서 로봇에 잡힌 후 강판에 충돌한 작업자가 사망하였고, 2016년 미국에서는 자율주행자동차의 오작동으로 운전자가 사망하였다. 이처럼 예기치 못한 신종사고가 이미 발생하였지만 이에 대응하거나 관련된 안전기준은 현재 충분히 마련되지 못한 상태이다. 또한 기존에 존재했던 안전기준이 새로운 기술분야와 산업분야에서 상충되기도 한다. 따라서 안전기준에서의 문제점(미비/상충/누락 등) 발생으로 국민안전이 위협받을 수 있다는 점을 심각하게 인식할 필요가 있다.

법령 내에 존재하는 안전기준은 법령의 제·개정에 따라 새롭게 생산되거나, 기존에 있던 내용이 끊임없이 수정되기 때문에 실시간으로 전체법령을 대상으로 안전기준을 추출분석하고 관리하는 일은 단순한 작업이 아니다. 이러한 이유로 사람이 직접 법령조문을 독해하여 안전기준을 분류하고 추출하여 분석하는 일은 비효율적일뿐만 아니라 시간과 비용이 많이 소요된다. 따라서 본 연구에서는 안전기준의 통합관리시스템 개발의 가능성을 확인하고, 더 나아가 안전기준 통합관리 시스템의 기반을 마련하고자 특정 사례에 한정하여 안전기준 자동검색 지원도구를 개발하고자 하였다.

## II. 이론적 배경

### 1. 안전기준 정의 및 관리현황

안전기준은 사전적으로는 어떤 유해한 환경 조건이 어느 한계를 넘어야만 안전할 때에 그 한계가 되는 값으로 정의될 수 있을 것이다. 「재난 및 안전관리 기본법」 제 3조4의 2에서는 안전기준을 각종 시설 및 물질 등의 제작, 유지관리 과정에서 안전을 확보할 수 있도록 적용하여야 할 기술적 기준을 체계화한 것이라고 정의하고 있다. 동법 시행령에서는 안전기준의 분야 및 범위를 총 8가지(① 건축시설/② 생활 및 여가/③ 환경 및 에너지/④ 교통 및 교통시설/⑤ 산업 및 공장/⑥ 사이버안전분야를 제외한 정보통신/⑦ 보건·식품/⑧ 그 밖의 분야)로 분류하고 있다. 여기서 ‘그 밖의 분야’는 안전기준심의회에서 안전관리를 위하여 필요하다고 정한 사항과 관련된 안전기준분야를 말한다. 안전기준심의회는 안전기준의 주요안건을 검토하고 지원하기 위한 회의체로서 안전기준심의회 운영규정(행정안전부령 제15호)에 근거하여 구성 운영되고 있다[1]. 세부적으로는 각 부처의 다양한 안전기준을 일괄 조사하여 심의, 등록함으로써 안전기준 상호간 중복, 상충되는 사항을 조정하고, 미비한 안전기준은 발굴하여 개선해나가는 업무를 수행하고 있다[2]. 여기서 안전기준의 등록 및 심의에 관한 내용은 재난 및 안전관리 기본법 제 34조의 7(안전기준의 등록 및 심의 등)에 명기되어 있으며, 중앙행정기관의 장이 안전기준을 신설하거나 변경하고자 할 때는 행정안전부 장관에게 안전기준의 등록을 요청해야하며, 행정안전부 장관은 등록요청 받은 안전기준을 안전기준심의회 심의를 거쳐 관계중앙행정기관 장에게 통보해야 한다는 것이 주요내용이다[3]. [그림 1]. 안전기준심의회는 2015년 11월 4일 제1회를 시작으로 2019년 9월 기준으로 총 9회까지 개최되었고, 안전기준심의회를 통해 현재까지 1,328개의 안전기준이 심의·등록되었다[표 1].

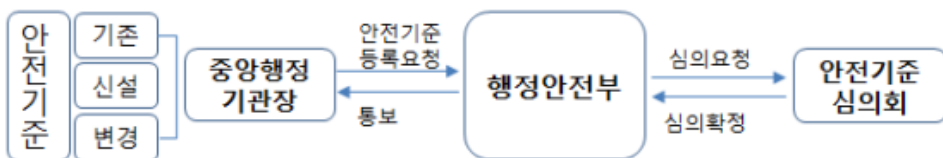


그림 1. 안전기준심의회 운영프로세스

표 1. 안전기준심의회 개최 현황

구분	날짜	내용
1회	'15.11.04.	- 안전기준심의회 운영방안 보고 및 논의
2회	'16.12.06.	- 안전기준 등록·심의 추진계획 보고
3회	'17.08.29.	- 국토부 소관 안전기준 200개 심의·의결
4회	'17.12.15.	- 환경부 등 17개 부처 소관 안전기준 243개 심의·의결
5회	'18.05.04.	- 국토부 등 9개 부처 소관 안전기준 80개 심의·의결
6회	'18.06.08.	- 크레인 안전기준 개선방안 심의 ① 건설기계 조종사면허 정기적성 검사규정 신설 ② 타워크레인 정기검사 대상 부품 기준 명확화 ③ 부처간 상이한 타워크레인 고정부품 기준 통일 ④ 타워크레인 지지점 개수 통일
7회	'18.08.09.	- 국토부 등 7개 부처 소관 안전기준 157개 심의·의결
8회	'19.04.24.	- 교육부 등 16개 부처 소관 안전기준 430개 심의·의결 - 생활 속 불합리한 안전기준 사례별 개선대책 심의 ① 피난 유도등 화재안전기준 개선 방안 ② 화목보일러 안전관리 기준 강화 ③ 휴대용 부탄겐 사고감축 대책 ④ 체육시설 조명시설 등 안전관리 강화
9회	'19.09.18.	- 218개 안전기준 심의 및 의결 - 분야별 안전기준 개선방안 심의 ① 석유가스 저장시설 안전기준 개선 ② 대학연구실 안전관리 강화 ③ 신산업 연구개발 관리체계 개선 ④ 수입 전동킥보드 안전관리 강화

