

빅데이터 분석을 통한 화재조사서류 내 전자제품명 출현빈도 분석

박 유 현*, 배 덕 곤*, 최 태 일**, 정 원 일***, 김 경 배^o

Analyzing the Frequency of Appearance of Electronic Product Names in Fire Investigation Documents through Big Data Analysis

Yoo-hyun Park*, Deok-gon Bae*, Tae-il Choi**, Weon-il Jeong***, Gyoung-bae Kim^o

요 약

화재조사서류는 화재발생종합보고서, 화재현장조사서, 화재현장출동보고서, 화재피해산정서 등 발생한 화재에 대해 다양한 관점으로 작성된 문서들의 집합이다. 현재까지 화재조사서류에 대한 분석이 이루어지기는 하였으나 대부분 화재조사서류에 대한 메타데이터를 기반으로 분석하였고, 실제 문서들의 내용기반의 분석은 거의 이루어지지 않았다. 본 논문에서는 화재 원인 중 부주의의 다음으로 많은 전기제품에 대하여 화재조사서류 문서 내용을 빅데이터 분석하였다. 분석결과 화재조사서류에 공통적으로 많이 출현하는 전자제품명으로는 콘센트, 냉장고, 가스레인지가 도출되었다.

키워드 : 화재조사서류, 전자제품, 빅데이터, 화재예방

Key Words : fire investigation documents, electronic product, big data, fire prevention communication

ABSTRACT

Fire investigation documents are a collection of documents prepared from various perspectives on fires that have occurred, such as comprehensive fire report, fire field investigation report, fire site dispatch report, and fire damage calculation report. Although analysis of fire investigation documents has been made so far, most of them have been analyzed based on metadata of fire investigation documents, and little analysis of content base of actual documents has been made. In this paper, the contents of fire investigation documents were analyzed big data for the next largest number of electrical products among the causes of fire. As a result of the analysis, socket-outlets, refrigerators, and gas ranges were derived from the names of electronic products that commonly appear in fire investigation documents.

* First Author : Dongeui University Computer Software Engineering Major, yhpark@deu.ac.kr, 정희원

^o Corresponding Author : Seowon University Department of Information Security, gbkim@seowon.ac.kr, 종신회원

* University of Seoul Department of Disaster Science, justdoit@korea.kr

** National Fire Agency, wao907@korea.kr

*** Hoseo University Division of Computer and Information Engineering, wnchung@hoseo.edu

논문번호 : 202001-255-C-RN, Received October 13, 2020; Revised November 4, 2020; Accepted November 4, 2020

1. 서론

재난관리에서 예측 문제를 예방이나 초기 대응에 적용하기 위한 노력은 우리나라 뿐만 아니라 해외에서도 많은 진행되고 있다. 미국과 영국 등의 국가에서는 스마트 소방을 위한 지능형 소방 체계로 다양한 유관 기관으로부터 대량의 정보를 수집하고 이를 서로 연결하여 다양하게 분석하고 있으며, 분석한 결과를 토대로 미래 상황을 예측하여 지역사회·소방본부·지휘소·소방관 등에게 제공하고 있다. 우리나라에서도 인공지능과 빅데이터 기반 소방 예측 서비스를 통하여 육상재난 컨트롤 타워와 화재 예방 대응을 위한 사전 예방과 완화 체계를 정립하기 위해 많은 노력을 하고 있다. 이를 위하여 지능형 소방 예측 시스템의 방향을 소방활동에 필요한 데이터 중 과거 소방활동에 대한 각종 기록과 소방점검의 결과를 통하여 소방 고유의 목적 데이터를 자체적으로 발굴하고 생성하면서 관리하고, 그 밖의 정보는 관련 유관 기관의 협조를 통하여 수집하고 통합하려는 방향으로 움직이고 있다¹⁾.

기존 소방시스템에 다양한 IT 기술을 접목하려는 시도는 다양하게 진행되었다. 먼저, 해외 선진 사례 분석을 통해 국내에 적용할 수 있는 스마트 소방안전 정보 시스템의 방향성을 제안한 연구가 있었다. 이 연구에서는 화재재난상황 분석에 필요한 다양한 정보를 일원화된 데이터웨어하우스를 구축하고, 이를 분석, 활용하는 체계적인 사전점검을 통해 예방 효율을 극대화할 수 있는 체계 구축이 필요하며, 실시간 수집 데이터를 활용하여 화재 재난 상황을 조기에 감지하고, 최적의 대응 시나리오를 자동 생성하여 현장에 적용할 수 있는 체계 개발이 이루어져야 한다고 주장하였다²⁾.

또한, 네트워크 분석을 기반으로 과학기술을 통해 우선적으로 해결이 필요한 재난 유형 및 재난안전 관리 단계, 그리고 재난 안전문제 해결을 위해 필요한 기술을 분석한 연구도 있고³⁾, 국민이 보다 적극적으로 다양한 공공서비스를 활용할 수 있는 기반으로 전기안전 분야를 중심으로 빅데이터 플랫폼 기반의 서비스 유형을 제시한 연구도 있다⁴⁾. 재난이 발생했을 때, 재난안전상황실로 정보들이 구분별하게 수집되고 전달되고 있어 신속하고 올바른 의사결정에 어려움을 겪고 있는 상황을 개선하기 위해서 현장대응팀과 상황실 모두 최적의 재난 상황인지 및 대응이 가능토록 재난관리 필수정보 요소를 도출하고자 한 연구도 있다⁵⁾.

