

Comparing the Usages of Vocabulary by Medias for Disaster Safety Terminology Construction

Jung-Eun Lee[†] · Tae-Young Kim^{**} · Hyo-Jung Oh^{***}

ABSTRACT

The rapid response of disaster accidents can be archived through the organical involvement of various disaster and safety control agencies. To define the terminology of disaster safety is essential for communication between disaster safety agencies and well as announcement for the public. Also, to efficiently construct a word dictionary of disaster safety terminology, it's necessary to define the priority of the terms. In order to establish direction of word dictionary construction, this paper compares the usage of disaster safety terminology by media: word dictionary, new media, and social media, respectively. Based on the terminology resources collected from each media, we visualized the distribution of terminology according to frequency weights and analyzed co-occurrence patterns. We also classified the types of terminology into four categories and proposed the priority in the construction of disaster safety word dictionary.

Keywords : Disaster Safety, Terminology Resource, Word Dictionary, Standardization

재난안전 용어사전 구축을 위한 미디어별 어휘 사용 양상 비교

이 정 은^{*} · 김 태 영^{**} · 오 효 정^{***}

요 약

재난사건의 신속한 대응은 다양한 분야의 재난안전 유관기관들이 유기적으로 관계함으로써 가능하며, 이 때 사용되는 재난용어의 표준화는 필수적이다. 따라서 재난안전 분야의 전문 용어사전 구축은 각 유관기관 간의 의사소통 및 국민에게 명확한 정보 전달을 위해 수반되는 핵심 요소이다. 더불어 효율적인 용어사전 구축을 위해서는 구축 대상 용어의 우선순위 선정이 필요하다. 본 연구에서는 구체적인 용어사전 구축방향의 설정을 위하여 용어 사용 주제로 대표되는 미디어를 각각 용어사전, 뉴스미디어, 소셜미디어로 선정하고 어휘의 사용 양상을 비교하였다. 이를 위해 각 미디어에서 수집된 어휘 자료를 바탕으로 미디어별 동시 출현 양상 및 빈도 가중치 분석을 통하여 어휘의 분포를 시각화하였다. 분석 결과를 통해 어휘의 사용 양상에 따라 용어사전의 구축대상이 될 수 있는 어휘의 유형을 4가지로 분류하고, 구축대상 기준별 용어사전 구축의 우선순위 방향성을 제안하였다.

키워드 : 재난안전, 어휘자원, 용어사전, 표준화

1. 서 론

특정 분야에서 활용되고 있는 용어를 분석하고 표준 어휘를 정리한 전문 용어사전은 인접 분야와의 의사소통을 효율적으로 이루어지게 한다[1]. 현재 행정안전부(前, 국민안전처)

와 연계된 재난안전 분야 관련 유관기관은 모두 55개로, 국가적 차원의 재난 관리 및 대응은 다양한 분야의 조직 및 기관들의 긴밀한 협력을 통하여[2] 유기적으로 수행되어야 한다. 그러나 이들 기관에서 재난사건에 대한 대응 시, 동일한 개념을 서로 다른 용어로 지칭하거나 잘못된 용어를 사용하여 의사소통에 차질이 발생할 경우[1], 그 피해는 고스란히 국민이 짊어지게 된다. 특히, 잘못된 정보의 공유와 확산으로 인하여 일반 국민들이 불안과 공포에 사로잡히는 경우가 종종 발생하곤 한다[3]. 따라서 재난안전 분야에 대한 용어 표준화 및 용어사전 구축은 매우 중요한 일이다.

행정안전부는 제각기 운영되던 국가재난정보센터, 재난심리상담센터, 재난훈련관리시스템 등을 2016년 3월에 '대국민 재난안전포털[4]'로 통합하였다. 이를 기반으로 행정안전부는 생활안전정보를 효율적으로 제공하고, 재난안전 유관기관의

※ 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단-재난안전플랫폼기술개발사업의 지원을 받아 수행된 연구임(No. NRF-2016M3D7A1912703).

※ 이 논문은 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2017M3C4A7068186).

† 비 회 원 : 전북대학교 기록관리학과 박사과정

** 비 회 원 : 전북대학교 기록관리학과 박사수료

*** 정 회 원 : 전북대학교 기록관리학과 조교수,
문화융복합아카데미, 의료정보융합연구소 연구원

Manuscript Received : January 5, 2018

First Revision : February 20, 2018

Accepted : March 19, 2018

* Corresponding Author : Hyo-Jung Oh(ohj@jbnu.ac.kr)

정보를 통합·연계하고자 많은 노력을 기울이고 있다. 더불어 재난안전 유관기관의 원활한 정보 공유 및 업무 협조체계 구축을 위하여 재난안전정보 용어사전을 구축하는 사업을 진행 중에 있다. 각 재난안전 유관기관에서 사용되는 용어들은 업무 실무자들이 주로 활용하는 전문용어와 일반인들도 쉽게 이해할 수 있는 일반용어가 혼재되어 있으며 그 의미조차 서로 다르게 기술되어 있다. 따라서 용어와 해당 용어의 정의문이 정리되어 있는 용어사전을 구축하는 일이 급선무이다. 나아가 활용도가 높은 용어사전을 효율적으로 구축하기 위해서는 그 이용 대상 및 어휘의 사용 양상에 따라서 용어의 우선 순위를 고민해야 한다.

재난안전 분야 어휘의 사용 양상 분석을 위해 관련 용어들을 추출할 수 있는 정보자원(information sources)을 검토하였다. 일반적으로 대형 재난사건 및 사고가 발생하게 되면 뉴스미디어가 주도적으로 정보를 제공하고 여론을 형성한다[5]. 더불어 소셜미디어는 공통의 사건을 중심으로 사람들을 연결해주는 실시간 정보 네트워크로, 특정 사회적 이슈에 대한 집단의 의견 및 관심을 반영하고 있다[6]. 따라서 각각 뉴스미디어와 소셜미디어로부터 재난안전 관련 어휘들을 추출한다면 정보제공자 및 정보이용자 관점에서 재난안전 분야의 어휘가 어떻게 활용되는지 분석이 가능하다.

본 연구에서는 재난안전 분야 어휘의 사용 양상을 업무 실무자, 정보제공자, 정보이용자 관점으로 분석하였다. 이를 위해 업무 실무자가 참조하는 재난안전 분야 용어사전과 정보제공자가 작성한 뉴스미디어, 일반이용자가 생성하고 소비하는 소셜미디어를 선정하여 다양한 미디어에 나타난 어휘의 특성을 분석하였다. 나아가 분석 결과를 토대로 향후 진행될 재난안전 분야의 용어 표준화 및 용어사전 구축방향에 도움이 되고자 한다.

본 연구의 선행연구로는 재난안전 유관기관에서 관리하고 있는 어휘자원 현황과 사용 양상을 분석한 연구들이 있다[7, 8]. [7]은 재난안전 분야에서 수집된 어휘자원을 대상으로 정보제공자 및 이용자 관점에서 활용도를 분석하였으며, 이를 통해 어휘 그룹별 특성을 파악하였다. [8]은 수집된 어휘자원의 이용자별 활용도를 분석하고자 소셜미디어에서 재난안전 어휘가 사용되는 양상을 비교·분석하고 표준화 방안을 제시한 바 있다.

