

재난 전조 정보 추출 알고리즘 연구

A Study of the extraction algorithm of the disaster sign data from web

이창열[†] · 김태환* · 차상열**

Lee, Changyeol · Kim, Taehwan · Cha, Sangyeul

Abstract Life Environment is rapidly changing and large scale disasters are increasing from the global warming. Although the disaster repair resources are deployed to the disaster fields, the prevention of the disasters is the most effective countermeasures. the disaster sign data is based on the rule of Heinrich. Automatic extraction of the disaster sign data from the web is the focused issues in this paper. We defined the automatic extraction processes and applied information, such as accident nouns, disaster filtering nouns, disaster sign nouns and rules. Using the processes, we implemented the disaster sign data management system. In the future, the applied information must be continuously updated, because the information is only the extracted and analytic result from the some disaster data.

Keywords Disaster, Disaster Sign Data, Accident Nouns, Disaster Filtering Nouns, Disaster Sign Nouns, Disaster Extraction

요 지 지구 온난화로 생활 환경이 급격히 변화하고 있으며, 대형 재난이 증가하고 있다. 이러한 재난 발생 시 복구에 많은 자원을 투입하고 있지만, 재난의 예방 만큼 효과적인 대책은 없을 것이다. 재난전조 정보란 하인리히 법칙에 따라 예고되는 재난에 대한 전조이며, 이에 대한 정보를 자동으로 추출하여 대비할 수 있게 하는 것이 본 논문의 초점이다. 웹에 산재된 정보로부터 전조 정보를 정확히 추출하기 위한 기반이 되는 단어(명사)를 구축하고 이를 기반으로 정확한 데이터를 추출할 수 있는 알고리즘을 연구하였다. 본 연구의 결과물로 도출된 단어는 분석적인 연구결과이기 때문에 장기적으로 실제 데이터를 적용하면서 지속적으로 보완되어야 할 것이다.

핵심어 재난, 재난전조정보, 사건 명사, 필터링 명사, 전조 명사, 재난 추출

1. 서론

최근 들어 세계적으로 대형 재난이 많이 발생하고 있으며, 지구 온난화로 기후가 변화하고 이로 인하여 생활 환경이 급격히 변화하면서 이에 적응이 힘들어지고 있다. 자연재해는 Table 1과 같이 발생건수는 4.6배, 피해액은 15.7배 정도로 늘어나고 있으며(정예모, 2005), 집중화된 도시 시설과 문명의 이기들이 인간의 관리 부실 또는 자연 재난에 의하여 이차적으로 우리에게 커다란 재난을 발생시키고 있다.

† 교신저자 : 정회원, 동의대학교 컴퓨터공학과 교수
E-mail : lcy@deu.ac.kr
TEL : 051-890-1726 FAX : 051-890-2629
* 정회원 용인대학교 경호학과 교수
** 정회원 메타라이츠(주) 연구개발부 부장

Table 1. 자연재해 발생 추이 (1950-2004년)

항목	50년대	60년대	70년대	80년대	90년대	최근
발생건수	20	27	47	64	91	63
피해액	449	805	1,476	2,280	7,036	5,668

벨기에의 르벤.카톨릭 연구소(CRED)의 자연 재해에 관한 통계 데이터(<http://www.em-dat.net/>)를 이용한 분석에 의하면, 전 세계에서 매년 약 1억6천만명이 재해로 인한 피해 보고 있다고 나타나고 있다(기술과 가치, 2006)