그리고, 유관기관에서 관리하고 있는 데이터를 활

용하기 위한 연구로 국토교통부 등에서 관리되고 있는 건물 데이터를 중심으로 화재 관련 융합 데이터셋을 생성하고 건물단위로 인공지능 기술을 적용하여 화재 발생 위험도를 예측할 수 있는 모델을 개발한 연구도 있다⁶⁾. 동시에 진행한 연구로, GIS건물통합정보 시스템(국토교통부)에서 관리하는 건물의 용도, 건폐율·용적률 등의 정보, 건축행정시스템(국토교통부)에서 관리하는 전기사용량 등의 정보, 소방방재청에서 관리하는 화재방생이력정보 등은 동일한 건물 객체에 대해 서로 다른 기관에서 서로 다른 목적으로 구축되어 형상(geometry)과 위치, 속성정보가 상이하여 상호 연계되기 어려운 문제점을 가지고 있어, 건물 객체를 중심으로 화재 예측이 가능하도록 이기종의 데이터를 융합 방법을 제안하였다⁷⁾.

마지막으로, IoT 장비 등을 활용하여 실시간 화재 감지를 위한 시스템을 제안한 연구들도 다수 있는데, IoT 관련 기존 연구 분석과 해외 사례분석을 통한 국내 현실을 비교하고, 국내 재난재해 서비스 발전을 위한 방향을 제안한 연구⁸⁾, 화재 감지의 정확도를 높이기 위해서 다중센서를 사용하는 화재 감지기를 제안한 연구⁹⁾, 화재감지를 실시간으로 할 수 있도록 화재 감지기 센싱 데이터를 수집, 빅데이터를 이용하여 실시간으로 판단(엣지컴퓨팅)하고, 중앙플랫폼(신규)이 전국에 걸쳐 이들 설비와 데이터처리 상태를 실시간 관리하고, 대응할수 있는 서비스를 제안한 연구도 있다¹⁰⁾.

한편, 2008-2017년 10년간 발생한 441,030건의 화재 원인을 살펴보면 ‘부주의’(215,245건, 48.8%)가 가장 많고, ‘전기적 요인’(41,905건, 100,823건, 22.9%), ‘기계적 요인’(9.5%) 등의 순이었다. 또한, 2017년 계절별 전기화재 발생 빈도는 ‘여름철’(6~8월) 2,300건(28.7%), ‘겨울철’(12~2월) 2,246건(28.0%), ‘가을철’(9~11월) 1,743건(21.8%), ‘봄철’(3~5월) 1,722건(21.5%) 순이었다. 여름·겨울철 전기화재가 전체 전기화재의 절반 이상(56.7%)을 차지하고 있는 것은 기온 변화에 따라 냉·난방기 사용 등 전력소비량이 급증하는 것이 원인인 것으로 분석하고 있다. 그리고, 전기화재 발생 최초 착화물은 ‘전기, 전자제품’이 6,156건(76.8%)으로 가장 많았고, ‘합성수지’ 1,077건(13.4%), ‘종이, 목재, 건조 등’ 238건(3.0%) 등의 순이었다¹¹⁾.

국가화재정보시스템(NFDS)는 대국민 서비스를 통해 화재정보 및 화재 관련 지식을 공유하는 국가적 안전관리 네트워크 시스템으로 화재의 원인, 발화, 발전, 통보 및 연소 확대 등의 화재 발생으로부터 피난상황,

소방설비의 작동 등 화재진압에까지 화재의 메커니즘과 관련된 화재정보와 이를 통계화한 정보를 국민에게 제공하고 있다. 국가화재정보시스템에서는 그림 1의 화재조사서류를 관리하고 있다.

화재조사서류란 소방기본법에서 규정하고 있는 화재조사의 결과를 사진이나 도면 등에 의하여 정확하게 기록하고 소방기관으로서의 최종의사결정을 기록한 문서이다. 화재조사서류는 화재현장을 영구적으로 보존하는 자료로서 화재 1건마다 작성된다. 이렇게 축적된 화재조사서류들은 다양하게 분석되고 유형화하여 시민에 대한 예방지도나 소방관계법령 등의 소방행정 제시의 기초자료로 하는 외에 소방활동자료로서 소방업무 전반에 활용된다^[12].

화재조사서류는 그림 1과 같이 화재발생종합보고서, 화재현장조사서, 화재현장출동보고서, 질문기록서, 화재피해액산정서, 감식감정결과통지서, 감정서 등으로 구성된다. 화재발생종합보고서는 화재현장조사서, 질문기록서 등의 내용을 집약하여 하나씩 정리하는 것으로 화재대상물의 종합적 내용에 소방활동을 추가한 것이다. 화재현장조사서는 발화원인, 연소확대원인, 사상자 발생원인 등을 조사한 서류로 유사화재 방지, 연소확대 및 인명피해방지 등의 화재예방을 중심으로 한 소방행정에 반영함을 목적으로 하며, 화재조사서류의 핵심이 된다. 화재현장조사서에 포함되는 내용은 기본사항(작성일, 작성자, 현장조사 일시, 현장조사 장소 및 물건, 현장조사시 입회인), 현장조사결과(현장의 위치 및 부근상황, 현장의 모양, 소손상황), 발화건물의 판정(현장관찰·확인상황, 화재현장출동시의 확인·조사상황, 발견상황, 결론), 발화지점의 판정, 발화원인의 판정(발화원과 착화물, 발화원으로부터 가연물로의 착화경과와 연소 경과, 발화에 이른 인적·물적 유인), 도면 및 사진으로 구성된다. 화재현장출동보고서는 소방대가 소방활동 중에 관찰·확인한 결과를 기록하여 화재원인판정에 있어서 발화건물의

판정 등의 자료로 활용하고자 출동도중의 관찰·확인 상황, 현장도착시의 관찰·확인 상황, 소방활동중의 관찰·확인 상황으로 구성된다^[12].

