

## 형태소 분석을 이용한 발화관련 기기의 새로운 입력 키워드 추출

김은주 · 최정우\* · 류정우\*<sup>†</sup>

특허법인 MAPS, \*(주)세이프티아 기술연구소

### New Input Keyword Extraction of Equipments Involved in Ignition Using Morphological Analysis

Eun Ju Kim · Jeung Woo Choi\* · Jung Woo Ryu\*<sup>†</sup>

MAPS Intellectual Property Law

\*Technical Research Center, SafeTia Co., LTD.

(Received March 20, 2014; Revised March 21, 2014; Accepted April 20, 2014)

#### 요 약

급변하는 사회에서 기존에는 없었던 새로운 화재 사고가 발생하거나, 또는 과거에는 자주 발생된 화재 사고가 발생하지 않는 경우가 많아지고 있다. 본 논문에서는 화재분류체계를 개선할 때 사용할 수 있는 새로운 명사를 화재조사데이터로부터 추출하는 방법을 제안하였다. 화재조사데이터는 화재 조사관에 의해 작성된 화재사고 관련 데이터이다. 제안한 방법은 기타 소분류에 포함된 화재조사데이터의 제품명과 발화개요를 형태소 분석하여 발화관련 기기의 소분류 범주로 사용할 수 있는 새로운 입력 키워드를 추출하였다. 본 논문에서는 제품명을 분석하여 발화관련 기기 대분류인 농업용장비와 계절용 기기의 새로운 소분류 범주로 “건조기”와 “보일러”를 추출하였다. 또한 발화개요를 분석하여 발화관련 기기 대분류인 상업장비와 영상, 음향기기의 새로운 소분류 범주로 “수족관”과 “모니터”를 추출하였다. 이 네 가지의 새로운 소분류 범주로 화재조사데이터를 재분류하였을 때 총 3,808건 중 14.39%인 548건이 재분류되었다.

#### ABSTRACT

New types of fire accidents appear or the existing types disappeared because of rapidly changing society. We proposed a methodology of extracting new nouns from fire investigation data each of which is an accident report produced by fire investigators. The new nouns could be used in modifying the existing categories for classifying fire accidents. We analysed morphology of the product names and the ignition summaries using the proposed method for the fire accidents classified as the etc sub-category of the category of equipments involved in ignition. In this paper, we found “dryer” as a new sub-category of the agricultural equipment category and “boiler” in the seasonal appliance category from the product names of the fire accidents. We also extracted the new input keywords of “aquarium” and “monitor” in the commercial facilities category and the video, audio apparatus category from the ignition summaries respectively. Using the four sub-categories, we reclassified 548 (14.39%) of 3,808 fire accidents assigned to the etc sub-category.

**Keywords :** Fire data analysis, Morphological analysis, Fire information system, Fire investigation data

### 1. 서 론

현재 사용 중인 국가화재분류체계 개정안은 2006년 소방방재청에서 작성 발표하고 2007년 1월 1일부터 시행된 것이다. 이 국가화재분류체계에 따라 단일 화재의 화재발생 장소, 일시, 인명과 재산 피해정도(사망, 부상 등), 발화관련 정보 등 총 170여개의 속성에 대한 값이 저장되며 이를 화재조사데이터라 한다<sup>(1,2)</sup>. 현재 이 개정안이 발표된 지 수년이 지났기 때문에 최근 발생하고 있는 화재의 정보

를 모두 제대로 관리하지 못하고 있다. 특히 시간이 지남에 따라 큰 변화가 없는 “발화열원”, “동력원”, “발화요인” 등 속성에 비하여 매년 새로운 종류의 기기가 생기고 없어지는 “발화관련 기기”의 경우에는 기존 분류체제로 분류하지 못하고 기타(etc)나 발화관련 기기가 비어있는 상태로 분류되는 경우가 증가하고 있으므로, 이를 보완할 필요가 있다. 또한, 발화관련 기기는 기존에 많이 사용되던 기기가 시간이 지남에 따라 사용 빈도가 낮아지므로 자연스럽게 화재의 빈도도 줄어들기도 하고, 기존에는 없었던 기기가

<sup>†</sup>Corresponding Author, E-Mail: ryu0914@safetia.com  
TEL: +82-2-2057-0555, FAX: +82-2-2057-5111

ISSN: 1738-7167  
DOI: <http://dx.doi.org/10.7731/KIFSE.2014.28.2.091>

새로 개발되고 사용됨에 따라 화재 빈도가 높아지기도 한다. 예를 들어, 국가화재분류체계 개정안 발표 이후 디지털 카메라(digital camera)와 스마트폰(smart-phone)의 급속한 보급으로 인하여 사진현상 비용이 줄어들었다. 이로 인하여 실제 화재조사데이터 상 “상업장비” 대분류의 “사진현상/인화기”의 화재는 2007~2012년 사이에 단 한건만 존재한다. 또한, 반대로 디지털 카메라나 스마트폰 등 스마트 기기(smart device) 등에 대한 분류는 국가화재분류체계에 포함되어 있지 않다. 따라서 변화하는 시대에 적응하기 위하여 발화관련 기기를 분석하고 새로운 입력 키워드를 추출하여 기존 국가화재분류체계를 보완할 필요가 있다.

