

Research Paper

## 빅데이터를 활용한 건축물 화재위험도 평가 지표 결정

# Determination of Fire Risk Assessment Indicators for Building using Big Data

주홍준<sup>1</sup> · 최윤정<sup>2</sup> · 옥치열<sup>3</sup> · 안재홍<sup>4\*</sup>

Joo, Hong-Jun<sup>1</sup> · Choi, Yun-Jeong<sup>2</sup> · Ok, Chi-Yeol<sup>3</sup> · An, Jae-Hong<sup>4\*</sup>

<sup>1</sup>Researcher, Department of Construction Test & Assessment Center, Korea Institute of Civil Engineering and Building Technology, Ilsan, 10223, Korea

<sup>2</sup>Senior Researcher, Department of Construction Test & Assessment Center, Korea Institute of Civil Engineering and Building Technology, Ilsan, 10223, Korea

<sup>3</sup>Researcher Fellow, Department of Construction Test & Assessment Center, Korea Institute of Civil Engineering and Building Technology, Ilsan, 10223, Korea

<sup>4</sup>Senior Researcher, Department of Construction Test & Assessment Center, Korea Institute of Civil Engineering and Building Technology, Ilsan, 10223, Korea

\*Corresponding author

An, Jae-Hong

Tel : 82-31-995-0811

E-mail : rehong@kict.re.kr

Received : April 30, 2022

Revised : May 16, 2022

Accepted : May 26, 2022

### ABSTRACT

This study attempts to use big data to determine the indicators necessary for a fire risk assessment of buildings. Because most of the causes affecting the fire risk of buildings are fixed as indicators considering only the building itself, previously only limited and subjective assessment has been performed. Therefore, if various internal and external indicators can be considered using big data, effective measures can be taken to reduce the fire risk of buildings. To collect the data necessary to determine indicators, a query language was first selected, and professional literature was collected in the form of unstructured data using a web crawling technique. To collect the words in the literature, pre-processing was performed such as user dictionary registration, duplicate literature, and stopwords. Then, through a review of previous research, words were classified into four components, and representative keywords related to risk were selected from each component. Risk-related indicators were collected through analysis of related words of representative keywords. By examining the indicators according to their selection criteria, 20 indicators could be determined. This research methodology indicates the applicability of big data analysis for establishing measures to reduce fire risk in buildings, and the determined risk indicators can be used as reference materials for assessment.

**Keywords :** building, fire risk assessment, big data

## 1. 서론

화재는 사회재난 중에서 주요한 재난으로써, 매년 전 세계적으로 많은 피해를 발생시키고 있다. 특히, 건축물에서 발생하는 화재는 재실자의 인명에 직접적인 영향을 미치고 여러 가지 기반시설을 비롯한 2, 3차적인 간접피해를 불러온다. 최근 10년간(2011~2020) 국내 화재사고 423,317건 중에서 주거시설, 산업시설 등의 건축물에서 발생하는 화재는 약 63%의 비율인 265,365건으로 가장 많은 화재발생이 발생하고 있다[1]. 이를 위해 건축물의 화재위험도에 영향을 미치는 원인을 검토하여 위험에 취약한 지표를 점검하고 그에 맞는 비구조적, 구조적 대책을 수립해야 한다.



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 non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

국외에서 건축물의 화재위험도를 평가하기 위한 지표 연구로 Mi et al.[2]은 5가지(Disaster-causing factors, Passive measures, Active measures, Property fire management, The rescue capability of fire brigade) 요소별로 건축물 화재위험도 지표를 구성하였다. NFPA[3]에서는 병원의 화재위험도를 평가하기 위해 3가지(Fire intervention, Hospital protection, Interface between fire brigade and hospital) 요소별로 지표를 구성하였다. Bryant[4]는 건축물, 소유자, 점유자 측면에서 4가지(Business, Life, Environment, Property) 요소별로 지표를 구성하여 다양한 이해관계를 고려한 화재위험도 평가 전략을 수립하였다. Frantzich[5]은 거주자가 받는 개별 화재위험과 대규모 화재에 따른 사회적 위험으로 나누어 지표를 구성한바 있다. 건축물 화재위험도 지표에 관한 국내 연구는 조금씩 차이가 있지만 대부분 화재통계연보, 재난연감 등에서 제시하는 화재 원인 중 일부 지표를 선정하고 있다[6-8].

빅데이터(Big data)는 인터넷의 발전, 모바일 디바이스의 대중화, SNS의 확대 등으로 인해 디지털 정보량이 기하급수적으로 증가함에 따라 2010년경에 등장한 용어로 불과 몇 년 사이에 전 세계 여러 국가들이 관심을 갖는 중요한 기술로 부상하였다[9]. 빅데이터 기술이 건축물 화재 위험 관리에 활용된 사례로는, 미국 뉴욕시 소방청(New York City Fire Department, FDNY)에서는 빅데이터와 AI 기술을 이용한 ‘Fire cast 2.0’ 모델을 통해 건축물의 화재위험도점수(fire risk score)를 산출하여 화재를 위한 대비책을 수립하고 있다. 활용 데이터(지표)는 건축물 구조, 재질, 위치와 더불어 건축물별 이격거리, 건축물 특성에 따른 인명 사망, 피해 금액 등이 고려되었으나, 뉴욕시의 인력 및 예산의 한계로 현재 화재안전점검 대상인 33,000여 개의 건축물 중 10% 수준인 30,000여개의 건축물만 구축되어 있다[10]. 국내의 경우 Park and Min[11]은 화재 통계 데이터를 가공하여 빅데이터 화재예측플랫폼에 구현하였으며, Kim and Jo[12]는 화재통계 데이터를 이용하여 상관관계 빅데이터 분석을 통해 화재현장의 위험도를 예측하는 모델을 개발하는 등 주로 정형화된 통계자료에 의존하는 연구가 수행되었다.