출처: 행정안전부 보도자료 등 발췌 및 수정

앞서 언급한바와 같이 안전기준의 규모는 방대하고 점차 전문화·복잡화되고 있지만 안전기준심의회를 통해서 안전기준 중 일부인 1,328개의 안전기준만이 논의되었다. 이에 전체 안전기준에 대해서 추출하고 분류하고, 더 나아가 통합관리 하고자 DB구축, 시스템 설계 등에 대한 연구가 진행되나 있다[4].

## 2. 안전기준 추출 및 분석 사례

안전기준을 추출하고 분류하기 위한 연구가 최근에 일부 진행되었는데 2015년 국민안전처에서는 우선적으로 안전기준 관리시스템 구축을 위해 안전기준의 코드화와 DB구축을 제안하였다[4]. 세부적으로는 안전기준을 크게 기술적 기준, 관리적 기준으로 분류하였고, 대상이 되는 법령, 조문별로 사람, 물질, 환경, 시설과의 관계성을 분류하여 설정하고 코드화하였다. 이처럼 안전기준을 코드화하여 관리하는 방법은 건설분야에서 이미 활용되고 있다. 건설기준에 대해 코드를 부여하고, 이를 기반으로 국가건설기준센터(DB)를 구축하여 제·개정 이력 등을 국민에게 공개하고 있다.

법령 내의 안전기준을 추출하고 분석한 대표적 사례

로는 국립재난안전연구원의 연구[5]를 들 수 있다. 해당 연구에서는 특정 안전사고 및 이슈(크레인사고, 살충제 계란파동, 낚시배 사고)를 선정하여 관련된 법령을 연구진이 직접 읽고, 안전기준을 추출·분석하였다. 하지만 해당 연구는 특정 시점에서의 법령을 대상으로 분석하였기 때문에 향후 법령이 제정되거나 개정될 경우 활용되지 못한다는 문제점이 있다. 또한 사람이 직접 수동으로 독해하는 방법이기 때문에 방대한 양의 모든 법령에 확장적용하기에는 한계점이 있다. 이러한 문제점을 개선하고자 빅데이터 처리 기술들을 안전기준 추출, 분석에 접목하는 연구가 일부 진행되었지만 단순 설계에 그친 연구가 대부분이었고, 실제로 안전기준 통합관리 시스템을 구축한 사례는 없었다. 따라서 본 연구에서는 일부 사례를 선정하여 실제 안전기준 자동검색을 위한 지원도구를 개발하였다.

## 3. 안전기준 추출 관련 기술

안전기준을 추출 관리시스템 개발을 위해서는 자연어 처리 기술, 법령문서 이해기술 등 다양한 기술들이 복합적으로 필요하다. 대표적으로 활용되는 기술인 자연어 처리(natural language processing)는 컴퓨터를 이용하여 인간 언어의 이해, 생성, 분석을 다루는 인공지능기술을 의미한다[6]. 특히, 자연어 처리 기술은 우리가 많이 사용하는 구글, 네이버 검색 등의 검색서비스에서 활용되고 있으며, 빅데이터 분석, 대화 인터페이스, 기계 번역 등 다양한 서비스에서도 활용되고 있다[7]. 본 연구에서 자연어 처리기술은 법령 내 존재하는 안전기준을 추출, 분류하는데 있어 전반적으로 활용되었다. 또한 법령내의 안전기준을 처리하기 위해서는 법령문서를 정제, 이해하는 기술도 별도로 필요하다. 법령문서의 경우, 맞춤법은 비교적 정확하나 소관부처별, 분야별 법령 간 통일되지 못한 띄어쓰기, 순화용어, 외래어, 의미를 혼동하여 사용하는 경우가 존재하기 때문에 법령문서를 정제하는 기술이 기본적으로 필요하다.

또한 안전기준 자동검색 지원도구를 개발하기 위해서는 안전기준 시소러스, 안전기준 온톨로지 개발도 필요하다. 시소러스(Thesaurus)에 대한 정의는 그동안 다양하게 제시되어 왔지만 대표적으로는 로제(P.M. Roget)가 Thesaurus of English Words and

Phrases란 책에서 지식의 창고(보고)라고 정의하면서 널리 소개되었다. 또한 기계에 의한 정보검색을 위하여 문헌 중의 언어와 디스크립터로 사용한 언어와의 관계로 구성된 언어를 편집한 것으로 정의하기도 한다[8]. 시소러스는 정보검색 분야에서 가장 많이 활용되는 지식베이스로서 색인 작업시에는 적절한 색인 표본의 선택과 색인어의 통제를 위해 필요하며, 검색 시에는 적절한 질의어의 선택을 위해 필요하다[9]. 일반적인 시소러스 구축과정은 크게 6단계로 구성된다. 첫 번째로는 주제영역을 설정, 두 번째로는 시소러스의 특성성 수준(색인이가 주제를 정확하게 표현할 수 있는 능력)과 조합수준 결정(복합개념을 표현하는 복합 용어를 시소러스에 어느 정도 포함시키는지에 대한 것), 세 번째로는 시소러스의 형태 결정, 네 번째로는 용어의 수집, 다섯 번째로는 용어의 표준화, 여섯 번째로는 용어의 상호관계 결정이다.