본 논문에서는 해당 화재에 대해 적당한 입력 키워드(input keyword)가 존재하지 않아 기타로 입력하는 경우를 기타화재(other-fires)로 정의하고 화재조사데이터에서 이러한 기타화재를 분석하여 새로운 국가화재분류체계에 개정을 위한 새로운 입력 키워드를 추출한다. 이를 위하여 형태소 분석(morphological analysis)을 이용한 화재정보시스템 입력 키워드 추출 방법을 제안한다. 이 방법은 화재에 직간접 원인이 될 수 있는 발화관련 기기 대분류(main-category), 발화관련 기기 소분류(sub-category) 내의 기타화재의 빈도 분석과 기타화재에 대한 “제품명”, “발화개요”의 형태소 분석 및 명사 추출, 추출된 명사 빈도 분석을 통한 새로운 입력 키워드 추출을 수행한다.

본 논문의 구성은 다음과 같다. 2장에서는 화재조사데이터의 전처리 방법을 설명한다. 3장에서는 새로운 입력 키워드 추출 방법에 대하여 설명하고 이 방법에 따라 “발화관련 기기” 속성을 분석하여 새로운 입력 키워드를 추출한다. 4장에서는 결론과 향후연구를 기술한다.

## 2. 전처리

본 논문에서는 입력 키워드 추출의 전처리 작업으로 한국어 형태소 분석기를 이용하여 전체 화재조사데이터의 “발화개요”와 “제품명”을 형태소 분석하고 명사를 추출하였고, 추출된 명사에 대한 화재 빈도를 분석하여 불용어를 추출하였다.

### 2.1 형태소 분석

화재조사데이터의 “발화개요”는 자연어(natural language) 문장이며, “제품명”은 단순명사 혹은 복합명사로 이루어져 있다. 이러한 자연어 문장이나 단순·복합 명사를 컴퓨터 상에서 의미의 최소단위인 형태소로 파싱(parsing)하고 품사를 판별하는 것을 형태소 분석(morphological analysis)이라고 한다<sup>(3)</sup>. 형태소 분석 추출 및 과정은 1) 전처리 모듈, 2) 형태소 분석 및 명사 추출 모듈, 3) 빈도 분석 모듈, 4) 후처리 모듈로 구성되어 있다. 전처리 모듈에서는 전체 화재조사데이터의 “발화개요”와 “제품명”을 대상으로 전처

리를 수행한다. 본 논문의 전처리 모듈에서는 한글, 영문 알파벳, 숫자를 제외한 문자를 제거하고, 형태소 분석 시 오류를 방지하기 위하여 띄어쓰기 오류나 오타를 미리 수정하였다. 또한, 개인정보 보호를 위하여 정규 표현식(regular expression)을 이용하여 혹시 포함되었을지도 모를 개인 정보(주민번호, 전화번호 등)를 제거하였다. 두 번째 형태소 분석 및 명사 추출 모듈은 형태소 분석기를 이용하여 “발화개요”나 “제품명”의 형태소를 분석하고 명사를 추출한다. 본 논문에서는 형태소 분석기로 한나눔(Hananum)을 사용하였다. 한나눔은 한국과학기술원 시맨틱 웹 첨단연구센터(SWRC)에서 개발한 형태소 분석기로 자동화된 명사 추출 프로세스를 제공한다<sup>(4)</sup>. 본 논문에서는 한나눔에서 제공하는 명사 추출 workflow를 이용하여 “발화개요”와 “제품명”에서 추출된 형태소 중 명사만 추출하여 저장하였다. 세 번째 빈도 분석 모듈에서는 추출된 명사 별로 화재 발생 빈도를 계산한다. 마지막 후처리 모듈에서는 화재발생 빈도가 2건 이하인 명사를 제거한다. 본 논문에서는 이렇게 분석된 결과를 이용하여 불용어 추출 및 새로운 입력 키워드 추출을 수행하였다.