본 논문은 과거 소방활동에 대한 각종 기록인 화재조사서류를 기반으로 빅데이터 분석하여 화재 예방에 필요한 정보를 얻기 위한 연구의 일부분이다. 본 논문의 내용과 유사한 기존 연구들은 대부분 화재조사서류의 메타정보를 활용하여 정량적인 값을 추출할 수 있는 부분에만 집중하였다. 즉, 기존의 연구들은 현재의 국가화재정보시스템에 화재정보를 입력할 때, 다수의 파일 형태로 작성된 화재조사서류 내용 자체를 분석하는 것이 아니라, 이미 유의미한 정보로 추출된 다양한 기준의 선택지에서 입력자가 주관적으로 선택한 사항들을 기준으로 분석하였고, 대부분 화재 관련 통계자료들은 이러한 방식으로 작성되고 있다. 하지만, 본 논문에서는 화재조사서류의 파일 자체 내용에 기반하여 빅데이터 분석을 수행하여 기존에 도출된 기준 이외의 사항에 대한 분석도 할 수 있다.

본 논문에서는 화재조사서류의 파일 자체 내용 중에서 분석할 대상을 전자제품명으로 선택하여 분석하였다. 전자제품명에 한정하여 분석한 이유는 화재원인 중 부주의 다음으로 많은 전기적 요인에서 최초 착화물이 ‘전기, 전자제품’이 가장 많았기 때문이다.

II. 본 론

2.1 분석 대상 데이터

본 연구에서는 국가화재정보시스템에 저장되어 있는 화재조사서류에서 모든 단어들을 뽑아내어 다양한 종류의 단어군으로 나누어 분석하였다. 본 논문은 단어군 중에서 전자제품명을 대상으로 분석한 내용을 포함하며 문서당 전자제품명의 출현빈도를 1회로 제한하는 경우와 그렇지 않은 경우, 사망사고가 있는 중대 화재의 경우와 그렇지 않은 경우를 비교 분석하였다.

이를 위하여 먼저, 국가화재정보시스템에 저장되어 있는 화재조사서류 중에서 2015년 1월부터 2020년 5월까지 최근 5년의 데이터를 중심으로 빅데이터 분석을 수행하였다.

국가화재정보시스템에서 화재조사서류를 저장하기 위해서는 먼저, 각 화재에 화재발생 연월일이 포함된 화재번호를 부여한다. 하나의 화재에 대해 생성되는 문서의 종류는 그림 1과 같이 다양하기 때문에, 국가화재정보시스템에서는 화재번호 이후에 일련번호를 추가적으로 덧붙인 형태로 관리한다. 즉, 화재번호 n에 대하여 화재발생종합보고서는 n_1, 화재현장조사서는

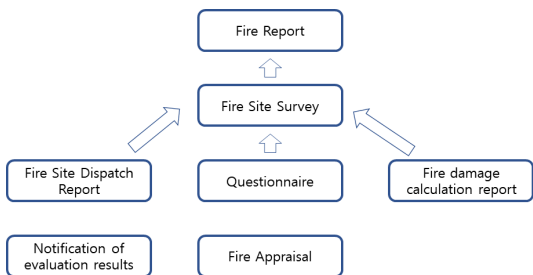


그림 1. 화재조사서류의 구성[12]
Fig. 1. Composition of fire investigation documents[12]

표 1. 분석대상 원본 데이터
Table 1. raw data for analysis

year	number of files
2015	131,710
2016	137,819
2017	162,578
2018	165,772
2019	158,513
2020	65,507
total	821,899

n_2, 화재현장출동보고서는 n_3과 같이 별도의 파일로 저장하고 있다. 저장하는 파일 수의 제한은 없으며 해당 화재에 관해 생성된 모든 자료를 포함하여 관리하고 있다.

표 1은 분석하기 전 화재조사서류 파일 개수를 년도별로 나타내었다. 화재보고서 파일은 한글, PDF, 이미지, 동영상, 엑셀, 텍스트 등 다양한 파일 형식으로 저장되어 있으며, 지역별로 일부 파일의 경우 암호화하여 저장되어 있기 때문에 본 논문에서는 암호화되지 않은 한글, PDF 파일만을 대상으로 분석하였다.

2.2 원본 데이터 가공

대부분의 화재조사서류는 화재현장조사서를 기본으로 포함하며 구성되지만, 작성자에 따라 추가적으로 다양한 파일을 포함하고 있는 경우도 많다. 본 연구에서는 관리의 편의성을 위해 화재당 하나의 파일로 병합하여 이 파일을 대상으로 키워드 추출을 하였다.

표 2는 화재당 하나의 파일로 병합된 파일의 수를 연도별 월별로 나타내었다. 표와 같이, 2015년-2019년 동안 1년에 5만건 이상의 화재가 발생하고 있으며 분석 대상 화재는 총 281,886건이다. 2020년의 경우는 5월 15일 이전까지의 데이터만을 대상으로 분석하였다.

표 2. 분석 대상 화재조사서류
Table 2. target data for analysis

year	1	2	3	4	5	6	7	8	9	10	11	12	total
2015	4,144	4,319	7,363	3,930	4,811	4,170	3,373	3,350	3,901	4,088	3,341	4,079	51,109
2016	5,459	5,568	5,797	4,423	4,355	3,663	3,387	4,410	3,402	3,409	3,937	4,486	52,476
2017	5,028	4,671	6,383	5,024	5,279	4,815	3,723	3,389	3,387	3,430	4,440	5,089	55,028
2018	5,231	5,700	4,885	4,694	3,695	4,330	4,351	4,785	3,245	3,922	3,752	5,255	53,875
2019	5,674	4,550	5,400	4,397	5,041	3,621	3,719	3,727	3,497	3,536	3,627	2,973	50,002
2020	4,080	3,902	5,177	5,225	1,012								19,396
total													281,886

2.3 키워드 추출 제외 규칙

병합된 화재조사서류 내의 키워드를 그대로 추출하게 되면 분석의 대상이 될 수 있는 유의미한 단어 뿐만 아니라 일반적인 용어들이 포함될 수 있어 분석 대상만 많아지게 하는 단점이 있다. 이를 해결하기 위해 다음과 같은 규칙을 통해 키워드 추출 단계에서 분석 대상을 축소하였다.

먼저, 그림 2와 같이 화재조사서류 문서 양식에 들어 있는 키워드들을 제외 하였다. 이러한 키워드는 표 3과 같다.