온톨로지(Ontology)는 언어로 표현된 개념간 연관관계 지식이 드러나는 망을 의미한다[10]. 일반적으로 온톨로지는 클래스에 대한 일련의 개별사례를 구성하며, 클래스는 대부분 온톨로지의 핵심요소로서 대상분야의 개념을 기술한다. 예를 들면 '안전기준' 클래스는 모든 안전기준을 대표하며, 특정 안전기준은 이 클래스의 사례에 해당한다.

온톨로지 구축과정은 크게 7단계로 구성된다. 가장먼저 대상분야와 범위를 결정하고, 기존에 구축되어 있는 온톨로지의 재활용을 검토, 온톨로지의 주요 용어를 선정 및 열거, 클래스와 클래스 계층의 정의, 클래스의 속성 정의, 슬롯(각 개념의 속성)의 패시(facet) 정의, 개별 사례를 생성하는 순으로 구축된다[9]. 1단계에서는 온톨로지가 다루는 대상분야와 용도를 명확히 하고 어떤 유형의 정보질의에 대한 응답을 제공할지, 온톨로지 이용과 관리자는 누가 되는지에 대해 결정하는 단계이다. 2단계에서는 기존 온톨로지의 재활용을 고려하는 단계로 타분야에 기구축된 온톨로지를 조사하고 확장하여 활용할 수 있는지에 대해 검토한다. 3단계에서는 온톨로지의 주요용어를 분류하고 용어들이 가지고 있는 속성을 파악한다. 4단계에서는 클래스와 클래스 계층을 정의하는 방식을 선정한다. 클래스와 클래스계층을 정의하는 방식에는 하향식, 상향식, 조합식이 있고,

온톨로지 개발자의 선호도, 구축 대상분야에 따라 방식이 정해질 수 있으며, 특정 방식이 우수하다고 정해진 것은 없다. 5단계에서는 클래스의 속성(슬롯)을 정의한다. 클래스만으로는 대상분야 및 범위 결정에 있어 적합성 질문에 응답할 수 있는 충분한 정보를 확보할 수 없기 때문이다. 마지막으로 슬롯의 패시(facet)을 정의한다. 슬롯은 값의 유형, 값의 범위, 값의 개수 등 해당 슬롯이 가질 수 있는 다른 특성을 기술하는 여러 패시를 가질 수 있는데, 예를 들어 부품의 무게 슬롯은 숫자 유형의 값을 갖는 슬롯이다.

이처럼 안전기준을 자동으로 검색하기 위한 지원도구를 개발하기 위해서는 다양한 기술을 복합적으로 활용하여야 하며 기본적으로는 언어의 특성과 활용에 관한 깊은 이해가 필요하다.

### III. 연구대상 및 방법

본 연구는 안전기준의 체계적 관리를 위해 효율적인 안전기준 검색 기술을 확보하는데 그 목적이 있다. 다만, 안전기준을 포함하고 있는 모든 법령을 대상으로 안전기준 검색시스템을 구축하기에는 시간, 인력, 비용 등의 물리적 한계가 있으므로, 특정 사고사례를 중심으로 안전기준 검색시스템을 시범구축 하고자 하였다. 특정 사고사례는 2018년 국립재난안전연구원에서 사람이 직접 관련 법령을 읽고 안전기준을 분류한 사례로 선정하였다.

연구는 크게 4가지 부분으로 구성된다(그림 2). 첫 번째는 기존에 타 연구에서 분류된 안전기준과 행정안전부에서 고시한 안전기준을 대상으로 특성을 분석한다. 두 번째는 2018년에 국립재난안전연구원에서 추출한 안전기준과 관련된 법령데이터를 수집하고, 세 번째는 수집된 법령데이터를 정제하고 형태소 분석을 실시한다. 마지막으로 이미 기존에 개발된 추출모형을 일부 활용하고, 안전기준 시소러스를 기반으로 자동검색 지원도구를 개발한다.

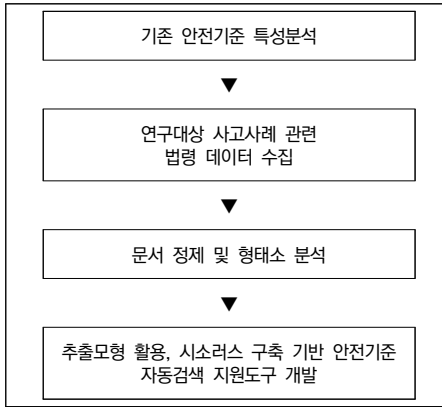


그림 2. 연구순서

#### IV. 연구결과

##### 1. 기존 안전기준 특성분석

법령내 안전기준의 특성은 일부 선행연구에서 언급된바 있다. 임수정과 박덕근(2019)의 연구[11]에서는 안전기준의 표현형태, 수록위치와 같은 구조적 속성과 강제적인 표현, 단위 및 범위를 나타내는 표현이 자주 사용되는 등의 언어적 특성에 대해 분석하였다. 선행연구에서는 언어적 특성을 중심으로 분석한 한계가 있어 본 절에서는 안전기준의 특성을 계량적으로 살펴보고자 하였다. 행정안전부에 기존 등록되어 있는 안전기준을 포함하고 있는 445개의 법령(27,065개 조문)을 대상으로 안전기준을 포함한 조문 1,108개와 일반조문 25,957개의 길이분포, 주요키워드 출현빈도를 계량적

관점에서 분석하였다.