### 2.2 불용어 추출

불용어란 정보 검색이나 텍스트 마이닝에서 문서나 문장 내에 높은 빈도로 출현하지만 특별한 의미가 없는 단어로 이를 미리 제거하여 검색 및 분석의 효율을 높일 수 있다. 일반적인 검색엔진에서 불용어는 “나”, “그”, “저”와 같은 대명사나 “은”, “는”, “이다”와 같은 조사이다. 그러나 특정 분야에 특화된 검색엔진이나 텍스트 마이닝 분석 시에는 해당 분야에서 고빈도로 발생하는 단어를 불용어로 선정하여 분석의 효율을 높이기도 한다.

본 논문에서는 전체 화재조사데이터의 “제품명”과 “발화개요”를 대상으로 형태소 추출 결과의 빈도 분석을 수행하여 빈도가 높은 50개 명사를 추출하였다. 화재 빈도가 높은 50개의 명사에는 “화재”(315,850건), “발화”(189,242건), “전기”(69,096건), “쓰레기”(60,343)와 같이 이미 알려진 화재 관련된 명사와 “추정”(175,277건), “조사”(90,654건), “진술”(87,127)과 같은 화재원인 조사 설명에 주로 사용되는 명사가 포함되었다. 이러한 명사를 본 논문의 화재조사데이터 분석 및 새로운 입력 키워드 추출을 위한 불용어로 분류하였다<sup>(3,5,6)</sup>.

## 3. 새로운 입력 키워드 추출

본 논문에서는 기타화재에서 새로운 입력 키워드를 추출하기 위하여 Figure 1과 같은 프로세스를 개발하였다. 새로운 입력 키워드 추출 프로세스는 Figure 1에서처럼 1) 발화관련 기기분류 분석 및 2) 기타화재 빈도/비율 분석을 통하여 분석 대상 발화관련 기기 대분류를 추출한다. 분석 대상으로 추출된 발화관련 기기 대분류에 대하여 3) “제품

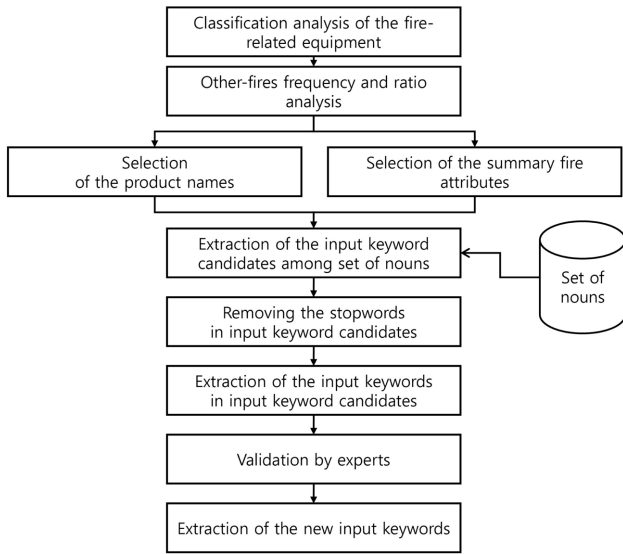


Figure 1. An extraction process of new input keyword.

명”이나 “발화개요” 중 분석에 사용할 속성을 선택하고 4) 선택된 속성에 포함된 명사 집합(set of nouns)을 입력 키워드 후보군(input keyword candidates)으로 추출한다. 5) 입력 키워드 후보군에서 불용어를 제거한 후 6) 입력 키워드 후보군에 대한 화재 빈도 분석을 수행하고 7) 전문가의 검증(validation by experts)을 통하여 8) 새로운 입력 키워드를 추출한다.

본 논문의 분석에는 소방방재청의 화재정보시스템에 의하여 수집된 2007년 1월 1일부터 2012년 12월 31일까지

의 총 273,821건의 화재조사데이터를 사용하였다.

### 3.1 “발화관련 기기” 분석

국가화재분류체계 매뉴얼<sup>(1)</sup>에서 “발화관련 기기”란 “발화에 관련된 불꽃 또는 열을 발생시킨 기기 또는 장치, 제품을 말한다. 발화열원을 제공한 기기 또는 제품에 대한 분석을 통하여 화재예방 대책 및 제조물 안전체계 개선에 활용함을 목적으로 한다.”라고 정의한다. “발화관련 기기”는 “계절용기기(seasonal appliance)”, “생활기기(life appliance)” 등 기기나 제품의 특징에 따른 대분류와 “에어컨”, “선풍기”, “다리미” 등 각각 대분류를 구분하는 소분류로 이루어져 있다. Table 1은 “발화관련 기기” 대분류 별 소분류, 화재빈도, 기타화재 빈도 및 기타화재 비율을 나타낸 것이다.