건축물의 화재위험도는 지표의 구성에 따라 상이한 결과가 도출될 수 있기 때문에 합리적인 검토가 필요하다. 하지만, 지표 선정 방법은 기존 연구의 지표 활용 빈도와 통계 자료에 근거하여 건축물 자체에서 발생하는 원인을 화재 위험도 지표로 고려해왔다. 따라서 지표를 확장하여 제시하기가 어렵고, 화재 위험에 영향을 미치는 지표를 주관적인 기준으로 고려했기 때문에 위험도 평가 결과에 대한 한계가 제기될 수 있다. 무엇보다 국내외 대부분의 연구는 건축물의 화재위험도를 평가하는 것에 집중되어 있으며, 위험도에 영향을 미치는 지표를 검토하는 연구는 다소 미미한 실정이다.

본 연구에서는 건축물의 내·외부 환경과 화재에 직·간접적으로 영향을 미치는 다양한 원인을 검토하여 새로운 위험도 지표를 제시하기 위해 최근 이슈가 되고 있는 빅데이터 분석 기법을 활용하여 건축물 화재위험도 지표를 결정하고자 한다. 이를 위해 질의어(query language)를 선정하고 검색 엔진을 통해 비정형 데이터(unstructured data)로 이루어진 전문 문헌을 웹 크롤링(web crawling) 기법을 이용하여 수집하였다. 그리고 사용자 용어사전 등록, 중복문헌 및 불용어 처리 등의 전처리 과정을 수행한 후, 요소를 분류하고 요소별로 대표 키워드를 추출하여 그에 대한 연관 검색어를 수집하였다. 마지막으로 연관 검색어를 대상으로 지표 결정 기준에 따라 최종적인 건축물 화재위험도 지표를 결정하였다. Figure 1은 본 연구의 과정과 흐름을 나타낸 것이다.

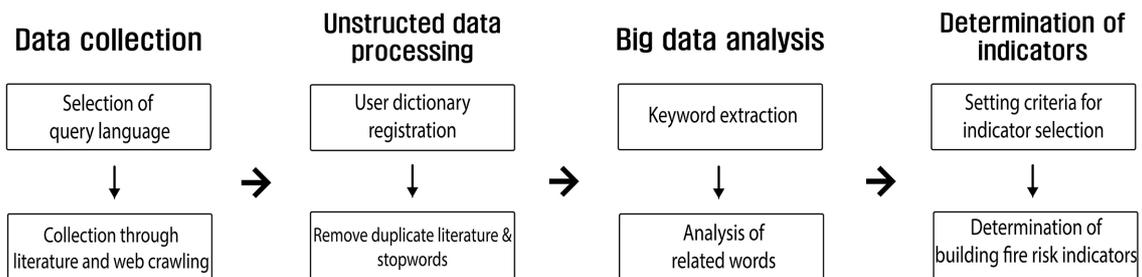


Figure 1. Research process & Flow chart

## 2. 이론적 배경 및 연구 방법론

### 2.1 빅데이터

빅데이터(Big data)에 대하여 다양한 정의가 존재하지만 일반적으로 ‘데이터를 수집, 저장, 관리, 분석하는데 있어서 기존의 처리 방식을 뛰어넘는 데이터의 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술’을 의미한다[13]. 빅데이터의 특징은 방대한 용량의 데이터 양(volume), 다양한 형태(variety), 빠른 생성 속도(velocity)인 3V로 표현되고 있으며, 최근에는 가치(value), 가치성(visibility) 등의 다양한 V가 추가되면서 특징이 계속 확대되고 있다. 이러한 빅데이터 기술을 통해 기상 및 기후, 전염병의 발병, 범죄율 등의 사회, 자연, 생활 재난 등의 리스크를 예측할 수 있다. 뿐만 아니라 의료, 제조, 통신, 산업 등 다양한 분야에서 새로운 시각이나 복잡한 법칙을 발견하기 위해서 빅데이터 기술이 요구되고 있기 때문에, 현재 4차산업시대를 견인하는 필수 기술로 자리매김하였다.

빅데이터는 자료 종류에 따라 크게 정형 데이터(Structured data)와 비정형 데이터(Unstructured data), 그리고 반정형 데이터(Semi-structured data)로 구분된다. 정형 데이터는 데이터베이스의 정해진 규칙에 따라 수치적으로 정의할 수 있으나, 비정형 데이터는 이와 반대로 정해진 규칙이 없고 수치적으로 정의할 수 없는 데이터를 의미한다. 따라서, 반정형 데이터는 정형 및 비정형 데이터가 정확히 구분되지 않고 혼합된 데이터로 이해할 수 있다. Figure 2는 빅데이터의 종류와 예시를 보여주며, 본 연구에서는 웹(Web)상에 존재하는 텍스트(비정형 데이터) 형식의 전문 문헌을 수집 및 분석하여 건축물 화재위험도 평가 지표를 결정하였다.