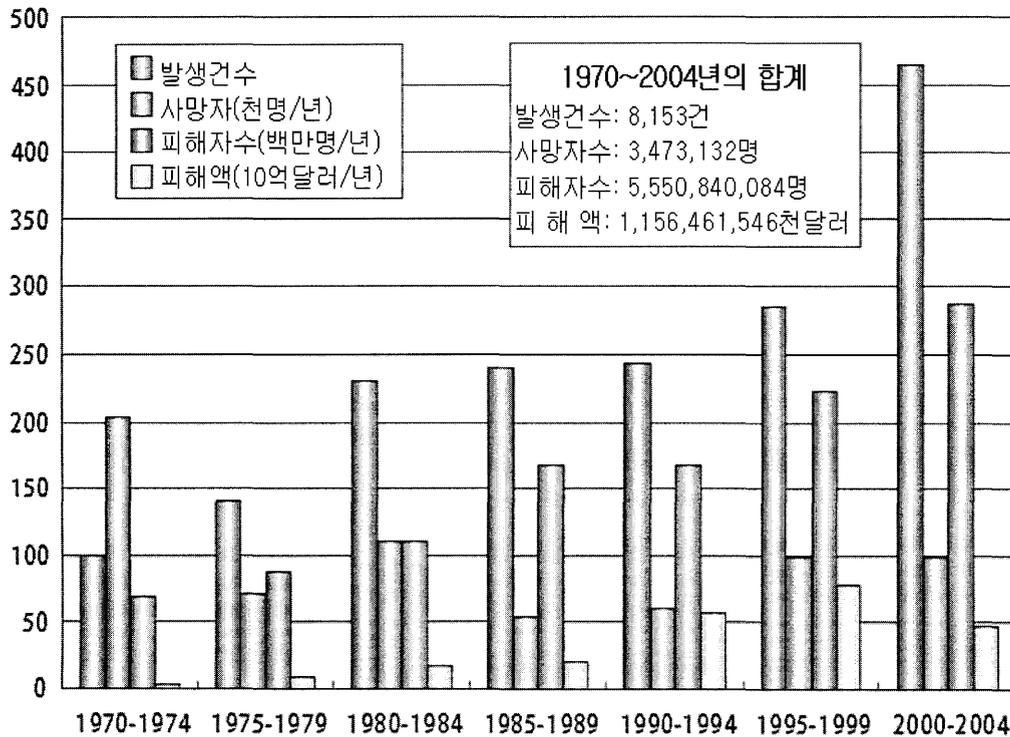


Fig. 1. 세계 자연재해 발생 및 피해 현황(연 평균)

재난전조정보란 이러한 재난이 발생하기 전에 예고가 발생한다는 하인리히 법칙(허버트 윌리엄 하인리히가 1931년 산업안전예방, 과학적인 접근이라는 책에서 1:29:300이라는 법칙을 제시함; 1건의 사고에 대하여 사전에 29건의 경미한 사고가, 그 이전에 300건의 경미한 사건이 발생한다는 사실)(김민주, 2008)에 기반을 두고 있다.(Fig. 1참조)

2010년부터 소방방재청에서는 이러한 재난전조정보를 수집하여 분석하여 발생할 사고에 대한 예방하는 작업을 진행(소방방재청, 2010)하고 있으며, 본 연구는 이러한 재난전조정보를 자동으로 추출하는 기술에 관한 것이다. 본 연구와 별도로 온톨로지에 기반한 재난사례를 분석할 수 있는 연구가 진행되었다(이영재, 2010; 윤남이 등, 2011).