Table 1에서 “발화관련 기기” 대분류가 “비어있음(empty)”으로 표시된 화재는 전체 화재의 약 59.85%인 163,892건으로 화재조사데이터 작성 시 발화관련 기기 대분류 속성의 값을 “해당 없음”으로 입력한 경우이다. 국가화재분류체계 매뉴얼에서는 발화관련 기기가 없는 화재인 경우에 “해당 없음”을 입력하도록 유도하고 있다. 이외 “기타(etc)”, “미상(unknown)” 화재는 실제로 해당 화재의 발화관련 기기 분류가 불명확하거나, 적합한 분류 기준이 없는 경우로 볼 수 있다. 본 논문에서는 발화관련 기기에 대한 정보가 없는 “비어있음”과 발화관련 기기의 정보가 불확실한 “기타”, “미상” 대분류는 분석 대상에서 제외하였다. “기타”, “미상”과 “비어있음”을 제외한 대분류는 “계절용 기기” 등 총 13가지이다. 이 각각의 대분류에는

Table 1. Status of the Fire-related Equipment Category

Main-category of the fire-related equipment	No. of sub-categories	No. of fires	No. of other-fires	Ratio of other-fires
Seasonal appliance	17	12,798	2,108	16.47%
Agricultural equipment	5	1,080	824	76.30%
Wiring/wiring appliance	9	14,299	2,037	14.25%
Business machine	4	288	55	19.10%
Industry equipment	31	13,070	2,461	18.83%
Commercial facilities	8	309	212	68.61%
Life appliance	11	2,727	1,045	38.32%
Video, audio apparatus	6	1,612	637	39.52%
Medical apparatus	6	139	81	58.27%
Electrical installation	20	7,703	1,251	16.24%
Lighting, sign	10	7,046	960	13.62%
Kitchen equipment	20	17,345	1,525	8.79%
Vehicle, vessel component	24	16,414	3,379	20.59%
Etc.	0	7,223	7,223	100.00%
Unknown	0	7,873	-	0.00%
Empty	0	163,892	-	0.00%
Total		273,818	23,798	8.69%

소분류로 “기타”가 존재한다. 이는 해당 대분류에 속하긴 하지만 소분류에 적당한 발화관련 기기 분류가 없는 기타 화재를 의미한다. 본 논문에서는 이렇게 “발화기기” 대분류는 “비어있음”이나 “기타”, “미상”이 아니지만, 소분류에 “기타”로 표기한 기타화재를 대상으로 하며, 이에 대한 분석을 통하여 새로운 입력 키워드를 추출하였다.

본 논문에서는 총 화재 빈도가 높은 대분류인 “차량, 선박부품”, “계절용 기기”, “배선/배선기구”와 기타화재 비율이 상대적 높고 기타화재의 빈도가 100건 이상인 “농업용 장비”, “상업용장비”, “영상, 음향기기”에 대하여 사전 분석을 수행하여 새로운 입력 키워드 추출 실험 대상 대분류를 선택하였다. 사전 분석결과 “발화개요”에 비하여 상대적으로 정보의 양이 적은 “제품명”에서 유용한 형태소 분석결과가 도출된 “농업용장비”와 “계절용기기”를 “제품명”을 이용한 입력 키워드 추출 대상 대분류로 선택하였다. 대분류 중 가장 화재 빈도가 높은 “차량, 선박부품”의 경우 형태소 분석결과 빈도순으로 상위 5개의 명사가 특정 회사이름이나 특정 차량 명이었고, “배선/배선기구”과 “산업장비”의 경우에는 추출된 명사 중 상위 5건이 “배선”, “전기”와 같은 불용어이므로 대상에서 제외하였다. 또한, 상대적으로 “제품명”의 수가 적은 “상업장비”(58건)와 “영상, 음향기기”(48건)는 “발화개요”분석 대상 대분류로 선택하였다.

3.2 제품명 분석 및 입력 키워드 추출 결과

3.2에서는 “농업용장비”와 “계절용기기” 대분류에 대하여 “제품명(product names)” 분석을 수행하고 입력 키워드를 추출하였다.

3.2.1 “농업용장비” 대분류 분석

“농업용장비” 대분류에는 총 5개의 소분류가 있다. “농

Table 2. Top 5 Noun Extraction Results for Product Names of “Agricultural Equipment” and “Seasonal Appliance” Categories

Rank	Agricultural equipment category		Seasonal appliance category	
	Noun	Frequency	Noun	Frequency
1	Dryer	14	Electricity	27
2	Spray	7	Boiler	25
3	Grain	7	Water heater	12
4	Horsepower	7	Heater	11
5	Tractor	5	Firewood	9

업용장비” 대분류에는 기타화재로 분류되는 화재의 비율이 76.30%인 824건이며, 다른 소분류인 “발효기”(1.11%), “수확/탈곡기/정미기”(8.61%), “양수기”(11.76%), “파쇄기”(2.22%)이다. “농업용장비” 대분류에서 “제품명”을 대상으로 형태소 분석을 수행하여 명사를 추출한 결과는 Table 2의 왼쪽 “agricultural equipment category”와 같다. 분석결과 가장 높은 빈도로 추출된 명사는 “건조기(dryer)”이며, 다음은 “분무기(spray)”, “곡물(grain)”, “마력(horsepower)”, “트랙터(tractor)” 순이다. 이 중 가장 빈도가 높은 “건조기”를 새로운 입력 키워드 후보로 선택하고 분석을 수행하였다.