2. 연구 방법

2.1 연구모델

본 연구에서는 분석 대상 소셜미디어로 다른 매체보다 정보의 신속성 및 확산성이 높고 자동 수집이 가능한 트위터[6]를 어휘 추출 대상으로 선정하였다. 본 연구에서 제안하는 재난안전 분야 어휘 사용 양상 분석 방법은 Fig. 1과 같이 설계하였으며, 재난안전 분야 용어사전 수집 및 뉴스와 트윗 자동 수집 방안을 기반으로 어휘의 동시 출현 양상과 미디어별 어휘 분포 분석을 수행하였다. 구체적인 내용은 다음과 같다.

첫째, 재난안전 분야에서 활용되고 있는 다양한 용어사전을 수집하여 표제어(entry)를 식별하고 어휘자원을 디지털화

한다. 둘째, 데이터 수집 기간 및 키워드를 선정하여 재난안전 분야와 연관된 뉴스와 트윗을 자동 수집한다. 셋째, 수집된 뉴스 및 트윗의 내용으로부터 어휘를 추출하여 자주 사용된 주요 어휘를 중심으로 분석한다. 넷째, 용어사전, 뉴스, 트위터에서 사용된 어휘 간의 중복도 및 어휘 분포를 분석하여 각 미디어별 어휘 특징을 도출한다. 이러한 과정을 통해 각 미디어에서 사용되는 재난안전 분야 어휘 유형을 분류함으로써, 재난안전 용어사전 구축 시 유용하게 활용할 수 있다.

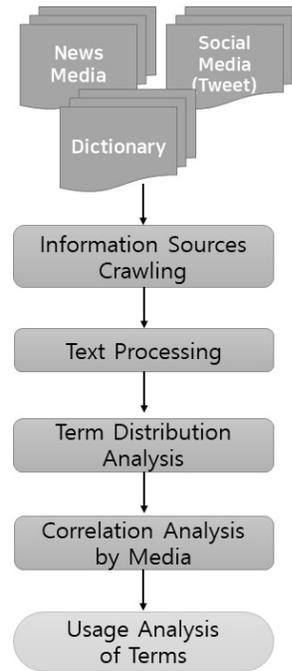


Fig. 1. Research Model

2.2 자료수집

본 연구에서는 재난안전 용어사전 구축에 필요한 어휘자원을 수집하고자 먼저 행정안전부와 연계된 다양한 재난안전 분야 유관기관에서 사용하고 있는 용어사전, 국가표준, 단행본, 법령, 매뉴얼 등으로부터 재난안전 관련 어휘를 온·오프라인으로 수집하고 디지털화하였다. 그 결과 Table 1과 같이 총 404개의 문서로부터 173,284건의 재난안전 관련 어휘들을 수집하였으며, 법령에서 나타난 어휘에 더 많은 가중치를 부여하였다. 또한 뉴스 및 트윗에서 재난안전 관련 어휘를 자동으로 수집하기 위해 수집대상 기간 및 키워드를 선정하였다. 뉴스의 경우에는 행정안전부에서 운영하는 ‘국민재난안전포털’ 웹사이트의 서브메뉴인 ‘안전뉴스’에 게시된 총 23,358개의 재난관련 뉴스를 자동으로 수집하였다. ‘안전뉴스’에 게시된 뉴스는 모두 ‘연합뉴스’에서 제공되는 기사로 단일 출처인 한계를 가지고 있으나, 정부가 재난안전과 관련하여 대국민 서비스를 목표로 운영하고 있는 웹사이트에 게시된 뉴스라는 점에서 재난관련 어휘의 추출 대상으로 선정하였다. 안전뉴스의 수집은 JAVA에서 제공하는 웹 데이터 수집 API를 활용하여 제목 및 내용, 일시 등을 자동으로 추출하였다.

Table 1. Statistics of Word Dictionary

Subject	Number of	Terms Count
Terminology	47	164,310
National Standards	4	120
Book	3	3,059
Law	299	5,046
Manual	48	642
ETC.	3	107
Total	404	173,284

2017.11.15. 기준

재난안전과 관련된 트윗은 2015년에 발행한 재난연감과 재해연보에 기재된 최근 10년간의 ‘대형’ 자연 및 사회재난의 발생기간, 사건명을 기준으로 수집하였다[9, 10]. 선정된 재난 사건은 태풍, 강풍, 풍랑, 호우, 대설 등 총 21건의 자연재난과 ‘고양중합터미널 지하 공사장 화재 사고’, ‘중동호흡기증후군(메르스)’, ‘용흥지구 산불피해’ 등 총 12건의 사회재난이며, 2016년 재해연보에 추가된 ‘태풍 차바’와 ‘경주 지진’을 추가하여 약 395,193개의 트윗을 자동으로 수집하였다[11]. 트윗 수집방법으로는 오픈소스 소프트웨어 저장소인 GitHub[12]에 공개된 트윗 수집 프로그램의 코드를 일부 수정하여 사용하였다. 각 미디어에서 최종적으로 수집된 결과에 대해서는 일부 중복되는 데이터를 제거하는 등 전처리를 수행하였다.

2.3 어휘 추출

본 연구에서는 자동으로 수집된 안전뉴스 및 트윗 데이터를 RHINO 형태소 분석기[13]를 사용하여 안전뉴스 및 트윗 내용으로부터 명사형 어휘를 추출하였다. Table 2는 각 미디어로부터 추출된 어휘 현황에 대한 내용이며, 안전뉴스에서는 2,701,973건(중복제거 32,186건), 트윗에서는 4,411,564건(중복제거 44,452건)의 어휘가 추출되었다. 어휘 수집결과 통계

Table 2. Statistics of Terminology by Media

	Dictionary	Safety News	Tweets
Collection Scope	Dictionary, Standards, Law, Manual etc. for Disaster Safety Agencies	Safety News in the Safety News Portal	Large-scale Natural and Social Disaster
Collection Period	-	2016.03.10. ~2017.09.30	Disaster occurrence date~2016.12.31./2017.09.30. (Typhoon Chaba, Gyeongju Earthquake)
Collection Count	404	23,358	395,193
Total Terms	173,284	2,701,973	4,411,564
Unique Terms	134,615	32,186	44,452
Average Terms per Event	-	115	11

를 통해 소셜미디어보다 뉴스미디어에서 재난안전 분야의 어휘를 더 풍부하게 활용하고 있음을 파악할 수 있었다. 이렇듯 본 연구에서는 각 미디어로부터 추출된 주요 어휘를 대상으로 어휘 특징 도출을 위한 분석을 실시하였다.

2.4 중복도 및 어휘 분포 분석

본 연구에서는 앞서 추출된 어휘를 대상으로 중복 출현 양상과 더불어 각각의 개별 어휘에 대한 분포도를 3차원 그래프로 시각화하여 미디어별로 분석하였다. 3차원 어휘 분포 분석은 각각의 개별 어휘에 대한 가중치 값을 토대로 R[14] 패키지에서 제공하는 scatter 3D 함수를 이용하여 수행하였다. 분석 결과를 바탕으로 미디어별 어휘 사용 양상 및 특징을 파악하였다.

2.5 분석 대상 어휘 유형 분류

본 연구에서는 각 미디어에서 파악된 어휘의 분포 및 특징을 통하여 재난안전 분야의 어휘를 유형별로 분류하였다. 분류한 어휘의 유형은 다음과 같다.