두번째, 문서 작성자의 특성으로 인해 화재조사서류 양식의 단어를 띄어 써서 발생하는 경우 등 1개 길이 단어를 제외하였다. 즉, 화재조사서류 양식에 ‘거리’를 ‘거 리’와 같이 띄어쓰기하여 기계적으로 키워드 추출할 때 ‘거’와 ‘리’가 다른 키워드로 추출되는 경우를 제거하였으며 일반적으로 많이 사용하는 ‘및’ 등의 단어도 자동 제외하였다.

세번째, 작성자의 오타 등을 제거하기 위해 국립국어원 국어대사전의 2020년 3월 기준에 포함되지 않는 단어를 제외하였다.

마지막으로 위 세 가지 규칙으로 뽑아낸 키워드들을 대상으로 화재전문가가 분석의 의미가 적다고 판

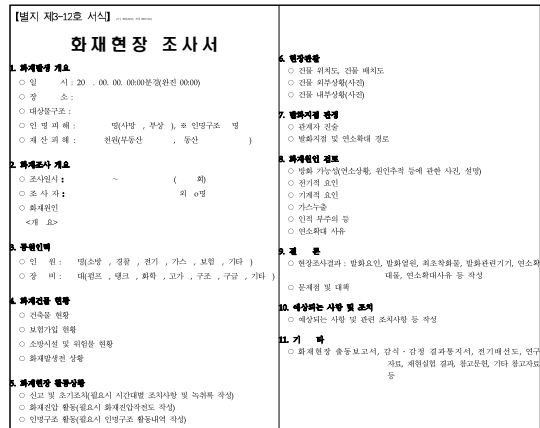


그림 2. 화재현장조사서 양식
Fig. 2. The Form of Fire Site Survey

표 3. 화재조사서류 양식에 나오는 키워드들
Table 3. The keywords in the fire investigation documents

거리 장소 건물 재산피해 관계인 조사일시 대상명 조사자 대상물구조 진술 발견 추정 발생 특이사항 보험 관계 현장 상황 화원 소방시설작동여부 화재 신고 화재원인 인명피해 화재현장 일시

표 4. 화재전문가에 의해 제외된 키워드들
Table 4. Keywords excluded by fire experts

가능성 가득 가운데 감식 감안 강원도 개요 개인 결과 경계 경로 경상남도 경상북도 경찰 계급 곤란 관련 광역시 구급 구급대 구분 구청 금액 기적 나이 나타남 날짜 남아 네이버 년도 년생 년월 다른 다소 담당 담당자 당시 대구 대상 대책 대표자 도로 동시 동영상 동원 따라서 로부터 로서 마을 마치 모두 모습 목격 문제 문제점 민원 반대편 발견 방문 방위 배제 번지 보고 보고서 보아 보험 본부 부근 부대 부분 분사 비고 비공개 사업 사용 사용자 사유 사진 사항 상기 상동 상이 상존 상태 상황 상황보고 생년 생략 서명 성명 성분 성은 세부 센터 소결 소방 소방대 소방본부 소방장 소방차 소속 소유자 시스템 신고 신원 실시 실제 아래쪽 아반떼 안전 압축기 양주시 양호 어머니 예상 완전 완진 완전함 용인시 우측 운행 원인 월일 위치 위해 위험 이득 이용 이하 인력 인명구조 인원 인적 입학 자신 자체 작전 재난 재산 저장 접근 접수 정부 조금 조사 조치 종료 종합 좌우 좌측 주소 주식회사 준비 중앙 지구대 지나 지역 지점 지휘 진출 진압 진압대 진압대원 진압합 진입 진행 진화 참고자료 참석자 참조 처음 천원 초기 촬영 최종 충청북도 특이 판단 평면도 평상시 품명 피해 피해자 해당 현장 현장사진 현재 현황 형태 화살표 화성 화재원인 화재현장 확인 활동 후작

표 5. 키워드 추출 Pseudo 코드
Table 5. Pseudo Code for extracting keywords

```
var filelist // store a list of files in a dictory
var extend // file extension for a certain file
var textdata
var noun_list // keyword nonu list
filelist ← get a list of all files in a dictory
while filelist :
    extend ← get the extension of the file
    if extend == "hwp"
        do convert hwp to html file
        (textdata, frequency) ← extract each none
            and frequency from html file
    else if extends == "pdf"
        (textdata, frequency) ← extract each none
            and frequency from pdf file
```

표 6. 시스템 환경
Table 6. System Configuration

OS : Windows 10
Processor : Intel i5-7500 3.4GHz
RAM : 16GB
SSD : 128GB

단한 키워드들을 제거하였다. 이 단계는 대상 키워드가 너무 많아서 모든 단어를 대상으로 검토하지는 못하였고, 빈도수가 높은 단어를 대상으로 검토하여 표 4의 단어들을 제외 하였다.

III. 실험

3.1 실험 환경

II에서 정의한 제외규칙을 적용하여 전체 화재조사서류를 대상으로 키워드를 추출하였다. 키워드를 추출하기 위해서 표 5와 같은 방법으로 Python 코드를 작성하여, PC 환경에서 실험하였다.

실험한 사용된 PC 환경은 표 6과 같다.

3.2 추출 키워드

실험결과 추출된 키워드 중 상위 100개는 그림 3과 표 7과 같다. 여기에서는 하나의 화재조사서류 상 키워드가 중복해서 여러 번 나타나는 것을 모두 더한 결과를 보여준다.

그림 3에서는 상위 100개의 키워드에 대한 빈도수를 그래프로 표현하였고, 표 7에서는 상위 100개의 모든 키워드들을 표시하였다. 지면상의 이유로 그림 3의 가로축에 나타나는 키워드들은 상위 100개의 단어 중

일부만을 표시하였다. 여기에서 가장 빈도수가 높게 나오는 ‘내부’라는 키워드는 전체 파일 281,886개 중에서 1,011,747번 나타나 하나의 화재조사서류 파일에서 평균 3.5회 나타남을 알 수 있다.