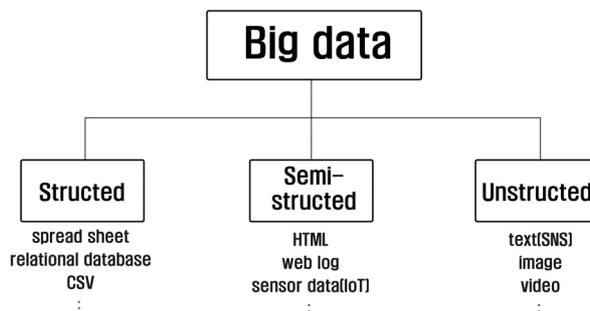


Figure 2. Big data types & examples

### 2.2 웹 크롤링

웹 크롤링(Web Crawling) 또는 웹 스크래핑(Web Scraping)은 컴퓨터 소프트웨어 기술 등을 통해 웹 사이트에서 원하는 정보를 추출하는 것을 의미한다[14]. 뉴스 및 전문 자료와 같은 사실에 근거한 객관적인 데이터는 정확하고 신뢰적인 분석 결과로 이어진다. 이러한 자료는 웹이라는 공간에 방대하고 복잡하게 혼재되어 있지만, 원하는 자료를 파일 등과 같은 정리된 형태로 수집하기에는 인력적, 물리적인 시간 측면에서 한계가 있다. 따라서, 빅데이터를 구축하기 위해서는 웹 자료를 효과적으로 수집할 수 있는 자동화된 동적(dynamic) 웹 크롤링 기법이 필수적이라 할 수 있다.

### 2.3 사용자 용어사전 등록

수집된 비정형 데이터가 찾고자하는 용어, 즉 질의어(query language)와 면밀한 의미를 갖기 위해서는 용어들에 대한 기준이 필요하다. 기준 용어의 누락은 왜곡된 데이터가 수집되는 문제가 발생할 수 있다. 예를 들어, 본래의 ‘건축물화재’와

‘내화구조’ 등의 의미가 각각 ‘건축물’ + ‘화재’, ‘내화’ + ‘구조’ 등으로 분리되어 형태소 단위의 의미로 과분절화가 이루어질 수 있다. 따라서, 복합명사와 전문용어 등을 포함하는 사용자 용어사전을 별도로 구축하여 분석에서 의미를 갖는 단어가 누락되지 않도록 해야한다[15]. 이를 통해, 수집 데이터의 의미를 명확하게 하고 분석 결과의 신뢰성을 확보하며 의도치 않는 불용어가 발생하는 현상을 사전에 방지할 수 있다.

## 2.4 불용어 처리

비정형 데이터 중에서 분석 방향과 상이하거나 정보로서의 가치가 없는 단어가 존재하는데 이를 불용어(stopword)라 한다[16]. 불용어는 문헌에서 출현 빈도가 높으나 의미가 없거나 반복되는 경우가 많으며, 데이터 분석 결과에 악영향을 미친다. 불용어의 처리를 위해서는 동의어와 유의어에 대한 기준 검토가 필요하며, 표준화된 언어사전 등을 통해 분석에 가치가 없는 용어를 제거해야한다. Table 1은 조사, 지시어 등 대표적으로 간주되고 있는 불용어를 나타낸다[17].

**Table 1.** Examples of stopwords

Type	Stopwords
Determiners	a, the, etc.
Conjunctions	is, am, are, etc.
Pronouns	he, him, this that, they, them, etc.
Interrogative	what, where, when, how, etc.
Auxiliary	for, of, etc.
Prepositions	from, at, on, in, to, etc.

## 2.5 연관 검색어 분석

연관 검색어는 이용자가 검색한 용어를 파악하여 관련된 검색어를 제공하는 확장 서비스이다. 연관 검색어 서비스의 활용은 용어에 대한 직접적인 정보뿐만 아니라 다양한 상세 키워드를 제공 받음으로써 다각적인 차원에서 정보 접근이 가능하다[18]. 예를 들어 웹 검색 엔진을 통해 ‘내화’라는 단어를 검색하면 내화구조, 내화충전구조, 내화충진재, 층간방화, 방화실리콘, 방화폼, 방화실란트, 층간방화재, 차열재 등 다양한 연관 검색어를 제공한다. 이러한 상세 키워드를 통해 중요한 정보를 획득하고 건축물 화재위험도를 평가하는 지표를 검토 및 선정하는데 활용할 수 있다.

## 3. 전문 문헌 수집 및 전처리

본 연구에서는 분석에 필요한 질의어를 선정하고 웹 크롤링 기법을 이용하여 학술자료 검색 사이트에서 화재와 관련된 전문 문헌 등의 비정형화된 빅데이터를 수집하였다. 수집된 데이터는 분석을 위한 형태가 아니기 때문에 일관되고 구조화된 형태로 변환해야 한다. 이를 위해 사용자 용어사전을 등록하고 중복 문헌 및 불용어 처리하는 일련의 데이터 전처리를 수행하였다.

### 3.1 질의어 선정

검색을 통해 수집되는 데이터는 질의어의 선택에 따라 정보의 크기 및 범위가 달라지기 때문에 위험도 지표 결과에도 영향을 미칠 수 있다. 따라서 객관성을 확보하고 범용적인 기준이 된다고 판단되는 화재 관련 법령의 단어를 검토하였다. 현재

건축물 화재와 관련된 주요 법령은 ‘건축법’과 ‘소방시설법’이 존재하며, 추가적으로 건축물 방화 규칙을 다루고 있는 ‘건축물방화구조규칙’과 재난 관련 법령인 ‘재난안전법’을 모두 검토하여 아래와 같은 질의어를 선정하였다.