- ① 유형 1 : 모든 미디어에 공통으로 자주 사용되는 고빈도 어휘
- ② 유형 2 : 용어사전 고빈도 어휘
- ③ 유형 3 : 용어사전과 안전뉴스 동시 출현 고빈도 어휘
- ④ 유형 4 : 안전뉴스와 트위터 동시 출현 고빈도 어휘

유형 1은 모든 미디어에서 공통적으로 자주 사용되는 고빈도 어휘로 재난안전 분야 어휘 중 가장 활용도가 높은 어휘 유형을 의미한다. 유형 2는 각 재난안전 유관기관에서 공통적으로 사용되어 중복 횟수가 높은 어휘로 주로 업무 실무자가 사용하는 어휘 유형이다. 유형 3은 용어사전과 안전뉴스에 동시에 출현하는 고빈도 어휘로 주로 국민에게 재난안전 관련 예방, 대응 등의 활동을 알리기 위해 활용되는 어휘 유형이다. 유형 4는 안전뉴스와 트위터에 동시에 출현하는 고빈도 어휘로 일반 이용자가 주로 활용하는 어휘 유형을 의미한다.

본 연구의 궁극적 목적은 재난안전 분야의 어휘들을 위 4가지로 분류하고 유형에 따른 어휘 활용 양상을 파악함으로써 정보제공자 및 정보이용자 모두가 효율적으로 활용할 수 있는 재난안전 용어사전 구축 방향을 제안하는데 있다.

3. 재난안전 분야 어휘 사용 양상 분석

재난안전 용어사전 구축을 위한 미디어별 어휘 사용 양상 분석을 위해 비교 대상인 3개의 미디어(용어사전, 안전뉴스, 트위터)에서 수집된 418,955개 문서로부터 약 720만 개의 어휘를 수집하였다. 이 중에서 중복을 제거한 약 13만 개의 개별(unique) 어휘가 본 연구의 주요 분석대상이다. 분석 방법은 먼저, 각 미디어에서 사용하고 있는 고빈도어 상위 100개 어휘를 추출한 후, 이들 어휘를 기준으로 타 미디어에서 동시 출현하는 양상을 확인하였다. 그리고 용어사전, 안전뉴스, 트위터를 X, Y, Z 축으로 하여 각각의 개별 어휘에 좌표를 부여한 후, 3차원 그래프로 시각화하였다. 예를 들면, X-Y 평면

은 용어사전과 안전뉴스, Y-Z평면은 안전뉴스와 트위터, X-Z평면은 용어사전과 트위터의 어휘가 동시 출현한 영역이다. 각 축은 미디어별 출현 빈도를 해당 미디어의 최빈값으로 나누는 값으로 0부터 1까지 정규화한 가중치이다.

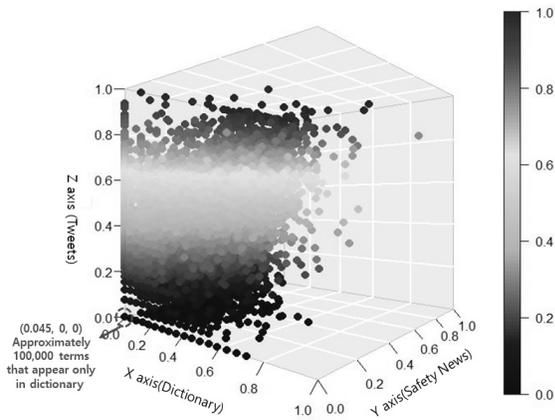


Fig. 2. Terminology Distribution by Media

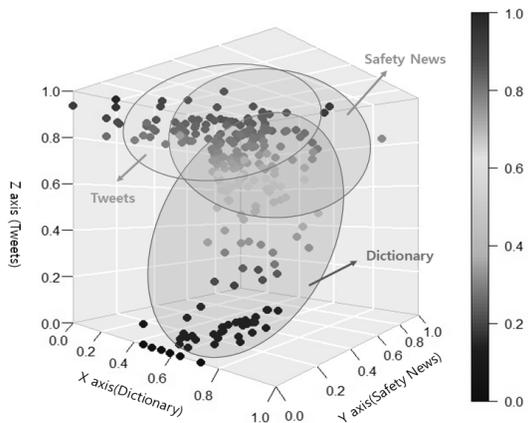


Fig. 3. Top 100 Frequently Terms Distribution by Media

Fig. 2는 분석대상인 13만 개의 어휘에 대한 미디어별 어휘 분포도이다. 전체 어휘 중 (0.045, 0, 0)값을 갖는 어휘가 약 10만 개로, 이는 모두 용어사전에만 출현하는 표제어이다. 이 중 각 미디어별 고빈도어 상위 100개씩 총 300개의 어휘를 중복 제거한 결과, 260개의 어휘가 도출되었으며 이들 어휘의 분포를 시각화하면 Fig. 3과 같다. 260개 고빈도어휘의 빈도 가중치 범위를 살펴보면 용어사전은 0.5~1, 안전뉴스는 0.80~1, 트위터는 0.86~1 사이의 값을 갖는 것으로 파악되었다. 이는 트위터 미디어에 고빈도로 나타난 어휘가 매우 집중된 반면, 용어사전에 나타난 고빈도어휘의 경우 빈도 편차가 큰 것을 의미한다.

3.1 용어사전 고빈도어

용어사전 고빈도어는 각 재난안전 유관기관에서 중복적으로 사용되어 활용도가 높은 어휘로, 실무자가 재난안전정보를 수집하는데 주로 활용하는 어휘이다[7]. 그러나 용어사전

의 어휘는 특정 재난안전 분야의 업무 실무자들이 사용하는 전문용어와 각 유관기관에서 중복적으로 활용되거나 일반인이 정보검색 시 사용하는 일반용어가 혼재되어 있다[8]. 따라서 용어사전의 어휘가 다른 미디어에서 어떻게 활용되는지를 분석함으로써 전문분야 용어의 사용 양상을 파악하는 것이 필요하다.

용어사전으로부터 수집된 어휘는 총 134,615개(중복 제거)이며, 이중 상위 100개 고빈도 용어사전 어휘가 타 미디어에 얼마나 출현하는가에 대한 양상을 Fig. 4와 같이 분석하였다. 용어사전과 안전뉴스에서 동시 출현한 어휘는 총 84개, 용어사전과 트위터에서 동시 출현한 어휘는 48개이다. 이를 통해 용어사전의 고빈도어가 트위터보다 안전뉴스에서 더 많이 사용되었음을 확인할 수 있으며, 이는 용어사전의 어휘가 공적 미디어에서 정보제공 시 주로 활용됨을 의미한다.

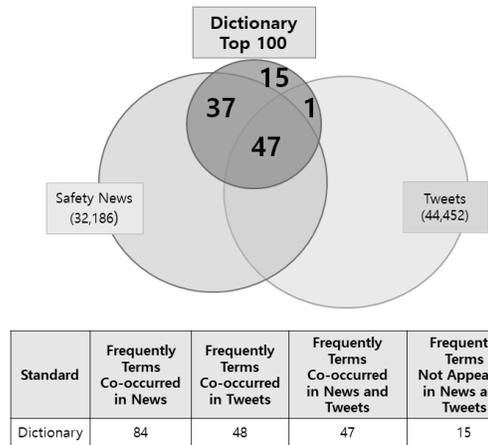


Fig. 4. Top 100 Frequently Terms in Word Dictionary Comparing News and Tweets

용어사전 고빈도어 상위 100개의 어휘 분포도는 Fig. 5와 같다. 상위 100개의 용어사전 어휘의 빈도 가중치 범위는 0.5~1이며, 평균 빈도 가중치는 0.57이다. 반면 용어사전 고빈도어별 Y축(안전뉴스)과 Z축(트위터)의 빈도 가중치를 살펴보면, 모두 0~1까지 매우 다양하게 분포되어 있음을 알 수 있다.