먼저 일반조문과 안전기준을 포함한 조문의 길이를 비교한 결과, 일반조문의 평균길이 417.53자이고, 안전기준을 포함한 조문의 평균길이는 527.64자로 안전기준을 포함한 조문이 일반조문에 비하여 길다[표 2]. 하지만 조문의 길이가 매우 짧거나 긴 특정 조문이 평균값에 영향을 미칠 수 있기 때문에 단순 산술평균만을 가지고 일반조문과 안전기준을 포함한 조문의 길이를 비교하여 안전기준의 모든 특성을 분석할 수는 없다.

표 2. 일반조문과 안전기준 조문의 길이비교

구분	일반조문	안전기준 조문
최대 길이	10,552	4,170
최소 길이	2	16
평균 길이	417.53	527.64
중간값	295	361

일반조문과 안전기준을 포함하고 있는 조문의 길이 분포를 살펴보기 위해 히스토그램을 통해 그 분포를 표현하였다(그림 3). 일반조문과 안전기준을 포함하고 있는 조문의 길이 분포 히스토그램을 살펴보면 안전기준을 포함한 조문은 약 4,000자 이내로 분포하고 있으며, 일반조문의 경우는 4,000자를 초과하는 조문도 많을뿐더러 최대 10,000자를 초과하는 조문도 있는 것으로 나타났다. 히스토그램 상으로는 안전기준을 포함한 조문의 길이가 일반조문보다 짧을 수 있는 것으로 파악할 수 있다. 이러한 분석결과는 안전기준을 포함한 조문과 일반조문을 분류하는 데에 중요한 단서가 될 수 있다.

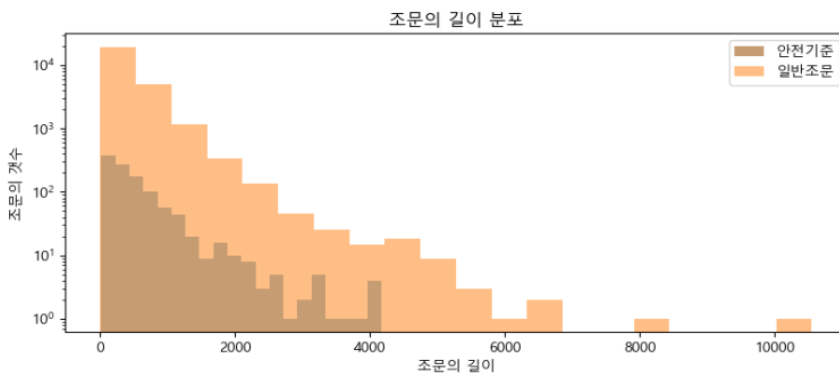


그림 3. 일반조문과 안전기준을 포함한 조문의 길이 분포

안전기준을 포함한 조문과 일반조문에서 나타나는 주요 키워드 파악을 위해 'WordRank'를 이용하였다. WordRank는 부분문자열로 구성된 그래프를 만든 후, 그래프 순위화 알고리즘을 학습시키는 방법으로 주요 키워드를 추출한다. 주요키워드 분석을 실시하여 키워드 중요도에 따라 20위까지 정리하고, 워드클라우드로 표현하였다[표 3][그림 4].

표 3. 일반조문과 안전기준을 포함한 조문의 키워드 출현빈도

순위	일반조문		안전기준을 포함한 조문	
	키워드	점수	키워드	점수
1	따른	1441.51	따른	138.72
2	또는	845.97	또는	102.89
3	따라	768.49	따라	86.74
4	한다	463.04	한다	79.64
5	경우	360.22	정하는	61.78
6	정하는	277.86	경우	52.46
7	이하	265.39	20	38.06
8	20	264.76	이하	32.49
9	제1	232.79	해당	29.59
10	해당	229.23	개정	28.68
11	관한	222.93	설치	25.58
12	개정	211.72	제1	23.47
13	별지	183.77	있는	22.53
14	제2	153.52	위하여	21.61
15	다음	149.79	다음	19.84
16	위하여	144.03	필요한	19.12
17	필요한	135.42	이상	18.76
18	대한	129.50	다만	17.33
19	받은	128.89	관한	17.19
20	사항	117.25	하는	15.82

분석결과, 안전기준을 포함한 조문과 일반조문의 뚜렷한 특징이 보이지 않는 것으로 나타났다. 따라서 단편적으로 키워드 출현을 고려하여 안전기준을 분류하고 추출하기에는 많은 한계점이 있으며, 안전기준의 분류를 위해서는 안전기준을 포함하고 있는 조문과 일반조문을 구분하게 하는 추가적인 특징을 찾고, 이를 활용한 기계학습 기반의 문서분류 모형을 개발하여야 한다.

## 2. 연구대상 사고사례 관련 법령 수집

본 연구대상으로 설정한 3가지(살충제 계란 파동, 크

레인 사고, 낚싯배 사고) 사례에 관한 법령 문서를 국립재난안전연구원(2018)의 연구[5]를 참고하여 국가법령정보센터를 통해 수집하였다. 살충제 계란파동 관련 법령문서는 1,157개, 크레인사고 관련 법령 문서는 1,105개, 낚싯배 사고 관련 법령 문서는 799개로 총 3,016개의 법령 문서를 수집하였다.