- 건축법: 화재, 방화, 내화, 안전, 예방, 관리
- 화재예방, 소방시설 설치·유지 및 안전관리에 관한 법률(소방시설법): 화재, 방화, 안전, 예방, 대비, 대응, 관리
- 건축물의 피난·방화 구조에 관한 규칙(건축물방화구조규칙): 화재, 방화, 내화, 피난, 안전
- 재난 및 안전관리 기본법(재난안전법): 화재, 안전, 예방, 대비, 대응, 복구, 관리

### 3.2 전문 문헌 검색 및 수집

화재위험도에 관한 비정형 빅데이터를 수집하기 위해 전문 문헌(자료)을 제공해줄 수 있는 학술검색 사이트를 이용하였다. 학술검색 사이트별로 제공하는 문헌의 형태가 다르고 존재 유무 또한 다르기 때문에, 다양한 사이트에서 전문 문헌을 수집할 필요가 있다. 예를 들어, 어느 한 문헌이 A사이트에서는 제공되지만 B사이트에서는 제공되지 않을 수 있으며, C사이트에서는 자료 검색은 되지만 원문 보기 등의 정보 획득이 제한될 수 있다. 따라서, 전문 문헌의 접근성 및 활용도가 높다고 판단되는 3개의 사이트(RISS, KISS, DBPIA)를 선정하였으며, 질의어와 관련된 보고서, 학술지, 특허 기술 및 단행본 등을 검색하였다(Table 2).

사이트별로 검색된 전문 문헌을 살펴보면 RISS(학술연구정보서비스)는 11,380건, KISS(학술정보연구서비스)는 18,930건이 검색되었다. 학술콘텐츠플랫폼(DBPIA)은 학술지 논문만이 제공됨에 따라 184건의 가장 적은 문헌이 검색되었다. 따라서, 3개의 학술 검색 사이트에서 총 30,494건의 전문 문헌이 검색되었으며, 비정형 데이터로 구성된 각 문헌의 주요 내용을 수집하기 위해 웹 크롤링 기법을 이용하여 제목, 저자, 핵심어, 초록 순으로 정리하였다. 그 결과 Table 3과 같이 원문 보기, 핵심어, 초록 등이 제공되지 않아 정보 수집이 어려운 문헌을 제외한 총 9,768건의 문헌을 수집할 수 있었다.

**Table 2.** Search and collection of professional literature

Search & Collect data	Web site for academic research			Total
	RISS	KISS	DBPIA	
Search	11,380	18,930	184	30,494
Collect	4,102	5,574	92	9,768

### 3.3 비정형 데이터 전처리

#### 3.3.1 사용자 용어사전 등록

화재와 관련된 용어를 포함한 복합명사 등을 구분하기 위해서는 일반적으로 국립국어원의 ‘표준국어대사전’에서 제공하는 용어를 활용하고 있다. 하지만 표준국어대사전은 화재 및 재난 관련과 같은 전문 용어가 제공되지 않기 때문에, 전문 용어를 포함하는 사용자 용어사전을 구성하기가 어렵다. 따라서, 본 연구에서는 대한건축학회 및 한국화재보험협회에서 제공하는 ‘건축용어사전’과 ‘방재용어사전’을 분석도구에 포함하는 사용자 용어사전을 등록하였다.

#### 3.3.2 중복 문헌 및 불용어 처리

빅데이터 분석에서 의미있는 결과를 도출하기 위해서는 수집된 데이터에서 중복 문헌 및 불용어를 제거해야 한다. 각 전문 학술사이트에서는 내용이 동일한 문헌이 중복되어 제공될 수 있다. 따라서, 특정 용어가 고빈도화가 되는 현상을 방지하고자 중복 문헌 및 의미와 내용이 유사한 문헌인 7,556건을 제거하였으며, 중복되지 않은 고유한 2,505건의 문헌을 대상으로

로 불용어 처리를 수행하였다.

일반적으로 조사나 접속사를 포함한 의미없는 고빈도어를 불용어로 정의할 수 있으나, 현재 한국어 불용어는 별도의 기준이 존재하지 않는다. 이에 대한 차선택으로 국립국어원의 ‘모두의 말뭉치’와 RANKS NL[19]에서 제공하는 불용어 리스트를 참고하여 ‘은’, ‘는’, ‘이’, ‘가’ 등을 포함한 불용어 675개를 제거하였으며, 추가적으로 학술지와 보고서 등에서 의미 없이 쓰이고 있는 ‘본’, ‘연구’, ‘논문’ 등의 불용어들도 제거하였다.

#### 4. 빅데이터 분석 결과를 통한 건축물 화재위험도 지표 결정

본 연구에서는 빅데이터의 분석을 통해 건축물의 화재위험도에 영향을 미치는 지표를 결정하였다. 이를 위해, 수집된 비정형 데이터를 대상으로 의미가 유사한 지표 중 대표 키워드를 추출하고 연관 검색어를 수집하였다. 그리고 지표의 선정 기준에 따라 건축물 화재위험도 평가 지표를 결정하였다.