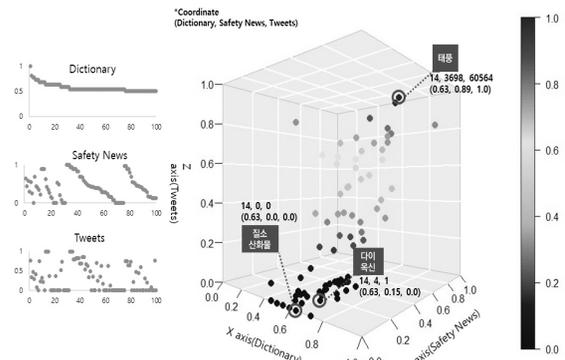


Fig. 5. Top 100 Frequently Terms in Word Dictionary

Table 3. Frequently Terms of Word Dictionary Co-occurred in News and Tweets (17 of 47 Terms)

No.	Dictionary Rank	Dictionary Terms	Dictionary Overlap	News Appearance	Tweets Appearance
1	1	조류	22	362	1,201
2	2	생태계	18	59	23
3	3	지하수	17	206	77
4	3	관계인	17	16	14
5	5	접지	16	10	2
6	5	방사선	16	147	46
7	5	호안	15	56	8
8	5	전선	15	541	970
9	5	등상	15	44	18
10	5	경도	15	25	7
11	13	태풍	14	3,698	60,564
12	13	적조	14	338	28
13	13	자외선	14	3	31
14	13	다이옥신	14	4	1
15	13	공동구	14	51	1
16	23	소독	13	90	821
17	23	분류	13	140	118
...
Co-occurring Terms Average			12.95	1,061	2,050

Table 3은 용어사전을 기준으로 안전뉴스와 트위터에 동시에 출현한 47개 어휘 중 평균 중복도 12.95회 이상인 17개 어휘 현황이다. 해당 어휘를 자세히 살펴보면 같은 가중치 값을 갖는 용어사전 어휘라도 안전뉴스와 트위터에서는 서로 다른 사용 양상을 보이고 있다. 예를 들어 Fig. 5에서 ‘태풍’과 ‘다이옥신’의 경우, 용어사전 중복도는 총 14회로 같은 중복도를 가지고 있다. ‘태풍(좌표:0.63, 0.89, 1.0)’은 용어사전에는 14회, 안전뉴스에는 3,698회, 트위터에서는 60,564회의 출현빈도를 보이는데, 이를 좌표화하면 모두 ‘1’에 가까운 위치에 배치된다. 이를 통해 ‘태풍’이 각 미디어 모두에서 높은 활용성을 가진 어휘임을 알 수 있다. 반면, ‘다이옥신(빈도:14, 4, 1/좌표:0.63, 0.15, 0.0)’은 Z값이 ‘0’이므로, X-Y평면에 배치되어 트위터보다는 용어사전 및 안전뉴스에서 주로 활용되는 어휘임을 알 수 있다.

Table 4는 용어사전에 출현하였으나, 안전뉴스와 트위터에는 출현하지 않은 15개의 용어사전 어휘를 나타낸 표이다. 대부분이 하나의 어휘에 하나의 개념을 나타내는 ‘전문용어’로 이루어져 있음을 확인할 수 있다[8]. 이 중 ‘질소산화물(빈도:14, 0, 0/좌표:0.63, 0, 0)’을 보면, ‘태풍’, ‘다이옥신’과 마찬가지로 용어사전 중복도는 14회이며 기상청, 해양수산부, 안전보건공단 등 9개의 기관에서 중복적으로 사용되는 고빈도 어휘이다. 이상의 분석결과는 용어사전 고빈도어에 일반용어와 전문용어가 혼재되어 있다는 선행연구[8]의 결과와 맥을 같이 하며, 용어사전 고빈도어 중에서도 안전뉴스와 트위터에서 동시 출현하는 어휘는 정보이용자가 일반적으로 활용하는 어휘임을 유추할 수 있다.

Table 4. Frequently Terms of Word Dictionary Not Appeared in News and Tweets (15 Terms)

No.	Dictionary Rank	Dictionary Terms	Dictionary Overlap
1	13	화학적산소요구량	14
2	13	질소산화물	14
3	13	부영양화	14
4	23	용존산소	13
5	23	산성비	13
6	23	기상위성	13
7	30	허용오차	12
8	30	품질보증	12
9	30	연안류	12
10	30	생물다양성	12
11	67	직류	11
12	67	잔류성유기오염물질	11
13	67	기후변화협약	11
14	67	고조	11
15	95	환경호르몬	10
Not Appeared Terms Average			12.2

3.2 안전뉴스 고빈도어

뉴스미디어를 통한 재난안전 관련 보도는 재난의 내용을 보다 많은 사람들에게 전달하여 재난의 피해확산 방지 및 방제가 그 목적이다. 따라서 뉴스를 포함한 공적미디어에서 활용되는 재난안전 분야 어휘는 정보이용자가 곧바로 소비하는 어휘로서 정보 전달력이 높아야 한다. 뉴스미디어에서 활용하는 어휘에 대해 분석함으로써, 정보를 전달하는 ‘제공자’적 측면에서 주로 어떤 어휘자들이 활용되는지 시사점 도출이 가능하다. 이를 위해 안전뉴스에서 주로 사용되는 32,186개(중복제거)의 어휘 중, 고빈도어 상위 100개 어휘를 대상으로 용어사전과 트위터에서의 사용 양상을 분석하였다(Fig. 6).

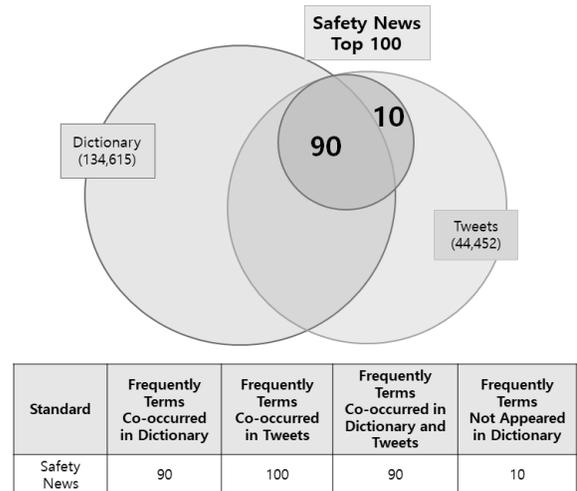


Fig. 6. Top 100 Frequently Terms in News Comparing Dictionary and Tweets

안전뉴스와 용어사전에서 동시 출현한 어휘는 총 90개로 안전뉴스에서 사용된 재난안전 분야 어휘는 대부분 용어사전에도 기술된 어휘인 것으로 파악되었다. 특히, 트위터와의 동시 출현 여부를 살펴보면 안전뉴스의 고빈도어가 트위터에 모두 사용되었다. 이는 공적미디어에서 사용되는 재난안전 분야 어휘가 소셜미디어에서도 활발히 활용됨을 의미한다.