이 실험에서 추출된 키워드를 그대로 해석하는 데는 많은 문제점이 있을 것으로 보인다. 내부, 구조, 주변, 주택 등 공간적인 정보는 물론, 확대, 소실, 연소, 관찰 등 화재 상황에 대한 기술내용, 가스, 담배 등 화재 원인 추정 대상 등이 함께 포함되어 있으며, 특정 키워드가 화재 원인인지, 화재의 결과인지 등에 대해서도 해석하기 어려운 점이 있다.

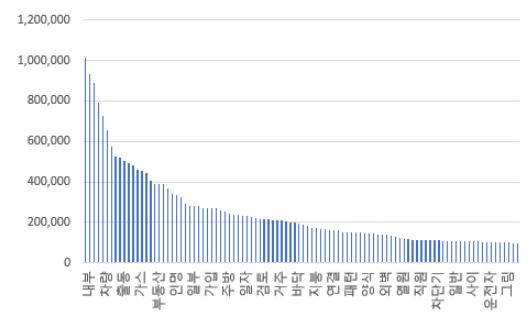


그림 3. 화재조사서류 빈도 상위 100개 키워드
Fig. 3. Top 100 keywords frequency in fire investigation documents

표 7. 화재조사서류 빈도 상위 100개 키워드
Table 7. Top 100 keywords frequency in fire investigation documents

no	keyword	no	keyword	no	keyword	no	keyword
1	내부	26	면적	51	엔진	76	제품
2	확대	27	판정	52	공장	77	직원
3	소실	28	흔적	53	지붕	78	장비
4	연소	29	가입	54	비닐하우스	79	냉장고
5	차량	30	동산	55	지방	80	그을음
6	구조	31	담배	56	건축물	81	차단기
7	주변	32	탄화	57	연결	82	개방
8	도착	33	주방	58	가연	83	화장실
9	출동	34	소화기	59	화학	84	관계
10	주택	35	작동	60	컨테이너	85	일반
11	관찰	36	작업	61	패턴	86	거실
12	식별	37	일자	62	물탱크	87	전면
13	가스	38	벽면	63	시간	88	피복
14	연기	39	공초	64	탱크	89	사이
15	최초	40	선착	65	양식	90	냄새
16	기계	41	검토	66	전원	91	아파트
17	부동산	42	설비	67	천장	92	적재
18	외부	43	특이점	68	지하	93	운전자
19	펌프	44	인근	69	외벽	94	출입구
20	설치	45	거주	70	지상	95	목격자
21	인명	46	보일러	71	하부	96	진경
22	배선	47	시설	72	출입문	97	그림
23	창고	48	방향	73	열원	98	인접
24	전선	49	바닥	74	박스	99	사무실
25	일부	50	상부	75	용융	100	창문

3.3 전자제품명 출현빈도

앞에서 언급한 바와 같이 표 7의 내용을 그대로 해석하는 데는 어려운 점이 많아 본 논문에서는 전자제품명을 기준으로 추가적인 분석을 하였다. 그림 3과 표 7의 내용 중 전자제품만을 재추출하여 그림 4와 표 8의 결과를 얻었다. 그림 4와 표 8은 화재조사서류상 나타난 전자제품을 모든 빈도수를 합한 결과를 보여 주고 있다.

그림 4는 화재조사서류 상에 나타나는 전자제품에 대한 총 빈도수를 그래프로 표현하였고, 표 8은 모든 전자제품명의 순서를 표시하였다. 지면상의 이유로 그림 4의 가로축에 나타나는 전자제품명은 일부만을 표시하였다.

화재조사서류를 작성한 작성자에 따라 화재발생종합보고서, 화재현장조사서, 화재현장출동보고서, 화재

표 8. 화재조사서류상 전자제품 빈도
Table 8. Frequency of electronic products in fire investigation documents

no	keyword	no	keyword	no	keyword	no	keyword
1	보일러	24	오븐	47	전기포트	70	계입기
2	콘센트	25	충전기	48	오디오	71	스티커
3	냉장고	26	디지털카메라	49	백열등	72	카세트
4	건조기	27	모니터	50	해어프라이어	73	카메라
5	에어컨	28	가전제품	51	제습기	74	하드디스크
6	히터	29	휴대폰	52	다리미	75	마우스
7	가스레인지	30	라디오테이머	53	전화기	76	태블릿
8	모터	31	전자레인지	54	약탕기	77	마이크
9	배터리	32	전기밥솥	55	포트	78	턴테이블
10	세탁기	33	진구	56	프린터	79	리시버
11	김치냉장고	34	핸드폰	57	압력밥솥	80	프로젝터
12	선풍기	35	그릴	58	인터폰	81	이어폰
13	형광등	36	청소기	59	드라이어	82	흡수어터
14	전등	37	석유난로	60	복합기	83	연수기
15	전기장판	38	스텐드	61	토스터	84	그래픽카드
16	버너	39	커피포트	62	라디오	85	녹음기
17	전화	40	스피커	63	재봉틀	86	헤드셋
18	멀티탭	41	파워	64	무전기	87	전동칫솔
19	카메라	42	텔레비전	65	가습기	88	포인터
20	전기레인지	43	쿨러	66	스마트폰	89	메모리카드
21	컴퓨터	44	노트북	67	공유기	90	헤드폰
22	정수기	45	믹서기	68	앰프	91	전자사전
23	조명등	46	비데	69	에어컨디셔너	92	타블렛