## 3. 법령 문서정제 및 형태소 분석

본 절에서는 안전기준 자동검색 지원도구를 위한 기초작업으로 법령 문서를 정제하고, 형태소 분석을 실시하였다. 기존에 개발된 맞춤법/문법 검사기, 형태소 분석기, 구문 분석기 등 자연어 처리도구를 활용하고, 필요시 별도로 수작업도 병행하였다.

가장 먼저, 수집된 법령문서는 한글문서(.hwp)로 작성된 파일이기 때문에 본 연구에서 개발하고자 하는 안전기준 자동검색 지원도구에 적용하기 위해서는 텍스트문서(.txt)로 변환해야 한다. 이에 본 연구진은 ㈜한글과컴퓨터의 기술지원을 통해 텍스트문서(.txt)로 일괄 변환하였다[그림 5].

또한 안전기준 검색 시스템에서는 검색결과를 법령을 문서단위로 반환할 뿐만 아니라 법령 문서의 조/호/항 단위로 반환하는 것도 필요하다. 따라서 개별 법령문서를 조/호/항 단위로 구분할 수 있는 메타 태그(meta tag)를 법령 문서에 부착하였다[그림 6].

수집된 법령에 대해 형태소 분석을 실시하였다. 형태소 분석단계에서는 중의성이 있는 단어(어절)에 대해 그 어절내의 형태소 분리 위치와 그 형태소들의 품사(범주)에 따라 가능한 모든 분석 후보를 생성한다[그림 7].

법령 문서의 정제를 위해 기존의 한국어 맞춤법/문법 검사기를 활용하여 수집한 법령 문서에 대해서 검사를 실행하고, 그 결과를 수작업을 통해 검토하였다. 한국어 맞춤법/문법 검사기는 한국어 문서에 잘못 쓰인 단어나 표현을 찾아서 대치어와 도움말 등 교정정보를 제공하는 시스템이다[그림 8]. 검토결과에 따라 미등록어를 형태소 분석 사전에 추가하고, 맞춤법, 문법 교정 규칙을 수정하였다. 맞춤법, 문법 검사오류의 유형으로는 철자 오류(띄어쓰기, 오타, 외래어 표기법 위반), 의미 문체 오류(문맥의존 철자오류, 문장부호 오류 등)가 있고, 특



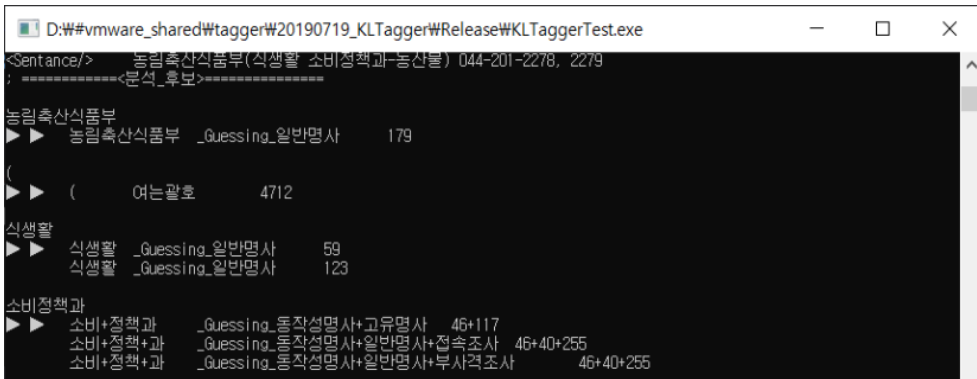




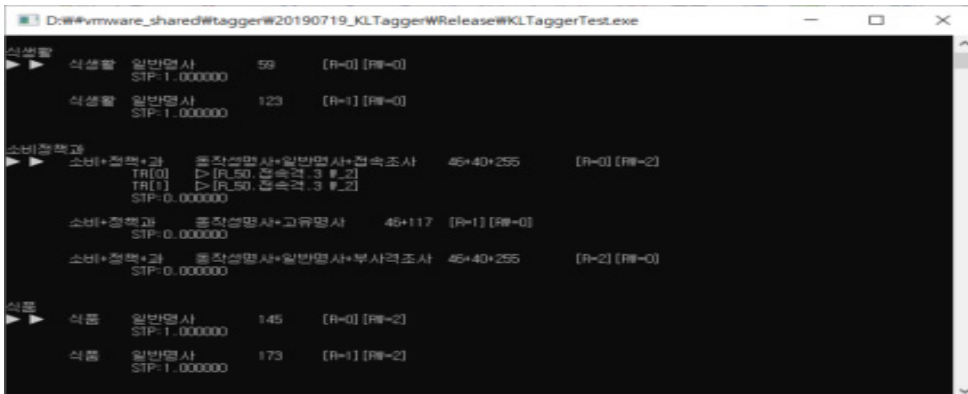




그림 8. 한국어 맞춤법·문법 검사기(산업안전보건법 시행규칙 적용예시)



규칙수정 전



규칙수정 후

그림 9. 한국어 품사 태거의 실행결과 일부

히 법령 문서의 경우는 띄어쓰기 오류로 판단된 경우가 가장 빈번한 것으로 나타났다.