Fig. 7은 안전뉴스 고빈도어 상위 100개를 시각화한 것으로, 어휘의 빈도 가중치 범위는 1~0.80이고, 이들의 평균 빈도 가중치는 0.89이다. 안전뉴스 고빈도어별 X축(용어사전)과 Z축(트위터)의 빈도 가중치를 살펴보면, X축의 가중치 범위는 0~0.63인 반면, Z축의 가중치 범위는 0.45~1이다. 이를 통해 안전뉴스에서 사용하고 있는 어휘는 용어사전과 트위터에서 사용하는 어휘를 함께 활용하고 있으며, 용어사전의 어휘보다는 트위터의 어휘에 근접하게 분포하고 있음을 시각적으로 확인할 수 있다. 어휘 분포 관점에서 보면, 용어사전 고빈도어보다 안전뉴스의 고빈도어 분포도가 좀 더 군집되어 있다. 이는 안전뉴스에서 활용되는 어휘가 용어사전 및 트위터에서도 유사한 빈도로 사용되고 있음을 의미한다.

안전뉴스 고빈도어 상위 100개 중 용어사전과 트위터에 동시에 출현한 어휘는 90개로, 이 중 상위 100개의 평균 출현빈도(3,872회) 이상의 어휘를 Table 5와 같이 추출하였다. 해당 어휘들을 살펴보면, ‘사고’, ‘피해’, ‘안전’, ‘소방’, ‘작업’, ‘조사’, ‘재난’, ‘대책’ 등의 어휘는 각 재난안전 유관기관에서 재난업무 수행을 위해 공통적으로 사용되는 어휘이거나 매뉴얼 및 업무처리지침 등과 관련된 어휘로 볼 수 있다. 따라서 이들 어휘가 각 재난안전 유관기관의 실무에 활용될 때에는 하위 개념을 가지는 다양한 어휘가 존재할 것으로 추측된다. 실제로 Fig. 7과 같이, 안전뉴스에서 12,394회의 고빈도를 보인 ‘작업(빈도:3, 12394, 1038/좌표:0.13, 1.0, 0.75)’의 경우, 용어사전의 중복도는 3회로 낮다. 하지만 ‘작업’을 포함하는 하위 개념의 어휘를 조사해 보면, ‘분진작업’, ‘도복작업’, ‘대별작업’ 등 ‘작업’이란 어휘가 총 25개의 기관에서 440개의 다양한 복합명사로서 활용됨을 확인할 수 있다. 이는 용어사전의 어휘가 실무자가 활용하는 실용용어로서 보다 세부적이고 전문적인 지침을 가리키는 용어가 다수임을 시사한다.

Table 5. Frequently Terms of News Co-occurred in Dictionary and Tweets (24 of 90 Terms)

No.	News Rank	News Terms	News Appearance	Dictionary Overlap	Tweets Appearance	Dictionary Compound Noun
1	1	사고	37,645	4	8,946	280
2	2	피해	20,177	2	20,331	151
3	3	안전	19,469	2	6,323	835
4	4	지진	16,613	12	19,025	291
5	6	소방	12,402	4	764	142
6	7	작업	12,394	3	1,038	440
7	8	화재	12,030	5	5,787	165
8	9	차량	10,277	5	884	133
9	10	조사	7,551	10	3,852	365
10	11	도로	7,443	11	1,039	232
11	12	병원	7,306	2	35,644	71
12	13	재난	6,580	9	4,544	223
13	14	수색	6,568	2	1,062	23
14	15	해경	5,969	1	908	14
15	16	구조	5,770	12	1,572	628
16	17	산불	5,282	10	3,064	53
17	18	신고	4,995	5	1,790	116
18	19	교통	4,950	2	1,499	287
19	20	공사	4,921	2	572	361
20	21	인명	4,678	1	1,097	12
21	22	불	4,577	1	833	1
22	24	대책	4,408	1	10,527	152
23	25	물	4,371	7	1,196	7
24	27	선박	4,003	11	369	240
...
Co-occurring Terms Average			3,872	463	4,082	-

Table 6은 안전뉴스와 트위터에 출현하였으나 용어사전에는 출현하지 않은 어휘들로, 주로 해양사고와 관련된 어휘가 대부분이다. ‘세월호’, ‘선체’, ‘인양’, ‘참사’, ‘침몰’, ‘시신’, ‘유가족’ 등의 표제어는 최근 우리나라에서 발생한 특정 재난인 ‘세월호’와 관련된 어휘로써 정보제공자인 뉴스 및 정보이용자인 트위터에서는 높은 관심을 보이며 고빈도어로 사용된 반면, 실무자가 사용하는 용어사전에는 아직 반영되지 않았음을 알 수 있다. Fig. 7에서 추출된 어휘를 자세히 살펴보면, ‘원전(빈도:0, 1397, 1886/좌표:0.0, 0.78, 0.81)’의 경우 두 가지의 의미로 사용되고 있는데, 원자력 에너지를 이용해 전기를 얻는 기술적 방법을 일컫는 ‘원자력 발전’을 의미하거나, 또는 고리원전, 월성원전 등과 같이 고유명사 뒤에 사용될 때에는 ‘원자력 발전소’의 줄임말로 시설 기관을 의미하며 쓰이고 있다. 안전뉴스에 출현한 ‘원전’의 어휘 수는 총 1,397개로 활용예시를 살펴보면 Table 7과 같다. 안전뉴스에서는 줄임말인 ‘원전’ 이외에도 용어사전의 어휘인 ‘원자력 발전’으로 204회, ‘원자력 발전소’로 13회의 출현빈도를 보였다. 이는 뉴스미디어가 용어사전의 용어를 활용함과 동시에 정보이용자인 소셜미디어에서 사용하고 있는 어휘도 즉각적으로 반영하고 있다는 것을 의미한다.

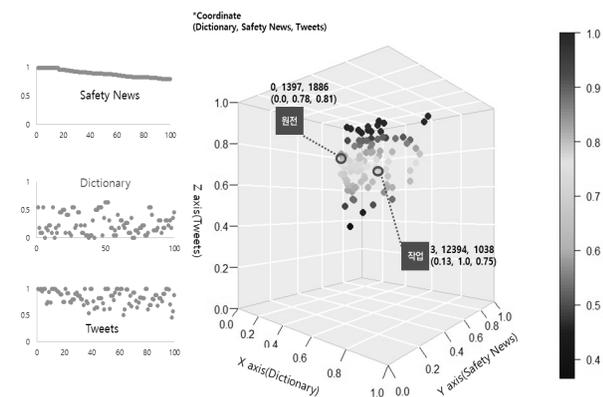


Fig. 7. Top 100 Frequently Terms in News

Table 6. Frequently Terms of News Not Appeared in Dictionary (10 Terms)

No.	News Rank	News Terms	News Appearance	Dictionary Overlap	Tweets Appearance
1	5	세월호	13,305	0	16,138
2	23	선체	4,417	0	139
3	26	인양	4,042	0	321
4	41	참사	2,323	0	5,123
5	63	국민안전처	1,615	0	1,240
6	71	침몰	1,493	0	1,517
7	77	원전	1,397	0	1,886
8	81	시신	1,319	0	646
9	82	유가족	1,318	0	1,150
10	85	추돌	1,276	0	2,096
Not Appeared Terms Average			3,251	0	3,026