제품명에 “건조기” 명사를 포함하고 있는 화재를 “건조기” 화재라 정의하고 기타화재 내에서 “건조기” 화재를 재분류하였다. 재분류 결과는 다음 Figure 2의 왼쪽 “agricultural equipment category”와 같다. Figure 2에서 “previous classification”은 기타화재를 제외한 기존 분류체계에 의해서 기 분류된 화재를 의미하며, “other-fires”는 기타화재, “dryer”는 “건조기” 화재를 의미한다. Figure 2에서처럼 “건조기” 화재는 전체 “농업용장비” 화재의 8.43%인 91건으로 “건조기” 화재 분류 이후 76.30%였던 기타화재의 비

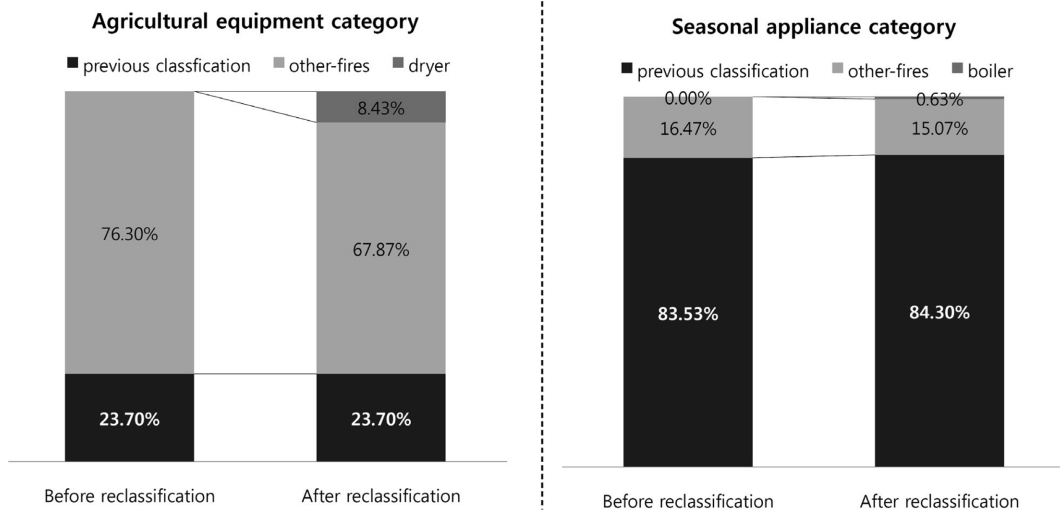


Figure 2. Before and after reclassification of “agricultural equipment” and “seasonal appliance” main categories. “Previous classification” means already classified fires, except other-fires.

율은 67.87%로 감소하였다. “농업용장비” 대분류에서 “건조기”는 매년 평균 15건 이상 화재가 발생하며, 이는 평균 2건 발생하는 발효기, 매년 평균 4건 발생하는 “피쇄기”보다 높은 화재 빈도를 보이는 것이며, “수확/탈곡기/정미기”와 비슷한 빈도를 보이므로, “농업용기기” 대분류에서 “건조기”를 새로운 입력 키워드로 추출하였다.

3.2.2 “계절용기기” 대분류 분석

“계절용기기” 대분류는 “가스난로/스토브”, “기습기” 등 총 17개의 소분류가 있으며, 기타화재의 비율이 16.47%로 총 2,108건이다. “계절용기기”의 경우 총 화재의 빈도는 연도별로 별 차이가 없으나, 기타화재는 2007년 199건에서 2008년 307건, 2012년 488건으로 매년 증가추세이므로 매년 다른 소분류에 속하지 못해 기타로 분류되는 화재의 빈도가 증가한다고 볼 수 있다. “계절용기기” 대분류에서 제품명을 대상으로 형태소 분석하여 명사를 추출한 결과는 Table 2의 오른쪽 “seasonal appliance category”와 같이, “전기(electricity)”, “보일러(boiler)”, “온수기(water heater)”, “히터(heater)”, “화목(firewood)” 순이다. 추출된 명사 중 가장 빈도가 높은 것은 “전기”이지만 “전기”는 불용어 목록에 포함되어 있으므로 분석에서 제외되었다. 본문에서는 “전기” 다음으로 화재빈도가 높은 “보일러(boiler)”에 대하여 재분류를 수행하였다. 재분류 결과는 “보일러” 화재는 총 80건 추출되었으며, 이는 전체 “계절용기기” 화재의 0.63%, 기타화재의 4.15%이다. 재분류이전과 이후를 비교한 그래프는 Figure 2의 오른쪽 “seasonal appliance category”와 같으며 Figure 2에서 “보일러” 화재는 “boiler”로 표기된 것이다.