피해예산정서 등 그림 1의 다양한 문서를 충실히 포함한 경우에는 하나의 화재에 대해서 같은 키워드가 반복해서 나타나는 경우가 있기 때문에, 위의 결과 그림 4와 표 8을 기준으로 판단하기에는 어려움이 있다. 즉, 화재조사서류 A의 경우에는 화재발생종합보고서, 화재현장조사서, 화재현장출동보고서, 화재피해예산정서가 모두 포함되어 있고, 각각의 문서에 중복된 표현이 있으며 그 내용 중에 전자제품명이 표현되었다면, 해당 전자제품명의 빈도수는 단순히 화재현장조사서만 포함하고 있는 화재조사서류 B에 비하여 매우 많

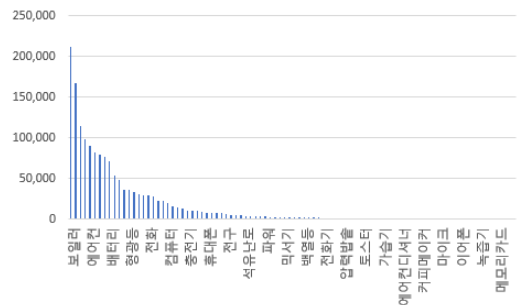


그림 4. 화재조사서류상 전자제품 빈도
Fig. 4. Frequency of electronic products in fire investigation documents

이 나타날 수 있다. 따라서, 하나의 화재조사서류 파일 내에 빈도수가 많은 키워드들도 1회 빈도로만 인정하는 추가 실험을 하였다. 이에 대한 결과는 그림 5와 표 9와 같다.

그림 4와 그림 5, 표 8과 표 9의 경우를 함께 살펴보면, 콘센트, 냉장고, 전화, 에어컨, 보일러, 모터, 가스레인지, 전등, 배터리, 히터, 세탁기, 형광등, 선풍기, 건조기, 멀티탭 등은 양쪽 모두 20위 안에 포함되어 화재조사서류에 주로 나타나는 전자제품으로 판단할

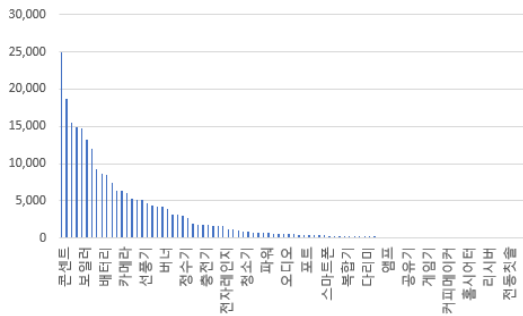


그림 5. 화재조사서류상 전자제품 빈도(문서당 최대 1회)
Fig. 5. Frequency of electronic products in fire investigation documents (Up to 1 per document)

표 9. 화재조사서류상 전자제품 빈도(문서당 최대 1회)
Table 9. Frequency of electronic products in fire investigation documents (Up to 1 per document)

no	keyword	no	keyword	no	keyword	no	keyword
1	콘센트	24	휴대폰	47	석유난로	70	카세트
2	냉장고	25	정수기	48	백열등	71	토스터
3	전화	26	핸드폰	49	포트	72	하드디스크
4	에어컨	27	전구	50	인터폰	73	게임기
5	보일러	28	모니터	51	믹서기	74	마이크
6	모터	29	충전기	52	무전기	75	마우스
7	가스레인지	30	전기레인지	53	스마트폰	76	태블릿
8	전등	31	전기밥솥	54	압력밥솥	77	카메라
9	배터리	32	라디에이터	55	전기포트	78	프로젝터
10	다지털카메라	33	전자레인지	56	비데	79	이어폰
11	히터	34	텔레비전	57	복합기	80	터레이블
12	세탁기	35	오븐	58	해상도어	81	홀시어터
13	카메라	36	스탠드	59	채송기	82	연수기
14	조명등	37	청소기	60	라디오	83	그래픽카드
15	형광등	38	그릴	61	다리미	84	녹음기
16	컴퓨터	39	노트북	62	야채관리기	85	리시버
17	선풍기	40	전화기	63	드라이어	86	메모리카드
18	가전제품	41	피워	64	스터커	87	헤드셋
19	건조기	42	커피포트	65	램프	88	포인터
20	멀티탭	43	클러	66	가스기	89	진동찾솔
21	버너	44	스피커	67	재봉틀	90	헤드폰
22	김치냉장고	45	오디오	68	악랑기	91	전자사전
23	전기장판	46	프린터	69	공유기	92	타블렛

수 있다.

또한, 본 연구에서는 사망사고가 있었던 화재만을 대상으로 같은 실험을 수행하였다. 그림 6과 표 10은 사망사고가 있었던 화재조사서류에 나타난 전자제품의 모든 빈도수 합한 결과를 나타내고, 그림 7과 표 11는 같은 경우에서 화재조사서류 당 1회씩의 빈도로 추출된 전자제품들을 보여주고 있다.

두 경우 모두 콘센트, 냉장고, 가스레인지가 상위 1-3위에 공통적으로 포함되고 있다.

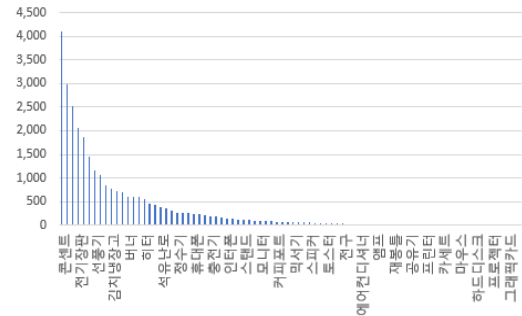


그림 6. 화재조사서류상 전자제품 빈도(사망사고)
Fig. 6. Frequency of electronic products in fire investigation documents (Fire with the dead)