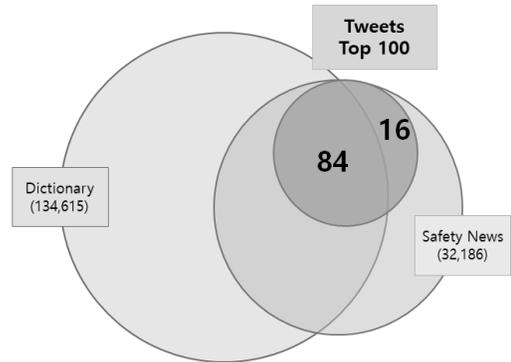
Table 7. 'Nuclear Power' related Terms in News

Terms (Appearance)	Practical Example about 'Nuclear Power'
Nuclear Power (204)	-원전에서 탈피하고 재생에너지를 확대하는 등 에너지 정책의 대전환 필요(16.09.16) -국민의 머리맡에 핵폭탄을 설치한 자들에게 철퇴를 가하고 핵발전정책을 전면 재검토해 원전 제로국가로 가야 한다고 덧붙였다.(16.09.22)
Nuclear Power (1,397)	-종합개선대책에 국가 활성단층 연구와 원전 안전 과제를 신규로 포함하기로 했다.(16.10.04)
Nuclear Power Plant (13)	-월성원전과 고리원전 등 지진 발생 인근 지역의 원전과 주요 산업시설을 점검한 바 있다.(16.09.07) -“한울원자력발전소를 비롯한 모든 원전이 정상 운영을 하고 있다”고 밝혔다(17.03.08) -한수원 “경주 지진에도 모든 원전 정상 운영”(17.03.31)

3.3 트위터 고빈도어

트위터 고빈도어는 주로 일반 이용자가 활용하는 어휘로서, 이를 통해 정보를 소비하는 이용자가 어떠한 재난안전 분야의 어휘에 관심을 가지고 있는지 유추할 수 있다[7]. 트위터에서 주로 사용되는 어휘는 44,452개(중복 제거)이며, 이 중 상위 100개 고빈도어를 대상으로 용어사전과 안전뉴스에서의 사용 양상을 분석하였다(Fig. 8). 트위터 고빈도어 상위 100개 어휘를 기준으로 용어사전 및 안전뉴스에서 동시에 출현한 어휘는 총 84개이며, 트위터에서 사용된 어휘는 안전뉴스에 모두 출현하였다. 이는 트위터의 고빈도어 역시 용어사전 및 안전뉴스에서도 활용되고 있음을 의미한다.

트위터 고빈도어 상위 100개의 어휘 분포를 시각화하면 Fig. 9와 같다. 트위터 어휘의 빈도 가중치 범위는 0.86~1이



Standard	Frequently Terms Co-occurred in Dictionary	Frequently Terms Co-occurred in News	Frequently Terms Co-occurred in Dictionary and News	Frequently Terms Not Appeared in Dictionary
Tweets	84	100	84	16

Fig. 8. Top 100 Frequently Terms in Tweets Comparing Dictionary and News

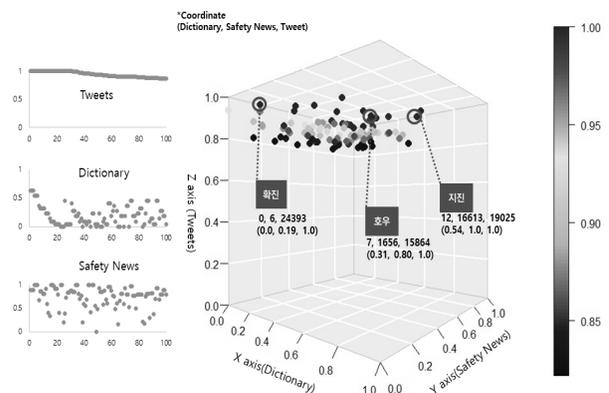


Fig. 9. Top 100 Frequently Terms in Tweets

며, 이들의 평균 빈도 가중치는 0.93으로 높다. 이는 트위터의 고빈도어가 정보이용자의 관심도가 높은 어휘에 집중되어 있으며, 타 미디어의 어휘에 비하여 반복적으로 활용되는 어휘가 많다는 것을 의미한다. 트위터 상위 100개의 어휘의 X축(용어사전)과 Y축(안전뉴스)의 빈도 가중치를 살펴보면, X축의 빈도 가중치 범위는 0~0.63로 매우 낮은 반면 Y축의 범위는 0.80~1이다. 다시 말하면, 트위터에서 사용되는 어휘는 용어사전의 어휘와는 괴리감이 있으며, 안전뉴스에서 활용하는 어휘와는 밀접한 관계가 있음을 시사한다.

트위터의 상위 고빈도 100개의 어휘 중 용어사전 및 안전뉴스에 동시에 출현하는 어휘는 Table 8과 같다. 이들 어휘 중 Fig. 9의 '지진(빈도:12, 16613, 19025/좌표:0.54, 1.0, 1.0)' 과 '호우(빈도:7, 1656, 15864/좌표:0.31, 0.80, 1.0)'를 비교해 보면, 두 개의 어휘 모두 Z값이 '1' 로서 트위터에서 고빈도로 활용되었다. 반면, Y값이 '지진'은 1.0, '호우'는 0.80으로 안전뉴스에서 활용되는 양상에 따라 서로 다른 위치에 배치되었음을 확인할 수 있다. 이는 일반인들이 어떤 재난 분야에 관심을 보이고 있는지를 대변하고 있는데, 매년 반복되는 자연재난

Table 8. Frequently Terms of Tweets Co-occured in Dictionary and News (20 of 84 Terms)

No.	Tweets Rank	Tweets Terms	Tweets Appearance	Dictionary Overlap	News Appearance
1	1	메르스	309,951	3	22
2	2	태풍	60,564	14	3,698
3	3	신종	34,203	1	40
4	5	피해	20,331	2	20,177
5	6	감염	19,536	5	215
6	7	<u>지진</u>	19,025	12	16,613
7	8	격리	16,712	5	40
8	10	<u>호우</u>	15,864	7	1,656
9	11	확산	15,065	8	502
10	12	발생	12,838	5	4,753
11	13	대응	11,820	7	2,796
12	14	대책	10,527	1	4,408
...
Co-occurring Terms Average			10,119.6	4.4	2,642.5

Table 9. Frequently Terms of Tweets Not Appeared in Dictionary (16 Terms)

No.	Tweets Rank	Tweets Terms	Tweets Appearance	Dictionary Overlap	News Appearance
1	4	<u>확진</u>	24,393	0	6
2	9	<u>세월호</u>	16,138	0	13,305
3	24	독감	7,117	0	6
4	29	파스	5,822	0	44
5	32	감염자	5,459	0	1
6	35	참사	5,123	0	2,323
7	46	탄저균	4,139	0	4
8	47	대처	4,079	0	429
9	49	접촉	4,007	0	228
10	52	의료진	3,895	0	121
11	58	감기	3,647	0	7
12	77	산간	2,664	0	297
13	90	추돌	2,096	0	1,276
14	97	완치	1,971	0	4
15	98	의료원	1,957	0	72
16	100	면역력	1,943	0	15
Not Appeared Terms Average			5,903	0	1,134

인 ‘호우’ 보다는 최근 우리나라에서 갑작스러운 이슈가 되고 있는 ‘지진’에 더 많은 관심을 보이고 있음을 시사하고 있다.