“보일러” 화재의 경우 “농업용기기”의 “건조기” 화재에 비하여 다른 특징을 가진다. 이미 “계절용기기” 대분류에 포함된 소분류 중에 “가정용보일러”가 존재한다. 실제 화재조사데이터의 제품명을 분석한 결과 기타화재에 속하는 보일러는 “가정용보일러”가 아닌 “화목 보일러” 등 상업용 혹은 산업용보일러이었다. 즉, “가정용보일러”를 제외한 보일러 화재를 기타로 분류한 것으로 추정된다. 그러므로 “가정용 보일러” 소분류를 일반적인 “보일러”로 변경하여 가정용과 다른 보일러를 통합관리하거나, 별도의 “산업용보일러” 새로운 입력 키워드로 추가할 필요가 있다.

3.3 “발화개요” 분석 및 입력 키워드 추출

3.3에서는 “상업장비”와 “영상, 음향기기” 대분류에 대하여 “발화개요” 분석을 수행하고 입력 키워드를 추출하였다. 발화개요 분석의 경우에는 화재조사데이터 중 발화개요 속성 값이 존재하는 2007년 1월 1일부터 2012년 6월 30일까지 데이터를 대상으로 하였다.

3.3.1 “상업장비” 대분류 분석

“상업장비” 대분류에는 “가스온수기”, “노래방화재” 등

Table 3. Top 5 Noun Extraction Results for Summary Fire Attributes of “Commercial Facilities” and “Video, Audio Apparatus” Categories

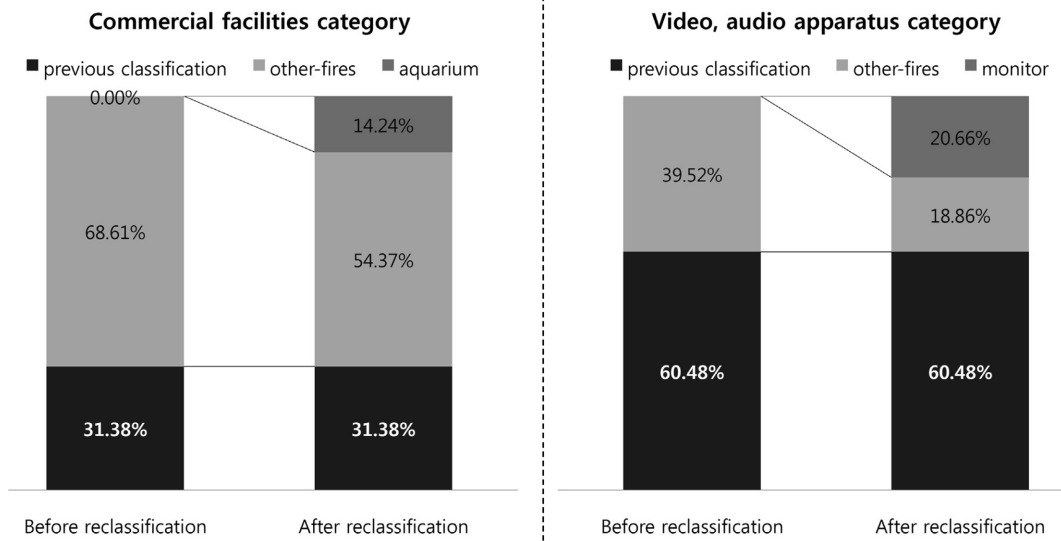
Rank	Commercial Facilities Category		Video, Audio Apparatus Category	
	Noun	Frequency	Noun	Frequency
1	Aquarium	44	Monitor	333
2	Fire Extinguishing	33	Singing	288
3	Functioning	26	Karaoke	235
4	Dryer	23	Business	222
5	Machine	20	Power	146

총 7개의 소분류가 있다. “상업장비”에는 2007년부터 2012년까지 단 한 건의 화재만 존재하는 소분류(“놀이기구”, “사진현상/인화기”)가 2개 있으며, 기타화재가 총 212건으로 총 “상업장비” 화재의 과반수 이상인 68.61%를 차지한다. “상업장비” 대분류에서 “발화개요”를 대상으로 형태소 분석하여 명사를 추출한 결과는 다음 Table 3의 왼쪽 “commercial facilities category”와 같다. 분석결과 가장 빈도가 높게 추출된 명사는 “수족관(aquarium)”이며, “소화(fire extinguishing)”, “작동(functioning)”, “건조기(dryer)”, “기계(machine)” 순으로 추출되었다. 본문에서는 가장 화재 빈도가 높은 “수족관”을 입력 키워드 추출 후보로 선정하여 화재를 재추출하였다.