형태소 분석과정에서 발생하는 품사 중의성과 어휘 중의성을 해소하기 위해 품사 태깅을 수행하였다. 기본적으로 기존에 개발된 한국어 품사 태거(Part of speech tagger)를 활용하였고, 고도화를 위해 수집한 법령 문서에 대하여 자동검사를 수행한 다음 수작업 검수를 진행하였다. 일부 품사 태깅 오류 중 고유명사나 전문용어의 미등록 오류는 앞서 실시된 맞춤법 및 문법 검사기의 고도화 과정에서 형태 분석 사전의 보완으로 해결하였고, 나머지 품사 태깅 오류는 관련 규칙을 수정하거나 통제정보를 보완하는 방식으로 수정하였다 [그림 9]. 품사태거는 향후 법령 문서 검색을 위한 색인어 추출에 활용될 수 있다.

#### 4. 자동검색 지원도구 개발

##### 4.1 안전기준 추출모형 차용 및 활용

본 연구에서 안전기준을 분류하기 위해 기계독해 기반으로 안전기준 추출모형을 개발한 김민호 등(2020)의 연구[11]를 활용하였다. 해당 연구에서는 BERT(Bidirectional Encoder Representations from Transformers) 기반 전이학습을 이용하여 안전기준 분류모형을 구축하였다. 이 모듈을 활용하여 1단계로 전체 법령 문서 중에서 질문과 연관성이 큰 법령 문서 100개를 검색하고, 2단계에서는 질문과 연관성이 큰 조문들을 선별하였다. 본 연구에서는 김민호 등(2020)의 연구[12]에서 구축한 안전기준 검색과 분석 모듈을 활용하여 안전기준 검색 지원시스템을 구축하였다. 세부적으로는 안전기준 자동검색 지원도구에서 사용자가 입력한 키워드에 해당하는 법령을 보여줌과 동시에 조문단위로 표시하며 안전기준 여부를 함께 표시되게 하는 부분에 활용되었다.

##### 4.2 안전기준 시소러스 구축

안전기준 검색분석 시스템에서 특정 분야, 대상물과 관련된 법령 내 안전기준을 검색하기 위해서는 별도의 안전기준 시소러스를 구축해야 한다. 시소러스를 구축하기 위해서는 가장 기본적으로 문장을 구성 성분으로 분해하고 이들 사이의 관계를 분석하여 구조를 결정짓

는 구문분석 과정이 필요하다. 또한 법령에서는 동일한 대상물에 대해서도 다르게 표현하는 경우도 있기 때문에 같은 뜻을 가지는 동의어 집합(synonym set, 이하 신셋) 간의 관계들로 이루어진 어휘의미망도 필요하다.

가장 먼저, 문장을 구성 성분으로 분해하고 이들 사이의 관계를 분석하여 구조 결정을 위해 본 연구진이 기존에 개발한 구문분석기 'KLParse'를 활용하여 구문분석을 실시하였다. 구문분석기는 문장 성분간 의존관계 생성 규칙, 통사적 제약 규칙 및 선택적 제약 규칙을 통해 의존관계 중의성을 해소하는 시스템이다. 일반적으로 구문분석은 전처리, 형태적 중의성 제거, 통사적 중의성제거, 구문분석 트리 후보의 순위화의 순서로 진행된다[그림 10].

법령내에 존재하는 안전기준에서 사용하는 단어는 일상생활에서 사용하는 단어의 정의와 다르기 때문에 별도의 어휘의미망 구축이 필요하다. 예를 들면 승강기, 리프트는 표준국어대사전에서는 동일한 의미로 사용되고 있지만, 법령에서는 구분하여 사용하고 있다. 또한 앞서 법령 문서 분석을 통해 도출된 미등록어를 기존에 개발·사용되고 있는 한국어 어휘의미망인 한국어 워드넷(Korean WordNet) KorLex 1.5에 추가하여 안전기준 어휘의미망을 구축하였다. 한국어 워드넷 KorLex 1.5는 명사, 동사, 형용사, 부사, 분류사로 구성되며, 약 13만개의 신셋과 약 15만개의 어의를 포함하고 있다[그림 11].

##### 4.3 색인 작업 및 안전기준 검색기 개발

법령 문서의 색인 작업을 위해 원본 문서를 아래와 같이 색인기의 입력 파일 형식인 txt로 변환하였다[그림 12]. 우선 JavaScript, JQuery, CSS, HTML5를 이용하여 웹 사이트의 UI를 개발하고, 검색 시스템 등과의 연동작업을 통해 안전기준 검색을 위한 시스템을 구축하였다. [그림 13]은 안전기준 검색 시스템을 통해 법령 내 일부 안전기준을 검색한 결과이다.

안전기준 자동검색 지원도구의 성능을 평가하였다. 학습데이터와 평가데이터가 같은 도메인일 경우와 그렇지 않은 경우의 성능 차이가 다소 있었다. 평가데이터를 8:2로 나눠 80%로 학습을 하고 20%로 평가를 하였을 때, 53%의 정확도와 84%의 재현율을 보였다. 그

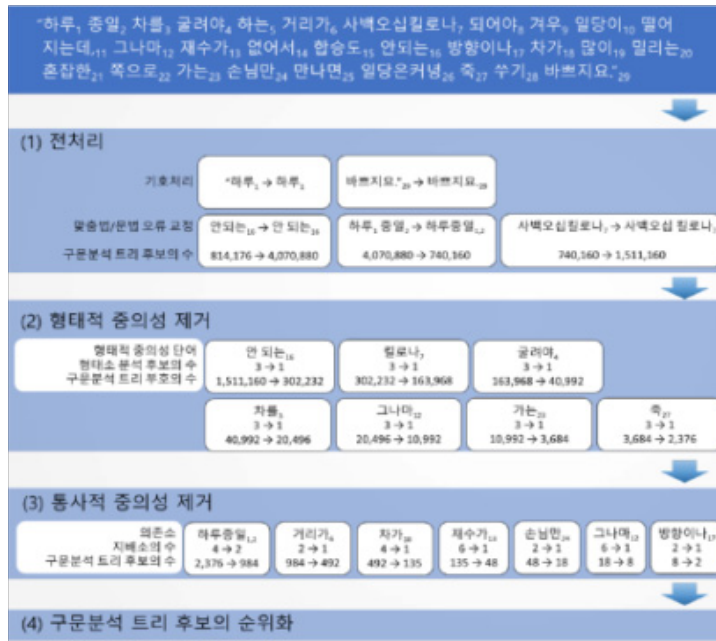


그림 10. 구문분석 전체 과정

## 다리 2, 교량 1

bridge 1, span 5 (a structure that allows people or vehicles to cross an obstacle such as a river or canal or railway etc.)