표 10. 화재조사서류상 전자제품 빈도(사망사고)
Table 10. Frequency of electronic products in fire investigation documents (Fire with the dead)

no	keyword	no	keyword	no	keyword	no	keyword
1	콘센트	22	정수기	43	믹서기	64	공유기
2	냉장고	23	전자레인지	44	라디오	65	태블릿
3	가스레인지	24	핸드폰	45	오븐	66	무전기
4	전기장판	25	휴대폰	46	스피커	67	프린터
5	보일러	26	전기밥솥	47	가습기	68	복합기
6	에어컨	27	카메라	48	그릴	69	마이크
7	선풍기	28	충전기	49	토스터	70	카세트
8	형광등	29	텔레비전	50	전자레인지	71	게임기
9	전등	30	조명등	51	스터커	72	이어폰
10	김치냉장고	31	인터폰	52	전구	73	마우스
11	세탁기	32	오디오	53	압력밥솥	74	백열등
12	전화	33	다지털카메라	54	라디에이터	75	카메라
13	버너	34	스탠드	55	야채관리기	76	하드디스크
14	모터	35	전기포트	56	다리미	77	피워
15	멀티탭	36	노트북	57	드라이어	78	악랑기
16	히터	37	모니터	58	램프	79	프로젝터
17	가전제품	38	전화기	59	스마트폰	80	터레이블
18	컴퓨터	39	청소기	60	바데	81	전자사전
19	석유난로	40	커피포트	61	재봉틀	82	그래픽카드
20	건조기	41	포트	62	채송기	83	클러
21	배터리	42	해상도어	63	리시버	84	헤드셋

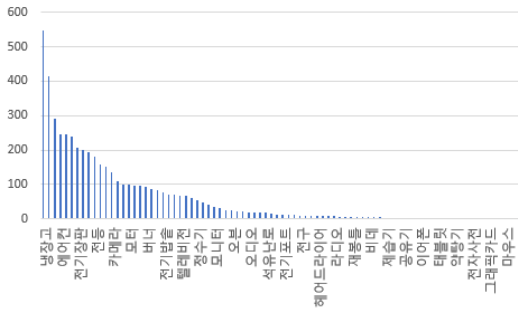


그림 7. 화재현장조사서 전자제품 빈도(사망사고, 문서당 최대 1회)
 Fig. 7. Frequency of electronic products in fire investigation documents (Fire with the dead, Up to 1 per document)

표 11. 화재현장조사서 전자제품 빈도(사망사고, 문서당 최대 1회)
 Table 11. Frequency of electronic products in fire investigation documents (Fire with the dead, Up to 1 per document)

no	keyword	no	keyword	no	keyword	no	keyword
1	냉장고	22	전기밥솥	43	전기포트	64	공유기
2	콘센트	23	히터	44	라디에이터	65	게임기
3	가스레인지	24	조명등	45	다리미	66	커피메이커
4	에어컨	25	텔레비전	46	전구	67	이어폰
5	전화	26	전자레인지	47	스마트폰	68	카세트
6	보일러	27	건조기	48	포트	69	하드디스크
7	전기장판	28	정수기	49	해상도어	70	태블릿
8	가전제품	29	충전기	50	압력밥솥	71	백열등
9	세탁기	30	배터리	51	무전기	72	야간안시거
10	전등	31	모니터	52	라디오	73	악당기
11	선풍기	32	스탠드	53	그릴	74	프로젝터
12	형광등	33	전화기	54	믹서기	75	턴테이블
13	카메라	34	오븐	55	재봉틀	76	전자사전
14	컴퓨터	35	청소기	56	스티커	77	리시버
15	김치냉장고	36	노트북	57	드라이어	78	앰프
16	모터	37	오디오	58	비데	79	그래픽카드
17	휴대폰	38	인터폰	59	프린터	80	쿨러
18	멀티탭	39	커피포트	60	복합기	81	파워
19	버너	40	석유난로	61	제습기	82	마우스
20	다자발베터	41	전자레인지	62	가습기	83	헤드셋
21	핸드폰	42	스피커	63	마이크	84	토스터

3.4 전자제품명 출현빈도 분석

본 논문에서 실험한 결과 화재조사서류에 등장하는 전자제품은 콘센트, 냉장고, 전화, 에어컨, 보일러, 모터, 가스레인지, 전등, 배터리, 히터, 세탁기, 형광등, 선풍기, 건조기, 멀티탭, 전기장판, 컴퓨터, 김치냉장고, 버너 등으로 나타났다. 특히 사망사고가 있었던 사고에 대해서만 살펴보면 콘센트, 냉장고, 가스레인지가 1-3위로 공통적으로 나타났다.

본 논문에서 화재조사서류상 출현 빈도가 높은 전자제품의 경우, 화재요인으로 고려되는 경우도 있지만, 화재피해액 산정서 등에서 피해물품으로 나타나는 경우도 많아 빈도수가 많은 제품이 화재원인이 된다고 보기는 어려운 점이 있다. 따라서, 화재조사보고서만을 대상으로, 화재조사보고서 중에서도 발화지점의 판정, 발화원인의 판정 부분만을 대상으로 분석하면 화재 발화에 관련한 전자제품을 보다 정확하게 분석할 수 있을 것으로 판단된다. 이러한 부분에 대한 연구는 추후 연구 과제에서 보다 구체적으로 연구할 예정이다.

IV. 결론

소방청은 인공지능과 빅데이터 기반 소방 예측 서비스를 통하여 육상제난 컨트롤 타워와 화재 예방 대응을 위한 사전 예방과 완화 체계를 정립하고 검증할 필요가 있다. 이를 위해 소방청은 지능형 소방 예측 시스템의 방향을 소방활동에 필요한 데이터 중 과거 소방활동에 대한 각종 기록과 소방점검의 결과를 통하여 정보 등을 획득하기 위한 목적을 달성하기 위한 소방 고유의 목적 데이터를 자체적으로 발굴하고 생성하면서 관리하여야 하며, 그 밖의 정보는 관련 유관기관의 협조를 통하여 수집하고 통합하여야 한다.

본 논문은 과거 소방활동에 대한 각종 기록인 화재조사서류를 기반으로 빅데이터 분석하여 화재 예방에 필요한 정보를 얻기 위한 연구의 일부분으로 화재원인 중 부주의의 다음으로 많은 전기적 요인에서 최초 착화물이 ‘전기, 전자제품’이 가장 많았기 때문에, 화재조사서류에 나타난 전자제품의 빈도수를 우선 분석하였다.