3.2절의 “제품명” 분석과 달리 “발화개요”는 자연어 문장이고, 현재 형태소 분석기에서는 추출된 명사에 대한 극성 분석을 수행하지 않으므로 “수족관” 화재로 추출된 화재가 실제로 “수족관” 화재인지 “발화개요”를 확인하였다. 기타화재 전체 확인 결과 “수족관” 화재로 분류된 모든 화재는 횃집, 상가 내 식당 등에 설치된 수족관의 배선, 부속품 등에서 일어난 화재임을 확인하였다. 실제 발화 개요의 예는 다음과 같다.

- ✓ 수족관에 설치된 온수히터, 공기주입기 모터 등 뭉쳐진 전선 과열, 과부하로 전선피복에 착화 추정
- ✓ ○○아구찜 수족관 전기모터 콘센트에서 단락되어 전선 피복에 착화 발화됨
- ✓ 화재발생 대상물을 1층 점포에서 취침 중이던 ○○○(여/○○년생)가 외부에서 들려오는 특이한 소리에 기상하여 확인한 바 점포 외부에 설치되어 있는 수족관에 발생한 화재를 목격함

Figure 3의 “commercial facilities category”는 “상업장비” 대분류의 재분류 이전과 이후 기타화재와 새로 입력 키워드로 추출된 “수족관” 화재의 비율을 비교한 것으로 그림에서 “aquarium”은 “수족관” 화재를 의미한다. “수족관” 화재는 전체의 14.24%인 44건이 추출되었다. 특히, “수족관” 화재의 경우 “상업장비” 대분류 중 “기타화재”, “노래방기기” 소분류 다음으로 화재빈도가 3위이며, “가스



**Figure 3.** Before and after reclassification of “commercial facilities” and “video, audio apparatus” main categories. “Previous classification” means already classified fires, except other-fires.

온수기”, “놀이기구”, “사진현상/인화기”, “전자오락기” 소분류보다 화재 빈도가 높아 새로운 입력 키워드로 추출하였다.

인 333건으로 44.91%의 텔레비전 다음으로 2위이므로, 새로운 입력 키워드로 추출하였다.

3.3.2 “영상, 음향기기” 대분류 분석

“영상, 음향기기” 대분류는 총 6건의 소분류를 가지며, 기타화재는 637건으로 전체의 39.52%로 “텔레비전” 다음으로 화재빈도가 높다. “영상, 음향기기” 대분류에 속하는 “발화개요” 속성에 대한 형태소 분석 결과는 Table 3의 오른쪽 “video, audio apparatus category”와 같다. 가장 높은 비율로 추출된 명사는 333건의 “모니터(monitor)”이며, “노래(singing)”, “노래방(karaoke)”, “영업(business)”, “전원(power)” 순으로 추출되었다. 본 논문에서는 “영상, 음향기기” 대분류에 속하는 화재의 발화 개요에서 추출된 명사 중 가장 많이 출현한 “모니터”를 입력 키워드 후보로 하여 추가 분석을 수행하였다. 재분류 결과는 Figure 3의 오른쪽과 같으며, 여기서 “모니터” 화재는 “Monitor”이다.

기타화재 중 “모니터” 화재로 추출된 총 333건은 “텔레비전” 다음으로 높은 빈도이며, 기존 기타화재의 50% 이상이 “모니터” 화재로 분류된 것이다. 재분류 이전과 이후를 비교한 결과 “모니터”가 포함된 화재는 전체의 20.66%