	KORLEX	WORDNET
상위개념	<ul style="list-style-type: none"> <li>▣삼체 1, 거체 1</li> <li>▣물체 1, 물건 1</li> <li>▣송체 1, 간체 1</li> <li>▣가공물 1, 가공품 1, 인공물 1</li> <li>▣건조물 1, 구조물 1, 구조 2</li> <li>▣다리 2, 교량 1</li> </ul>	<ul style="list-style-type: none"> <li>배일리식비상용조립교 1</li> <li>칸틸러버식다리 1</li> <li>가축발출방지용다리 1</li> <li>유개교 1</li> <li>승개교 1, 도개교 1, 가동교 1</li> <li>육교 1</li> <li>고가교 1</li> <li>부교 1, 배다리 2, 선창 3</li> <li>밧줄다리 1</li> <li>강철아치교 1</li> </ul>
하위개념	<ul style="list-style-type: none"> <li>Bailey bridge 1</li> <li>cantilever bridge 1</li> <li>cattle guard 1, cattle grid 1</li> <li>covered bridge 1</li> <li>drawbridge 1, lift bridge 1</li> <li>footbridge 1, overcrossing 1, pedestrian bridge 1</li> <li>overpass 1, flyover 1, flypast 1</li> <li>pontoon bridge 1, bateau bridge 1, floating bridge 1</li> <li>rope bridge 1</li> <li>steel arch bridge 1</li> </ul>	

그림 11. 한국어 워드넷(출처: <http://corpus.pusan.ac.kr>)

Tags	Description
<&s>	Document Start Tag
<i> 10 </i>	Document Number
<u> www.website.com/directory/ </u>	Document Inlink (URL)
<t> TITLE – titles [...] </t>	Document Title
<c> Contents [...] </c>	Document Contents
<l> \$http://URL/dir/file\$ [...] </l>	Document Outlink
<g> \$http://URL/dir/filename.gif\$ [...] </g>	Images <g>, <j>, <o> ..
<e>	Document End Tag

그림 12. ptxt의 tag 명세표



그림 13. 지능형 안전기준 관리도구의 활용

러나 학습데이터를 학습하고 평가데이터 전체로 평가를 하였을 때는 46%의 정확도와 63%의 재현율을 보였다[표 4]. 이를 통해 학습데이터가 안전기준 추출 모형을 일반화시킬 수 있을 만큼의 크기가 되지 않음을 알 수 있으며, 추후 학습데이터가 늘어난다면 성능 향상을 기대할 수 있다.

표 4. 안전기준 자동검색 지원도구의 성능

측정방법	검증방법	달성치	
		정확도	재현율
학습데이터와 평가데이터를 같은 도메인인 경우의 정확도/재현율	평가데이터	53%	84%
학습데이터와 평가데이터가 다른 도메인인 경우의 정확도/재현율	평가데이터	46%	63%

## V. 결론 및 고찰

현재 법령의 규모는 점차 증가하고 있고 분야도 세분화되어 가고 있어 법령에 포함된 안전기준의 내용도 더욱 복잡해지고 있다. 따라서 법령 내 안전기준을 추출하고 관리하는 일은 대규모 예산과 전문인력을 필요로 한다. 하지만 법령은 끊임없이 제·개정을 반복하면서 실시간으로 변화하고 있기 때문에 특정 시점의 법령을 대상으로 안전기준을 추출하고 분석하기보다는 법령의 제·개정에 맞춰 실시간으로 추출하고 분석하여야 한다. 이러한 이유로 사람이 직접 법령을 읽고 안전기준을 추출하는 것은 비효율적인 일이다. 따라서 본 연구에서는 안전기준을 효율적, 실시간으로 검색하고 추출하기 위한 안전기준 자동검색 지원도구를 시범적으로 구축하였다. 본 연구를 통해 안전기준 검색을 위한 문서정제 기술, 형태분석 기술, 구문분석 기술 등을 확보하였다. 각각의 기술을 새로 개발하는 것은 큰 비용이 소모되므로 기존에 개발되어진 기술을 기반으로 법령문서에 적용하도록 일부 수정 및 보완하여 활용하였다.