2. 재난전조정보 추출

2.1. 전체 프로세스

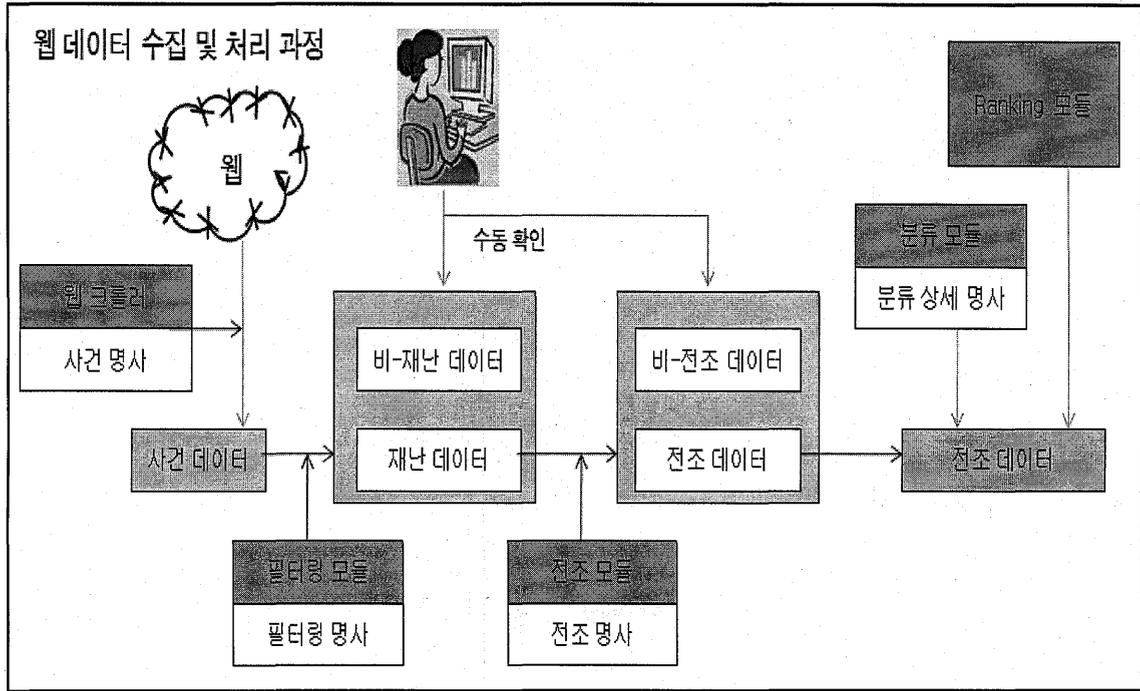


Fig. 2. 재난전조정보 수집 프로세스

재난전조정보의 수집은 관련된 사람에 의한 신고로도 수집될 수 있고, 자동으로 수집할 수도 있다. 본 연구에서는 자동으로 수집하는 것에 대하여만 살펴보기로 한다. 자동 수집은 대상이 되는 웹에서 자동 수집기(웹 크롤러)에 의해 자동으로 수집된 것을 좀 더 심층적으로 분석하는 과정을 포함한다. 전체 프로세스는 Fig. 2에 기술되었다. 각각 프로세스 사항을 살펴보기로 한다.

2.2. 사건 명사 도출

사건 명사란 재난 정보와 관련된 대표적인 명사 단어로 웹에서 재난 전조 데이터를 수집하기 위하여 웹 크롤러에게 제공하는 검색용 키워드라 정의할 수 있다. 1차 사건 명사는 2010년 소방방재청에서 수동으로 수집->분석->시행한 정보를 기반으로 다음과 같이 13개 키워드를 도출하였다.

- 1차 사건명사 : 사고, 재난, 폭발, 붕괴, 사건, 재해, 사망, 부상, 참사, 구조, 구출, 매몰, 전소
- 빈도분석 : 초안에 대하여 소방방재청이 보유하고 있는 재난전조 DB 269건에 대한 빈도 분석을 시도하였다. Table 2의 ‘열’은 초기 사건 명사가, ‘행’은 재난전조 DB 269건에 대한 것이고, 각 명사에 대한 빈도 분석을 시도하였다.

Table 2. 재난전조DB로부터 사건 명사 출현 빈도 계산

코퍼스	사건 명사													각각 코퍼스에서 나타나는 사건명사 수
	사고	재난	폭발	붕괴	사건	재해	사망	부상	참사	구조	구출	매몰	전소	
Total	147	6	7	46	3	8	19	11	5	12	0	5	3	
Num1	0	x	0	x	x	x	x	x	x	x	x	x	x	2
Num2	0	x	x	x	x	x	x	x	x	x	x	x	x	1
Num3	x	x	x	0	x	x	x	x	x	x	x	x	x	1
Num4	0	x	x	x	x	x	x	x	x	x	x	x	x	1
Num5	0	x	x	x	x	x	x	x	x	x	x	x	x	1
Num6	0	x	x	x	x	x	x	x	x	0	x	x	x	2
Num7	0	x	x	x	x	x	x	x	x	x	x	x	x	1
Num8	0	x	x	x	x	x	x	x	x	x	x	x	x	1
Num9	x	x	x	x	x	x	x	x	x	x	x	x	x	0
Num10	0	x	x	x	x	x	x	x	x	x	x	x	x	1
Num11	0	x	x	x	x	x	x	x	x	x	x	x	x	1
Num12	0	x	x	x	x	x	x	x	0	x	x	x	x	2
Num13	x	x	x	x	x	x	x	x	x	x	x	x	x	0
Num14	x	x	x	x	x	x	x	x	x	x	x	x	x	0
Num15	0	x	x	x	x	x	x	x	x	x	x	x	x	1
Num16	0	x	x	x	x	x	x	x	x	x	x	x	x	1
Num17	0	x	x	x	x	x	x	x	x	x	x	x	x	1