##### 4.1 요소 구성 및 대표 키워드 추출

수집되고 전처리된 비정형 데이터에서 성격이 비슷한 단어를 군집하기 위한 요소(components)를 결정하고 요소 내의 대표 키워드를 추출하기 위해 기존 연구에서 다루었던 화재위험도 지표를 검토하였다. 또한, 다양한 화재 영향 인자를 고려하고 확장하기 위해 재난 관련 위험도 지표를 추가적으로 검토하였다[5,20-26](Table 3).

**Table 3.** Search Review of previous studies related to disaster risk

Previous studies	Considerations for disaster risk assessment(indicators, etc.)
Frantzich[5]	fire protection facility, evacuation
Watts[20]	building location, fire frequency, fire probability
Shaw et al.[21]	physic, institutional, social, economic, natural
Cutter et al.[22]	community, economic, institutional, housing/infrastructural, social, environmental
Esnard et al.[23]	income, economic, race/ethnicity, age, residence, education
Burton[24]	social, economic, institutional, infrastructure, community capital, environmental systems
Beck et al.[25]	public infrastructure, nutrition, poverty and dependencies, economic capacity
Ko[26]	physical, social, economic, institutional, natural

검토를 통해 크게 4가지 요소인 건축물 특징-상태(Building characteristic-condition), 사회-경제(Socio-economy), 기후-기상(Climate-weather), 화재 방어(Fire protection)로 구분하였다. 이는 건축물 화재 위험도를 평가할 시 건축물 특징-상태(B), 사회-경제(S), 기후-기상(C) 요소가 증가할수록 위험도가 증가하고 화재 방어(P)가 감소한다는 식 (1)과 같은 개념식으로 나타낼 수 있다.

$$\text{Fire Risk Assessment} = B + S + C - P \tag{1}$$

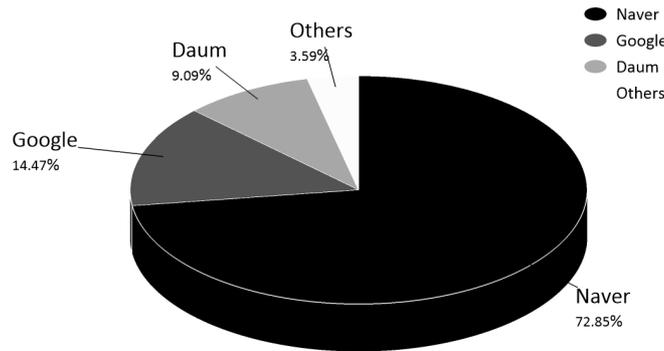
요소 내의 수많은 단어 중에는 연관 검색어를 수집하는데 적합하지 않은 추상적인 단어들이 존재하기 때문에, 선행연구에서 제시한 지표의 의미와 유사하고 요소 내의 의미를 포괄할 수 있으며 검색이 가능한 단어를 고려하여 대표 키워드를 추출하였다(Table 4).

**Table 4.** Key words extraction

Components	Key words
(B) Building characteristic-condition	building, electric, gas
(S) Socio-economy	building price, occupant
(C) Climate-weather	dry, weather
(P) Fire protection	fire protection facility, fire Resistance, evacuation

#### 4.2 연관 검색어 수집

본 연구에서는 대표 키워드에 파생되는 연관 검색어를 수집하기 위해 검색엔진을 이용하였다. 국내에서 연관 검색어 서비스를 제공하고 있는 대표적인 검색엔진으로는 구글(Google), 네이버(Naver), 다음(Daum), 네이트(Nate) 등이 있다. Figure 3과 같이 최근 약 10년(2010~2020)의 검색엔진 사용률은 네이버 72.85%, 구글 14.47%, 다음 9.09%, 네이트 포함 기타 3.59%인 것으로 나타났다[27]. 그중에서 네이버, 구글, 다음의 이용률이 96.41%로 국내 이용자들은 대부분 3개의 검색엔진을 사용하는 것을 알 수 있다. 따라서, 네이버, 구글, 다음의 검색엔진을 이용하여 대표 키워드에 파생되는 위험도 관련 지표를 수집하였다.



**Figure 3.** Internet search engine usage rate(2010~2020)

#### 4.3 지표의 선정 기준 설정

빅데이터 분석 결과에 기반하여 지표를 선정할 때 출현빈도가 높은 단어 및 용어를 선정하는 것이 방법이 될 수 있다. 하지만 단순히 출현빈도만을 고려하여 지표를 선정한다면, 지표 간의 의미 중복과 의도치 않은 다른 지표가 적용되는 문제가 발생한다. 수집된 연관검색어 중 건축물 화재위험도의 의미와 연관이 있고 이를 보다 구체적으로 표현해 줄 수 있는 지표가 선정되어야 한다. 따라서, 대표 키워드의 연관 검색어 중에서 인자선정원칙, 접근구득의 용이성을 고려하여 본 연구 목표와 부합하도록 아래와 같은 기준을 설정하고 지표를 선정하였다.

- 1) 연관성: 건축물 화재위험도에 영향을 미치는 지표여야 한다.
- 2) 대표성: 대표키워드와 연관 검색어 분석 결과를 대표할 수 있어야 한다.
- 3) 가용성: 건축물 단위의 통계 자료가 확보되거나 산정 가능해야 한다.
- 4) 지속성: 자료가 일정 시간 간격으로 지속적으로 갱신되고 생산되어야 한다.
- 5) 이해성: 간단하고 명료하여 쉽게 이해할 수 있어야 한다.
- 6) 방향성: 평가에 긍정적 혹은 부정적 영향의 일관성을 유지해야 한다.