트위터와 안전뉴스에 출현하였으나, 용어사전에는 출현하지 않은 어휘는 총 16개로 Table 9와 같다. 추출된 어휘를 살펴보면 주로 의료와 관련된 어휘로 ‘독감’, ‘파스’, ‘감기’, ‘의료

원’ 등은 정보이용자인 일반인이 일상적으로 사용하고 있는 어휘이다. ‘확진’, ‘감염자’, ‘의료진’, ‘완치’, ‘면역력’ 등은 2015년 우리나라에 발생하였던 ‘메르스’ 사건과 관련된 어휘로 해당 재난에 대해 소셜미디어 상에서 큰 주목을 받으며 자주 사용된 반면, 아직 용어사전에는 반영되지 않았음을 알 수 있다.

이상의 내용을 종합해 보면, 트위터의 고빈도 어휘는 정보이용자 측면에서 관심도를 파악할 수 있는 어휘로서, 공적 미디어인 뉴스에서 사용되는 어휘가 즉각적으로 반영됨을 알 수 있다. 반면, 트위터 고빈도어는 업무 실무자들이 활용하는 용어사전과는 다소 괴리가 있으며, 이를 통하여 특정 재난사건과 관련하여 누락된 재난안전 분야의 어휘를 확인할 수 있다.

4. 용어사전 구축 방안

최근 우리나라의 재난안전 사건 유형이 점점 복잡하고 다양해짐에 따라, 정부는 재난에 대한 신속한 대응과 빠른 복구 활동을 위해 여러 가지 방안을 강구하고 있다. 이 중 재난안전 분야의 용어사전을 구축하는 일은 각 유관기관 간의 원활한 의사소통 및 국민에게 명확한 재난안전정보를 전달하기 위해 필수적인 조치이다. 이에 본 연구는 구체적인 용어사전 구축방향 설정을 위하여 각 미디어에서 활용되는 어휘의 사용 양상을 살펴보았다. Fig. 10은 앞서 기술한 각 미디어별 고빈도 상위 100개의 어휘 사용 양상을 빈도 가중치의 평균 값으로 요약한 그림이다. 대체적으로 안전뉴스와 트위터 간에는 상관관계가 높으며, 용어사전과는 다소 괴리가 있음을 확인할 수 있다.

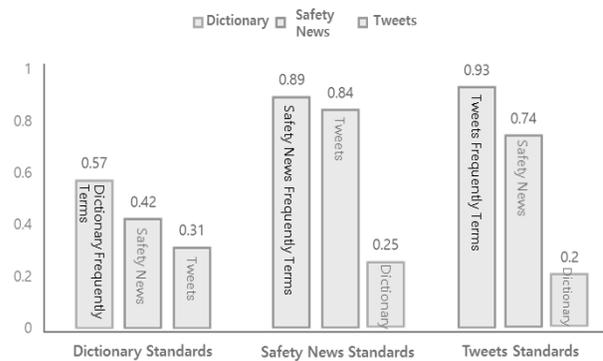


Fig. 10. Average Weights of Top 100 Frequently Terms by Media

앞 장의 분석 결과를 토대로 재난안전 분야의 어휘 양상을 유형별로 분류하면 Fig. 11과 같으며, 각 유형별 어휘의 특징을 살펴보면 다음과 같다.

유형 1은 모든 미디어에서 공통적으로 고빈도로 활용되는 어휘로 재난안전 업무를 수행하는 업무 실무자 및 재난 발생 상황을 전달하는 정보전달자, 정보를 소비하는 정보이용자 모두에게 활용도가 가장 높은 어휘이다. 각각의 미디어 모두

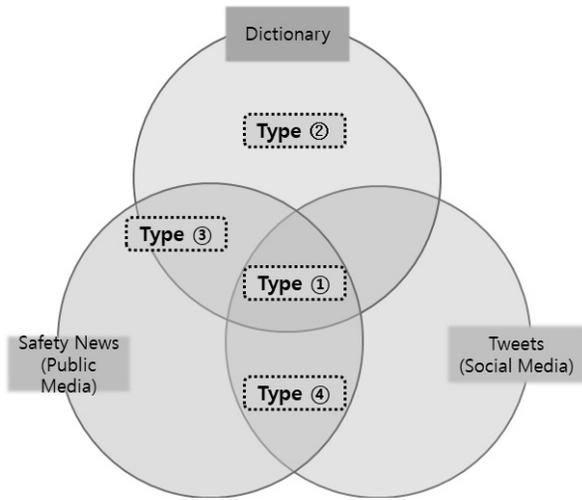


Fig. 11. Type of Disaster Safety Terminology

에서 중복적으로 사용되는 어휘는 누구나 개념을 이해하기 쉽고, 자주 활용되는 어휘로써 용어사전 구축 시 시소러스 구축을 위한 중심어가 될 수 있으며, 이러한 중심어로부터 상위어, 하위어, 관련어 등의 용어 관계를 설정할 수 있다[7]. 또한 정보제공자 및 정보이용자 모두에게 자주 사용되므로 대국민을 대상으로 하는 용어사전 구축 시에 주요 표준화 대상이 될 수 있다[8].

유형 2는 용어사전 고빈도 어휘이다. 각 재난안전 기관의 실무자가 주로 활용하는 어휘 중 중복 횟수가 높고, 특정 재난 발생 시 실제 업무를 처리하는데 유용하게 사용되는 어휘 유형이다. 어휘의 개념이 세부적이고 전문적인 내용을 담고 있어, 공적미디어와 소셜미디어에서는 활용도가 낮은 편이다. 해당 유형의 어휘는 실제 재난안전 관련 업무를 처리함에 있어서 필수적으로 활용되는 어휘로 특히, 유관기관 간의 의사소통 및 정보공유를 위해서는 표준화되어야 한다. 왜냐하면 서로 다른 어휘 정의는 재난안전 업무를 신속하게 처리하는데 큰 장애를 유발할 수 있기 때문이다. 또한 정보이용자 측면에서는 어휘의 복잡성으로 인해 정보에 접근하기가 어려우므로 어휘의 순화가 필요하다.

유형 3은 용어사전과 안전뉴스에 동시 출현한 고빈도 어휘이다. 국민에게 재난안전 예방 활동이나 특정 재난사건 발생으로 인한 대응 및 복구 활동을 알리기 위해 공적미디어에서 주로 활용하는 어휘이다. 따라서 해당 유형의 어휘는 일반인에게 정보를 전달하려는 목적을 가지고 활용되기 때문에 비교적 이해하기 쉬운 어휘이며, 재난사건의 시간적 흐름에 따라 반복적으로 사용되는 특징이 있다. 국민에게 일관된 재난안전 정보를 제공하기 위한 목적으로 표준화 대상이 되어야 한다.

유형 4는 안전뉴스와 트위터에 동시 출현한 어휘이다. 해당 어휘는 재난안전 분야 어휘 중에서도 일반 이용자에게 활발히 활용되는 소비자 중심의 어휘이다. 특정 재난사건 발생 시 정보이용자의 관심이 높아 활용도가 증가하였다가, 재난사건이 종료되는 시점에서는 활용도가 크게 감소하는 특징이 있다. 이처럼 공적미디어와 소셜미디어에서 등장하는 어휘는

복잡화·다양화되는 재난사건과 관련하여 즉각적으로 활용되는 어휘이므로, 이들 어휘를 파악하여 용어사전에 등재되지 않은 새로운 용어를 발굴해야 한다.