빈도분석 결과 초기 사건 명사에서 ‘사망’, ‘구조’, ‘부상’, ‘사건’은 저 빈도로 나타났으며, 예상치 못했던 ‘안전’, ‘위험’, ‘관리’, ‘대책’, ‘피해’, ‘우려’, 등이 고 빈도로 나타났다. 이에 따라 고 빈도 단어 중에서 전문가 회의를 거쳐서 다음과 같은 2차 사건 명사 리스트를 도출하였다.

- 2차 사건 명사
 - * 공통어 : 사고, 붕괴
 - * 대표어 : 안전, 위험, 재해, 재난, 피해
 - * 원인어 : 폭발, 균열, 화재, 추락, 부실

2차 사건 명사에 대하여 최종적인 자문을 거쳐서 비록 저 빈도이지만 중요성이 인정되는 ‘사망’, ‘구조’를 추가하기로 해서 사건 명사 최종 초안을 마련하였다.

- 사건명사 최종 초안
 - * 공통어 : 사고, 붕괴
 - * 대표어 : 안전, 위험, 재해, 재난, 피해, 사망, 구조
 - * 원인어 : 폭발, 균열, 화재, 추락, 부실

2.3. 필터링 명사

필터링 명사란 재난 정보와 관련없는 자료를 제거하기 위하여 선정된 명사로 Fig. 2에서 ‘사건 데이터’로부터 ‘재난 데이터’와 ‘비-재난 데이터’를 분리하는데 사용한다.

Table 3. 사건 데이터를 수동으로 분류

코퍼스	사건 명사												사건 명사 출현	도메인 관련	도메인
	사고	붕괴	안전	위험	재해	재난	피해	폭발	균열	파재	추락	부실			
	196	43	261	149	30	9	181	102	2	85	74	71			
1	1	0	0	0	0	0	1	0	0	1	1	0	4	o	
2	1	1	1	0	0	0	0	0	0	0	0	0	3	o	
3	0	0	0	0	0	0	0	0	0	0	0	0	0	x	경제
4	0	1	0	0	0	0	1	0	0	0	0	0	2	x	사회
5	1	0	0	0	0	0	0	0	0	1	0	0	2	o	
6	0	0	0	0	0	0	0	0	0	0	1	0	1	x	사회
7	0	0	0	1	0	0	0	0	0	0	0	0	1	x	정치
8	1	0	1	0	0	0	0	0	0	0	0	0	2	x	의학
9	0	0	0	0	0	0	0	0	0	0	0	1	1	x	경제
10	0	0	0	0	0	0	0	0	0	0	0	0	0	x	의학
11	0	0	1	0	0	0	0	0	0	0	0	0	1	x	기타
12	1	0	0	0	0	0	0	0	0	0	0	0	1	x	경제
13	0	0	0	0	0	0	1	0	0	1	0	0	2	x	스포츠
14	0	0	0	0	0	0	0	0	0	0	0	0	0	x	정치
15	0	0	0	0	0	0	1	0	0	0	0	0	1	x	사회
16	0	0	0	0	0	0	1	0	0	0	0	0	1	x	사회
17	0	0	0	0	0	0	0	0	0	0	0	1	1	x	사회
18	0	0	1	0	0	0	0	0	0	0	0	0	1	x	사회
19	1	0	1	1	0	0	0	0	0	0	0	1	4	o	
20	0	0	1	0	0	0	0	0	0	0	0	0	1	x	사회
21	1	0	1	1	0	0	1	0	0	0	0	0	4	o	

Table 3은 사건 명사를 적용하여 추출한 사건 데이터 1033개를 수동으로 재난 데이터인지 비-재난 데이터(도메인 정의)인지 분류한 정보(남지순, 2010)이다.