7) 비교가능성: 시간의 변화에 따라 비교할 수 있고 모든 건축물에 적용 가능해야 한다.

#### 4.4 건축물 화재위험도 평가 지표 결정

수집된 연관 검색어를 대상으로 연관성, 대표성, 가용성, 지속성, 이해성, 방향성, 비교가능성을 검토하여 지표를 선정하였다. 그 결과, 4개 요소 내의 20개의 지표로 구성할 수 있었다. 요소별로 구성된 지표를 살펴보면 건축물 성질-상태는 ‘건축물 용도’, ‘건축물 구조’, ‘건축물 규모’, ‘건축물 노후도’, ‘건축물 밀집도’, ‘가스 사용량’, ‘전력 사용량’으로 7가지로 구성되었다. 사회-경제 요소는 ‘공시지가’, ‘재실자 비율’, ‘재난 취약자 비율’, ‘방화범 비율’, ‘도로 상태’로 5가지로 구성되었으며, 기후-기상 요소는 ‘건조 주의보’, ‘건조 경보’, ‘평년 습도’, ‘평년 풍속’으로 4가지로 구성되었다. 마지막으로 화재 방어 요소는 ‘소화시설 비율’, ‘소방인력 비율’, ‘불연재료 비율’, ‘대피 능력’인 4가지로 구성되었다. Table 5는 건축물 화재위험도 지표를 요소별-지표별로 구분하고 각각의 의미를 나타낸 것이다.

**Table 5.** Determination of building fire risk assessment indicators

Classification		Meaning of indicators
Components	indicators	
A. Building characteristic-condition	(1) Building use	Fire risk according to the use of the building (ex. detached house, apartment house, type 1 and 2 neighborhood living facilities, etc.)
	(2) Building structure	Fire risk according to building structure (ex. wooden structure, masonry structure, steel structure, concrete structure, etc.)
	(3) Building scale	Fire risk according to the size of the building (ex. number of floors, height, floor space, area, etc.)
	(4) Building deterioration	Fire risk according to the age of the building
	(5) Building density	Fire risk according to the density between buildings
	(6) Gas usage	Fire risk according to gas usage by building or lot number
	(7) Electricity usage	Fire risk according to electricity usage by building or lot number
B. Socio-economy	(8) Property	Public notified price of building
	(9) Occupant	Ratio of occupants per unit area of a building
	(10) Fire vulnerable people	Percentage of dependent population(under 15 and over 65) among occupants
	(11) Arson	Number of arson cases per population in the administrative district where the building is located
C. Climate-weather	(12) Road condition	The width of the road near the building where firefighting vehicles can enter
	(13) Dryness advisory	The number of times the dryness advisory takes effect for a certain period of time
	(14) Dryness warning	The number of times the dryness warning takes effect for a certain period of time
	(15) Humidity	Average humidity for a certain period in the administrative district where the building is located
D. Fire protection	(16) Wind velocity	Average wind velocity for a certain period in the administrative district where the building is located
	(17) Fire protection facility	Ratio of fire extinguishing facilities(fire hydrant, sprinkler, etc.) per unit area of the building
	(18) Fire officer	Ratio of fire officer per population in the administrative district where the building is located
	(19) Noncombustible material	Amount of Noncombustible materials in the building (fireproof structure, finishing materials, composite materials, etc.)
	(20) Evacuation	Evacuation route size per unit area of building

## 5. 결론

본 연구에서는 건축물 화재위험도에 영향을 미치는 지표를 결정하기 위해 웹 상에 존재하는 비정형 데이터를 수집하고 빅데이터 분석 기법을 이용하였다. 이를 통해 효과적인 건축물 화재관리를 위한 위험도 지표를 결정하였다. 연구의 주요 내용을 요약하면 다음과 같다.

- 1) 화재 관련 법령을 참고하여 질의어를 선정하고, 전문 문헌을 제공하는 학술 검색 사이트에서 학술지, 연구보고서 등 총 30,494건이 검색되었으며, 웹 크롤링 자동화 기법을 이용하여 텍스트 기반의 비정형 데이터로 구성된 9,768건의 문헌을 수집하였다.
- 2) 건축 및 화재관련 전문 용어집을 참고하여 사용자 용어사전 등록하였다. 또한, 수집된 문헌에서 중복되는 문헌과 무의미하게 사용되고 있는 불용어를 제거하는 비정형 데이터의 전처리 과정을 수행하였다.
- 3) 기존 연구의 검토를 통하여 4개의 요소로 구분하고, 요소를 대표할 수 있는 대표 키워드를 선정하였다. 검색엔진의 사용률에서 대부분을 차지하는 네이버, 구글, 다음 검색엔진을 이용하여 대표 키워드별 연관 검색어를 수집하였다.
- 4) 수집된 연관 검색어는 인자선정원칙과 접근구독의 용이성을 고려하여 연관성, 대표성, 가용성, 지속성, 이해성, 방향성, 비교가능성을 검토하였으며, 이를 통해 4개 요소 내의 총 20개의 지표를 결정하였다.

본 연구에서의 한계점 및 시사점과 이를 해결하기 위한 향후 연구 계획을 정리하면 아래와 같다.