4. 결론 및 향후계획

본 논문에서는 발화관련 기기에서 새로운 입력 키워드를 추출하기 위하여 2006년 개정된 국가화재분류체계에 의해 수집된 2007년부터 2012년까지 6년간의 화재조사데이터의 “발화 개요”와 “제품명” 속성을 분석하여 새로운 입력 키워드를 추출하는 기법을 개발하였다. 또한, 이 방법을 이용하여 “발화관련 기기” 대분류와 소분류의 빈도 분석을 수행하고, 4개의 “발화관련” 기기 대분류를 선택하여 “발화관련” 기기의 새로운 입력 키워드를 추출하였다. 새로운 입력 키워드 추출 결과는 Table 4와 같다. Table 4와 같이 “제품명”을 이용한 새로운 입력 키워드 추출 결과는 “농업용장비” 대분류에서 “건조기”, “계절용기기” 대분류에서 “보일러”이며, “발화개요”를 이용한 새로운 입력 키워드 추출 결과는 “상업장비”에서 “수족관”, “영상, 음향기기”에서 “모니터”이다. 이렇게 추출된 키워드를 이용하여 4개 대분류에 포함된 기타화재 총 3,808건 중 14.39%인 548건의 화재를 재분류 하였다. 특히, “영상, 음향기기”

**Table 4.** List of New Input Keywords

Category	New Input Keyword	No. of Other-fires	No. of Reclassified Fires	Ratio of Reclassified Fires
Agricultural Equipment	Dryer	824	91	11.04%
Seasonal Appliance	Boiler	2,108	80	3.80%
Commercial Facilities	Aquarium	212	44	20.75%
Video, Audio Apparatus	Monitor	664	333	50.15%
Total		3,808	548	14.39%

의 경우 기타화재의 50.15%가 재분류 되었다. “상업장비”의 “수족관”과 “영상, 음향기기”의 “모니터”는 2007년부터 2012년 상반기까지의 데이터만 이용하였으므로 실제 화재의 빈도는 더 높을 것이다.

본 논문에서는 제안한 방법의 타당성 검증을 위하여 4건의 “발화관련 기기” 대분류를 대상으로 분석을 수행하였다. 향후 이 방법을 다른 “발화관련 기기” 대분류나 다른 속성에 적용하여 추가 입력 키워드를 추출하고 화재관련 전문가가 검토한 후 국가화재분류체계 개정에 활용할 수 있다. 또한, 이렇게 추출된 새로운 입력 키워드는 실제 화재의 발화관련 기기의 “제품명”이나, 발화 정보를 기술한 “발화개요”에서 추출된 묵시적 정보(implied information)를 이용한 것이므로 최신 화재조사데이터에 적용하여, 최근 화재 발생 트렌드를 반영할 수 있다.

본 논문에서 제안한 새로운 입력 키워드 추출 방법에 사용한 “발화개요”나 “제품명” 속성은 화재조사데이터 내의 다른 속성과 달리 비정형 데이터이다. 그렇기 때문에 한국어와 외국어를 혼용하고, 외래어의 경우 발음에 따라 다양한 형태로 입력이 존재하며, 오타 등의 입력 오류가 자주 발생된다. 이러한 경우 제안한 방법을 이용하여 분석하였을 때, 표기에 따라 추출된 명사가 달라지므로, 명사 발생 빈도가 왜곡되어 키워드로 추출되지 못하는 한계를 지니고 있다. 그러므로 기 수집된 화재조사데이터를 기반으로 화재의 대상이 되는 제품에 대한 계층 관계, 유의 관계 등 다양한 의미 관계를 정의하는 화재데이터 워드넷(WordNet)<sup>(4,6)</sup>을 구축하고, 이를 새로운 키워드를 추출에 이용하는 방법에 대한 연구가 필요하다.

## 후 기

본 연구는 소방방재청 차세대핵심소방안전기술개발사업

단에서 지원하는 2012년 차세대핵심소방안전기술개발사업(과제번호: 2012-NEMA14-009-01010000-2013)으로 이루어진 결과입니다.

## References

1. J. W. Ryu, Y. J. Kim, E. J. Kim and M. W. Kim, “A Generation Method of a Fire Probability Prediction Model Based on Weather Forecast”, Journal of Korea Information Science Society: Computing Practices and Letters, Vol. 20, No. 2, pp. 68-79 (2014).
2. W. K. Mun, “National Fire Classification System Manual”, Fire Investigation Team, National Emergency Management Agency of Republic Korea, Korea (2006).
3. P. Cimiano, “Ontology Learning and Population from Text: Algorithms, Evaluation and Applications”, Springer, Germany, pp. 35-55 (2006).
4. Semantic Web Research Center, “Hananum Korean Morphological Analyzer User’s Manual”, Korea Advanced Institute of Science and Technology, Korea (2011).
5. C. D. Manning, P. Raghavan, H. Schütze and E. Corporation, “Introduction to Information Retrieval”, Cambridge University Press Cambridge, United Kingdom, pp. 27-28 (2008).
6. E. J. Kim and J. W. Ryu, “Noun Extraction and Analysis on Fire Summary”, Proceedings of 2013 Spring Annual Conference, Korean Institute of Fire Science & Engineering, pp. 91-92 (2013).