분석결과 콘센트, 냉장고, 전화, 에어컨, 보일러, 모터, 가스레인지, 전등, 배터리, 히터, 세탁기, 형광등, 선풍기, 건조기, 멀티탭 등의 제품이 많은 빈도수로 추출되었다. 특히 사망사고가 있었던 사고에 대해서만 살펴보면 콘센트, 냉장고, 가스레인지가 1-3위로 공통적으로 나타났다. 이러한 전자제품들은 모두 화재원인으로 해석되는 것은 아니지만 화재조사서류에서 나타나는 빈도수가 많기 때문에 구체적인 연구가 필요할 것으로 보인다.

또한, 단순히 화재조사서류 전체를 대상으로 키워드를 뽑아서 분석하는 것보다는 화재조사서류의 특정 영역들만을 뽑아서 분석하거나 한 개 문장내에 나타나는 키워드들을 묶어서 분석하는 방법 등이 추가적으로 진행되어야 할 것으로 보인다.

본 논문의 결과에 추가하여 위에서 언급한 향후 연구로 전자제품 중 화재 요인이 되는 원인을 정교하게 분석하게 된다면, 화재점검 시 우선순위를 높여 점검하는 등의 행정적인 조치를 취할 수 있는 근거로 사용할 수 있을 것으로 예상된다.

References

[1] D. Lee, "Smart Firefighting Prediction System Based on Artificial Intelligence and Big data," Yunseongsa Press, 2020

[2] B. G. Choi, T. W. Lee, M. S. Kim, G. Y. Choi, B. K. Son, J. Lee, H. Kim, and G. B. Kim, "A direction for smart fire fighting system through case analysis," in *Proc. Symp. KICS*, pp. 251-252, 2019.

[3] J. Heo and C. H. Yang, "Science and technology networks for disaster and safety management : Based on expert survey data," *J. Korea Contents Assoc.*, vol. 18, no. 11, pp. 123-134, 2018.

[4] S. H. Kwon and D. H. Jeong, "A study on big data platform-based service for public safety: focus on electricity safety," *The e-Business Stud.*, vol. 20, no. 5, pp. 3-15, 2019.

[5] J.-Y. Cho, J.-I. Song, M.-Y. Jang, and C.-R. Jang, "A study on the essential information to collect disaster sites for effective disaster management : Focused on jecheon sports center fire case," *J. Soc. Disaster Inf.*, vol. 16, no. 1, pp. 70-78, 2020.

[6] K. Ko, J. Yang, D. Hwang, H. Ko, C. Ga, and J. Cho, "Building fire prediction model study using AI," *J. KICS*, vol. 5, no. 7, pp. 1210-1218, 2020.

[7] K. Ko, C. Ga, J. Cho, H. Ko, and H. Kim, "A study on convergence of attribute information of building objects for fire prediction," *J. KICS*, vol. 45, no. 7, pp. 1219-1227, 2020.

[8] Y. Yoon and H. Choi, "Study of IoT service strategy for prevent disasters," *J. Korea Contents Assoc.*, vol. 17, no. 11, pp. 102-109, 2017.

[9] J. K. Park and K. Nam, "Implementation of multiple sensor data fusion algorithm for fire

detection system," *J. Korea Soc. Comput. and Inf.*, vol. 25, no. 7, pp. 9-16, 2020.

[10] S. B. Hong, D. J. Yoo, D. H. Pack, and J. S. Yang, "A study on the integrated control platform for fire disaster prevention and response based on PS-LTE," *J. KICS*, vol. 45, no. 4, pp. 699-705, 2020.

[11] *Electric fire safety survey of large home appliances - Focusing on facilities for the vulnerable*, Korean Consumer Agency, pp. 1-31, 2018.

[12] J. Lee, *Theory & Practice in Fire Investigation*, Sinjeongsa Press, 95, 2007.

박 유 현 (Yoo-hyun Park)



1996년 : 부산대학교 전자계산학과 졸업
 1996년 : 부산대학교 전자 계산학과 석사
 2008년 : 부산대학교 전자계산학과 박사
 2000년 : 한국국방연구원 연구원

2001년~2009년 : 한국전자통신연구원 선임연구원
 2009년~현재 : 동의대학교 컴퓨터소프트웨어공 전공교수

<관심분야> 빅데이터, 클라우드

배 덕 곤 (Deok-gon Bae)



1996년 : 전북대학교 산업공학과
 2008년 : 서울시립대학교 방재공학과 석사

2019년~2020년 : 세종시 소방본부장

2020년~현재 : 소방청 소방분석제도과장

<관심분야> 화재안전, 빅데이터

최 태 일 (Tae-il Choi)



2001년 : 동서울대학교 전자통신과
2000년~2006년 : Cyberbank QC Team, Development 3Team
2009년~2015년 : 경기도 소방재난본부 정보통신팀 근무

2020년~현재 : 소방청 소방분석제도과
<관심분야> 소방안전, 빅데이터, AI

김 경 배 (Gyoung-bae Kim)



1994년, 2000년 : 인하대학교 전자계산공학과 석사/박사
2000년~2004년 : 한국전자통신연구원 선임연구원
2004년~현재 : 서원대학교 정보안학과 교수
<관심분야> 빅데이터, 무인항공기, 재난재해, 정보보안

[ORCID:0000-0001-6131-8846]

정 원 일 (Weon-il Jeong)



1999년 : 인하대학교 전자계산학과 학사
2004년 : 인하대학교 컴퓨터정보공학부 박사
2004년~2006년 : 한국전자통신연구원 선임연구원
2007년~현재 : 호서대학교 컴퓨터정보공학부 교수

<관심분야> 빅데이터, 클라우드컴퓨팅, 시스템 보안