- 1) 결정된 지표의 적절성과 신뢰성이 제기될 수 있다. 빅데이터 분석을 통하여 어느 현상을 예측하는 기술은 현재 많은 분야에서 활용되고 있지만, 검증할 수단이 마땅치 않기 때문에 분석 결과 자체가 사실이라고 가정되는 경우가 많다. 이를 해결하기 위한 대안으로 먼저 대상 건축물을 선정하여 각 지표에 따른 DB를 구축하고, 통계적 기법(요인, 주성분, 민감도 분석 등)을 이용하여 적절성과 신뢰성을 검증하는 방법도 고려해 볼 수 있다. 후속 연구는 우선적으로 이 부분에 많은 초점을 두고 수행할 계획이다.
- 2) 빅데이터 분석의 객관성과 과학성을 제고해야 한다. 수집된 연관검색어를 통해 인자선정원칙과 접근구독의 용이성을 고려하여 최종 지표를 결정하였지만, 이러한 과정이 배제되는 최적화된 빅데이터 모형을 활용할 필요가 있다. 이를 위해 빅데이터 분석 절차를 개선하거나, 머신러닝과 딥러닝을 연계한 새로운 빅데이터 모형을 개발하여 활용하는 등의 추가 연구를 통해서 본 연구의 방법론을 보완해야 할 것이다.
- 3) 결정된 지표를 바탕으로 건축물의 화재위험도를 평가하는 연구도 계속되어야 한다. 지표는 결국 평가를 하기 위한 하나의 과정이자 수단이다. 이를 위해 지표간의 중요도를 고려하여 가중치를 산정하고, 건축물의 효과적인 화재안전대책을 위한 위험도 평가 모형을 개발해야 한다.

본 연구에서 제시한 건축물 화재위험도 평가 지표는 기존 연구에서 고려되지 않았던 비정형 데이터를 활용하고 빅데이터 분석의 적용 가능성을 제시했다는 점에서 의미가 있을 것으로 판단된다. 또한 단순히 건축물 내부 요인만을 고려하는 평가 지표가 아닌, 사회과학 분야 등의 외부지표를 검토하는 복합적인 접근 방식을 추구하였다. 본 연구의 방법론은 건축물 화재에 영향을 미치는 다양한 차원을 고려하여, 건축물 화재 측면에서 간과되어 왔던 직·간접적인 영향 지표를 재설정하는 참고자료가 될 수 있을 것이다.

## 요약

본 연구에서는 빅데이터를 활용하여 건축물의 화재위험도 평가에 필요한 지표를 결정하였다. 건축물에서 화재위험도에

영향을 미치는 원인은 대부분 건축물만을 고려한 지표로 고착화되어 있기 때문에 제한적이고 주관적인 평가가 수행되어왔다. 따라서, 빅데이터를 활용하여 다양한 내·외부 지표를 고려한다면 건축물의 화재위험도 저감을 위한 효과적인 대책을 도모할 수 있다. 지표 결정에 필요한 데이터를 수집하기 위해 먼저 질의어를 선정하고, 웹 크롤링 기법을 이용하여 비정형 데이터 형식의 전문 문헌을 수집하였다. 문헌 내 단어를 수집하기 위해 사용자 용어사전 등록, 중복 문헌 및 불용어 제거의 전처리 과정을 수행하였으며, 선행 연구를 검토하여 단어를 4개의 요소로 분류하고 각 요소에서 위험도와 관련된 대표 키워드를 선정하였다. 그리고 대표 키워드의 연관검색어 분석을 통해 파생되는 위험도 관련 지표를 수집하였다. 지표의 선정 기준에 따라 수집된 지표를 검토한 결과, 20개의 건축물 화재위험도 지표를 결정할 수 있었다. 본 연구 방법론은 건축물 화재위험의 저감 대책 수립을 위한 빅데이터 분석의 적용 가능성을 나타내며, 결정된 지표는 건축물 화재위험도 평가를 위한 참고자료로 사용될 수 있을 것이다.

키워드 : 건축물, 화재 위험도 평가, 빅데이터

## Funding

This research was supported by Quality and Certification project(No. 20220035-001).

## ORCID

Hong-Jun Joo,  <http://orcid.org/0000-0003-2184-4440>

Yun-Jeong Choi,  <http://orcid.org/0000-0001-9415-7575>

Chi-Yeol Ok,  <http://orcid.org/0000-0002-1213-3623>

Jae-Hong An,  <http://orcid.org/0000-0003-0680-9233>

## References

1. National Fire Agency. 2020 National fire agency statistical year book. Sejong (Korea): National Fire Agency; 2021. 470 p.
2. Mi H, Liu Yaling, Wang W, Xiao G. An integrated method for fire risk assessment in residential buildings. *Mathematical Problems in Engineering*. 2020 Aug;2020:9392467. <https://doi.org/10.1155/2020/9392467>
3. National Fire Protection Association. NFPA 101A: Guide on Alternative Approaches to Life Safety. Quincy (MA): National Fire Protection Association; 2007. 96 p.
4. Bryant P. *Fire Strategies-Strategic Thinking*. London: Kingfell; 2013. 186 p.
5. Frantzich H. Risk analysis and fire safety engineering. *Fire Safety Journal*. 1998 Nov;31(4):313-29. [https://doi.org/10.1016/S0379-7112\(98\)00021-6](https://doi.org/10.1016/S0379-7112(98)00021-6)
6. Heo JE, Jeon GY, Hong WH. Study on the fire risk assessment in CBD based on the characteristic features of fire damage. *Journal of The Architectural Institute of Korea Planning & Design*. 2009 Mar;25(3):247-54.
7. Hwang HY, Baek KY, Park BH, Lee MH, Hwang JH, Ryu EL, Kim TH. Empirical application for the urban disaster risk assessment: fire, facility and escape cases in cheongju city. *Journal of The Korean Society of Hazard Mitigation*. 2001 Sep;1(2):123-37.
8. Kim YJ, Shin SY. *Developing a risk assessment method for the mitigation of urban disasters*. Seoul (Korea): Seoul Development Institute; 2009. 231 p.