본 연구의 특징은 두 가지로 정리할 수 있다. 첫째, 별도의 안전기준 어휘의미망 구축으로 동일 대상에 대해서 다르게 표현한 경우도 검색이 가능하기 때문에 특정 대상물과 관련된 모든 안전기준을 추출할 수 있다. 기존의 특정분야에서 법령정보를 다루는 시스템으로

한국건축규정 e시스템, 국제법령정보시스템 등이 구축된 사례가 있지만 법제처 법령정보시스템처럼 법령본문에 특정 용어가 포함되는 지에 대한 단순검색 기능만을 제공하고 있다. 따라서 기존의 법령정보를 다루는 시스템들은 법령 내에서 동일 대상을 서로 다르게 표현한 경우는 단순검색으로 찾아내지 못하는 한계점이 있다. 예를 들면 '바나나'라는 단어를 검색하면 '바나나'가 포함된 문장만 추출이 가능하고 '바나나'를 포함하고 있는 '과일'이라는 단어가 포함된 문장은 검색되지 않는 한계점이 있었다. 이에 본 연구에서는 안전기준 어휘의 미망을 별도로 구축하여 동일대상에 대해서도 다르게 표현한 경우도 검색이 가능하도록 하여 특정대상물과 관련된 모든 안전기준의 검색이 가능하도록 하였다. 둘째, 법제도 연구관점에서 특정 사례에 국한된 단편적인 연구가 아니라 향후 타 분야 및 사례에 적용하여 연구 범위를 확장할 수 있도록 그 기초를 마련했다는 특징이 있다. 법제도측면에서의 대부분의 연구는 특정현안 및 분야에 국한되어 일회성에 그친 연구가 대부분이었다 [13-15]. 하지만 본 연구에서 시범적으로 구축한 안전기준 자동검색 지원도구는 대상범위를 모든 법령, 행정규칙으로 확장하여 적용가능하다.

본 연구의 결과로 안전기준 관리에 필요한 시간과 비용의 절감이 가능하며 향후 인공지능 개념을 도입한 안전기준의 중복, 상충, 미비의 문제점을 찾아낼 수 있는 분석시스템 개발의 기초가 될 수 있다. 특히, 안전기준으로 인한 사고방지도 일조할 수 있을 것이다. 다만, 연구대상을 특정 사고사례 관련 법령으로 대상을 제한하여 우선 연구를 진행하였기 때문에 향후에는 모든 법령 뿐만 아니라 하위 행정규칙을 대상으로 적용범위를 확장하여 시스템을 구축할 필요가 있다.

### 참 고 문 헌

- [1] 국립재난안전연구원, *해외사례기반 국내 안전기준 분석 및 사전정비 기술개발*, 2019.
- [2] [https://www.mois.go.kr/firt/bbs/type010/comm onSelectBoardArticle.do?bbsId=BBSMSTR\\_000000000008&ntId=59414](https://www.mois.go.kr/firt/bbs/type010/comm onSelectBoardArticle.do?bbsId=BBSMSTR_000000000008&ntId=59414), 2020.04.09.
- [3] 임수정, 박덕근, "규모파악을 위한 국내 안전기준현황

- 예비조사 및 대표 문제사례 분석,” 한국방재학회논문지, 제18권, 제6호, pp.111-122, 2018.
- [4] 국민안전처, *안전기준의 통합적 관리운영체계 구축방안 연구*, 2015.
- [5] 국립재난안전연구원, *안전사고 사례기반 안전기준 분석 현업화 지원기술 개발*, 2018.
- [6] [https://opendict.korean.go.kr/dictionary/view?se\\_nse\\_no=740319&viewType=confirm](https://opendict.korean.go.kr/dictionary/view?se_nse_no=740319&viewType=confirm), 2020.04.05.
- [7] 김경선, “인공지능 기반 언어 처리 기술,” 새국어 생활, 제27권, 제4호, pp.9-37, 2017.
- [8] [https://www.kla.kr/jsp/fileboard/termdic.do?procType=view&f\\_termdic\\_seq=3662](https://www.kla.kr/jsp/fileboard/termdic.do?procType=view&f_termdic_seq=3662), 2020.03.20.
- [9] 국립재난안전연구원, *안전기준 추출분석을 위한 프로세스 설계연구*, 2016.
- [10] <https://stdict.korean.go.kr/search/searchView.do>, 2020.04.05.
- [11] 임수정, 박덕근, “구조적 속성과 어휘적 특징에 기반한 안전기준 고찰,” 한국융합학회논문지, 제10권, 제11호, pp.353-366, 2019.
- [12] 김민호, 조상현, 박덕근, 권혁철, “안전기준의 검색과 분석을 위한 기계독해 기반 질의응답 시스템,” 멀티미디어학회 논문지, 제23권, 제2호, pp.351-360, 2020.
- [13] 최봉문, “도시공간계획을 지원하는 정보시스템에 관한 법제도 실태연구,” 한국콘텐츠학회논문지, 제11권, 제1호, pp.396-403, 2011.
- [14] 노재철, 고준기, “독거노인에 대한 지원정책의 현황과 문제점과 법제도적 개선방안,” 한국콘텐츠학회논문지, 제13권, 제1호, pp.257-268, 2013.
- [15] 최봉문, 조병호, 박환용, “고령자 중심의 헬스케어 도입을 위한 관련 법제도 개선방안,” 한국콘텐츠학회논문지, 제13권, 제7호, pp.203-213, 2013.

저 자 소 개

임 수 정(Sujung Im)

정회원



- 2015년 2월 : 인하대학교 산업공학과(공학박사)
- 2016년 3월 ~ 현재 : 국립재난안전연구원 책임연구원

〈관심분야〉 : 안전기준, 재난안전산업 등

박 덕 근(Dugkeun Park)

정회원



- 1992년 12월 : 플로리다대학교 토목공학과(공학석사)
- 1998년 2월 : 코넬대학교 토목환경공학과(공학박사)
- 1998년 3월 ~ 현재 : 국립재난안전연구원 연구관

〈관심분야〉 : 재난관리, 재난안전산업, 안전기준 등