재난안전 용어사전의 구축은 정보제공자 및 정보이용자 모두가 재난안전정보를 효율적으로 활용하기 위함이다. 따라서 어휘를 사용하는 대상에 따라 표준화 및 순화 작업을 수행해야 한다. 먼저 Fig. 11의 모든 미디어에서 공통으로 사용되는 고빈도 어휘(유형 1)는 활용 대상에 관계없이 우선적으로 용어사전 구축 대상이 된다. 일반 이용자를 대상으로 한 대국민 서비스를 위한 용어사전의 경우, 정보이용자 측면에서 중요도가 높은 안전뉴스와 트위터 동시 출현 고빈도 어휘(유형 4)와 용어사전과 안전뉴스 동시 출현 고빈도 어휘(유형 3)의 우선순위가 높다. 반면 업무 실무자를 위하여 재난안전 용어사전을 구축한다면, 업무적 측면에서 중요도가 높은 용어사전 고빈도 어휘(유형 2)를 우선적으로 구축 대상으로 삼아야 한다.

5. 결론

재난안전 분야의 용어사전 구축은 재난안전정보의 통합적 관리를 위한 방안의 하나로, 각 재난안전 유관기관에 산재되어 있는 재난관련 용어를 표준화하는데 목적이 있다. 재난안전 용어사전 구축을 위해서는 구축 대상 선정이 우선되어야 한다. 이를 위해 본 연구에서는 각 미디어에서 활용되는 고빈도 어휘를 추출하고, 타 미디어와의 동시 출현 양상을 분석하였으며, 추출된 어휘의 빈도 가중치를 이용하여 어휘의 분포를 시각화하였다.

분석결과 용어사전 구축을 위한 표준화 대상이 될 수 있는 재난안전 분야 어휘 유형을 4가지로 분류하고, 각 유형의 특징을 도출하였다. 먼저 모든 미디어에서 사용되는 어휘(유형 1)는 활용도가 높은 어휘로 대상에 관계없이 용어사전 구축 시에 우선순위를 두어야 한다. 이후 재난안전 업무를 담당하는 실무자를 대상으로 하는 용어사전 구축 시에는 각 재난안전 유관기관에서 중복적으로 활용되는 고빈도 어휘(유형 2)에 우선순위를 두어야 한다. 반면 공적미디어인 뉴스에서 활용되는 어휘는 용어사전과 트위터를 매개하는 어휘로, 대국민을 대상으로 하는 용어사전 구축 시에는 안전뉴스와 트위터에 동시 출현하는 어휘(유형 4)가 우선순위를 갖는다. 그러나 업무 실무자를 대상으로 하는 용어사전 구축 시에는 용어사전과 안전뉴스에 동시 출현하는 어휘(유형 3)가 우선순위를 가진다.

본 연구는 보다 효율적인 재난안전 용어사전 구축을 위하여 재난안전 관련 어휘를 분석하고자 하였다. 이를 위해 각 미디어에서 사용되고 있는 어휘의 양상을 분석하고 재난안전 용어사전의 구축대상이 되는 어휘를 유형화하였다. 이는 용어사전의 활용 대상에 따라 용어사전 구축방향 수립의 기초 자료로 활용될 수 있음에 의의가 있다. 향후 연구방향으로는 본 연구에서 분석 대상으로 한 안전뉴스와 트위터의 수집기간이 2016년 3월 이후라는 한계점을 보완하여 보다 다양한 재난 유형이 포함되도록 수집 범위와 기간을 넓히고자 한다. 나아가 구축된 용어사전을 일반 이용자나 실무자가 재난관련

기록을 검토할 때 그 의미가 명확히 전달되도록 하는 검색 사이트나 플러그인 형태의 설명문 제시 기능 등의 도구로 활용하고자 한다.

References

- [1] The National Institute of The Korean Language, "A Study on the Specialized Terminology," Seoul: The National Institute of The Korean Language, 2007.
- [2] J. H. Lim, T. Y. Park, and S. J. Kim, "Development of a Facet Classification System for Integrated Management and Shared Use of Disaster and Safety Information," *Korea Libray And Information Science Society*, Vol.48, No.4, pp.375-399, 2017.
- [3] J. Y. Gang, H. J. Han, Y. Kim, and H. J. Oh, "Empirical Study of Disaster and Safety Record Information Resources Archiving Strategy based on Automatic Acquired Web Records from Related Organizations," *Journal of The Korean Biblia Society for Library and Information Science*, Vol.28, No.4, pp.223-246, 2017.
- [4] SafeKorea [Internet], https://www.safekorea.go.kr/idsiSFK/index_web.j.
- [5] C. M. Kim and C. I. Cho, "An Exploration of Issue Attention Cycle of Great Disasters," *Seoul Studies*, Vol.16, No.4, pp.147-162, 2015.
- [6] J. Y. Park, P. M. Ryu, and H. J. Oh, "A Collecting Model of Public Opinion on Social Disaster in Twitter: A Case Study in 'Humidifier Disinfectant,'" *KIPS Transactions on Software and Data Engineering*, Vol.6, No.4, pp.177-184, 2017.
- [7] H. C. Jeong, T. Y. Kim, Y. Kim, and H. J. Oh, "A Study on the Utilization Plan of Lexical Resources for Disaster and Safety Information Management Based on Current Status Analysis," *Korea Society for Information Management*, Vol.34, No.2, pp.137-158, 2017.
- [8] T. Y. Kim, J. E. Lee, and H. J. Oh, "A Study on the Analysis of Disaster Safety Lexicon Patterns in Social Media," *Journal of the Korea Contents Association*, Vol.17, No.10, pp.85-93, 2017.
- [9] Ministry of Public Safety and Security, "Disater Yearbook 2015," Ministry of Public Safety and Security, 2016.
- [10] Ministry of Public Safety and Security, "Statistical Yearbook of Natural Disaster 2015," Ministry of Public Safety and Security, 2016.

- [11] Ministry of Public Safety and Security, "Statistical Yearbook of Natural Disaster 2016," Ministry of Public Safety and Security, 2017.
- [12] GitHub [Internet], <https://github.com>.
- [13] RHINO morphological analyzer [Interne], <https://sourceforge.net/projects/koreanalyzer/>.
- [14] R [Internet], <https://www.r-project.org/>.



이 정 은

<https://orcid.org/0000-0003-2631-0245>

e-mail : lepina@naver.com

2017년 전북대학교 기록관리학과(석사)

2017년~현 재 전북대학교 기록관리학과
박사과정

관심분야: 이용자 분석, 빅데이터 분석,
네트워크 분석



김 태 영

<https://orcid.org/0000-0003-2624-7629>

e-mail : fnty127@hanmail.net

2015년 전북대학교 기록관리학과(석사)

2015년~현 재 전북대학교 기록관리학과
박사수료

관심분야: 이용자 행태, 빅데이터 분석,
온톨로지, 네트워크 분석,
시소러스



오 효 정

<https://orcid.org/0000-0001-8067-2832>

e-mail : ohj@jbnu.ac.kr

2008년 한국과학기술원 컴퓨터공학과
(공학박사)

2000년~2015년 한국전자통신연구원
지식마케팅연구실 책임연구원

2015년~현 재 전북대학교 기록관리학과 조교수
관심분야: 정보검색, 질의응답, 빅데이터정보처리