9. Kim YS. Development of resilience indicator based on big data analysis under climate change [dissertation]. [Incheon (Korea)]: Inha University; 2016. 452 p.
10. Desouza KC. Realizing the promise of big data: Implementing big data projects. WA: IBM Center for The Business of Government. 2014. 42 p.
11. Park ES, Min SH. Standardization of fire factor for big data. *Journal of Korean Society of Hazard Mitigation*. 2019 Aug;19(4):143-9. <https://doi.org/10.9798/KOSHAM.2019.19.4.143>
12. Kim DH, Jo BW. A study on the development of a fire site risk prediction model based on initial information using big data analysis. *Journal of the Society of Disaster Information*. 2021 Jun;17(2):245-53. <https://doi.org/10.15683/kosdi.2021.6.30.245>
13. Kim YS, Choi CH, Bae YH, Kim DH, Kim DH, Kim HS. Indicator development and evaluation of storm and flood resilience using big data analysis: (1) Development of Resilience Indicators. *Journal of Korean Society of Hazard Mitigation*. 2019 Aug;18(4):97-107. <https://doi.org/10.9798/KOSHAM.2018.18.4.97>
14. Sirisuriya SCM. A comparative study on web scraping. *Proceedings of 8th International Research Conference*; 2015 Nov 18-20; Seville, Spain. Valencia (Spain): General Sir John Kotelawala Defence University; 2015. p. 135-40.
15. Park GC. Big data analysis for civil complaints using text mining technique: Gangnam-gu. Seoul (Korea): Seoul Digital Foundation. 2020. 30 p.
16. Silva C, Ribeiro B. The importance of stop word removal on recall values in text categorization. 2003 *Proceedings of the International Joint Conference on Neural Networks*. 2003 Jul 20-24; Portland, OR. New York(NY): Institute of Electrical and Electronics Engineers; 2003. p. 1661-6.
17. Li M, Ch'ng E, Chong A, See S. Twitter sentiment analysis of the 2016 U.S. Presidential Election Using an Emoji Training Heuristic. *Applied Informatics and Technology Innovation Conference*; 2016 Nov 22-24; Newcastle, Australia. New York (NY): Springer; 2016. p. 1-16.
18. Kim YS, Kang NR, Jung JW, Kim HS. A review on the management of water resources information based on big data and cloud computing. *Journal of Wetlands Research*. 2016 Feb;18(1):100-12. <https://doi.org/10.17663/JWR.2016.18.1.100>
19. Dole D. Ranks NL [Internet]. Massat (France): RANKS NL; 1998 Jan 1 [updated 2014 Jan 1; cited 2021 Oct 13]. Available from: <https://www.ranks.nl/stopwords/korean>.
20. Watts JM. *SFPE Handbook of Fire Protection Engineering: Fire risk indexing*. Quincy (IL): National Fire Protection Association; 2016. p. 5-125. [https://doi.org/10.1007/978-1-4939-2565-0\\_82](https://doi.org/10.1007/978-1-4939-2565-0_82)
21. Shaw R, Takeuchi Y, Joerin J, Fernandez G, Tjandradewi BI, Chosadillia, Wataya E, McDonald B, Fukui R, Sharma A, Tsunozaki E, Matsuoka Y. Climate and disaster resilience initiative capacity-building program. Tokyo (Japan): United Nations International Strategy for Disaster Reduction; 2010. 28p.
22. Cutter SL, Burton CG, Emrich CT. Disaster Resilience Indicators for Benchmarking Baseline Conditions. *Journal of Homeland Security and Emergency Management*. 2010 Aug;7(1):1-22. <https://doi.org/10.2202/1547-7355.1732>
23. Esnard AM, Sapat A, Mitsova D. An index of relative displacement risk to hurricanes. *Natural Hazards*. 2011 Apr;59(2):833-59. <https://doi.org/10.1007/s11069-011-9799-3>
24. Burton CG. A validation of metrics for community resilience to natural hazards and disasters using the recovery from hurricane katrina as a case study. *Annals of the Association of American Geographers*. 2014 Nov 4;105(1):67-86.
25. Beck M, Shepard C, Birkmann J, Rhyner J, Welle T, Witting M, Wolfertz J, Martens J, Maurer K, Mucke P. *World Risk Report 2012*. Berlin (Germany): Alliance Development Works; 2012. 74 p.
26. Ko CS. A Study for the Improvement of Disaster Management Systems in Korea [dissertation]. [Seoul (Korea)]: Kyunghee University; 2012. 253 p.
27. Internet Trend [Internet]. Seoul (Korea): Bizspring. 2010 - [cited 2021 Nov 22]. Available from: <http://www.internettrend.co.kr/trendForward.tsp>