비-재난 데이터로 분류된 자료에 포함되는 비-재난 고빈도 단어를 Table 4처럼 분석하여 보았다. 6회 이상 출현한 단어가 4,097개로 나타났으며, 10회 이상은 1,112개로 이들 단어를 집중 분석한 결과 아래와 같이 8개 분야에 집중되고 있다 :

- 정치(북한, 정권, 대통령, ...) : 38개
- 사회(구제역, 사건범죄, 의식주, ...) : 500개
- 스포츠 : 30개
- 예능(예능프로, 연예인, ..) : 45개
- 의학(병, 약품, ...) : 76개
- 경제(코스피, 환율, 유가, 물가, 증권, 부동산, ..) : 212개
- 과학(순수과학, IT, ...) : 17개
- 기타(중복, 애매한 것) : 194개

Table 4. 비-재난 데이터에 포함되는 고빈도 단어 분석

순서	단어	빈도	분야	선택
1	기자	993		X
2	뉴스	666		X
3	경제	561	경제	○
4	투자	559	경제	○

고 빈도 단어 중에서 비-재난 분야를 대표하는 8개 분야에 확실하게 포함되는 단어를 수동으로 확인하고, 추가적으로 현재 데이터에는 없지만 포함될 가능성이 많은 다음 단어를 추가하여 1차 필터링 명사를 완성하였다.

- 임의 추가 단어 : 북한, 미국, 유럽, 커피, 도박, 서버
- 1차 필터링 명사 개수 : 1,163개

2.4. 전조 명사

재난을 예측할 수 있는 전조 역할을 하는 명사로 이를 기반으로 재난 데이터에서 전조 데이터를 추출한다. 전조 명사인 경우, 단어 1개로 해당 데이터가 전조라고 간주하기는 언어학적으로도 어려울 수 있으며, 또한 전조 자체가 1개의 명사로만 나타날 수 있다고 간주하기는 어려울 것이다. 예를 들어 다음과 같은 구조를 가질 수 있다.

- 구조에 의한 전조 : “붕괴될 우려”
- 동사로 실현 : “무너지다”
- 시제에 의한 사항 : “물이 넘치고 있어요” 등

전조 명사는 2010년 소방방재청이 구축한 재난전조정보 DB를 기반으로 수동으로 구축하였다. 1차적으로는 Table 5에서 처럼 소방방재청의 재난 정보 23개 분류(홍광희, 2010)에 해당되는 단어와, 전조 공통 어휘를 기반으로 전조 명사 26개를 도출하였다.

Table 5. 분류에 따른 어휘와 전조 공통 어휘

전조 공통 어휘		도메인 특수 어휘								
		도메인	어휘	랭킹	도메인	어휘	랭킹	도메인	어휘	랭킹
확률	통제	01. 도로교통	충돌	1	02. 화재	폭발	1	03. 산불	방화	1
인식	단속		과속	2		방화	2		불장난	2
한계	위험상황		음주운전	3		불장난	3		담뱃불	3
위협	관리실패		졸음운전	4		누전	4		건조	4
위기	사정조취		신호위반	5		연기	5		연기	5
대응	실태	04. 열차	충돌	1	05. 폭발	테러	1	06. 해양	기름유출	1
규제	경고신호		탈선	2		충격	2		전복	2
여파	의혹		이탈	3		붕괴	3		기상악화	3
경고	지적		단전	4		장비이상	4		파손	4
금지	정밀조사		기기결함	5		냄새	5		침수	5
대책	통보	07. 가스	폭발	1	08. 유도선	위법	1	09. 환경오염	기름유출	1
파손	재해노출		화재	2		안전관리	2		지구 온난화	2
가능성	문제		누출	3		기기결함	3		오염확산	3
안전관리	이슈		경보발생	4		무허가	4		재해노출	4
제기	대처		잔류가스	5		과적	5		기준치 초과	5

선정된 전조 명사 26개는 다음과 같다 :

- 전조 명사 : 위험, 전조, 가능성, 징후, 조짐, 위협, 불안, 위태, 우려, 불감증, 걱정, 조치, 안전, 방지, 염려, 방치, 대책, 처리, 방안, 무단, 소홀, 보수, 무방비, 불편, 불만, 부실

선정된 전조명사를 재난 데이터에 적용하여 출현 빈도를 Table 6에서처럼 확인하였다.

Table 6. 재난데이터에 전조 명사 출현 빈도 적용

코퍼스 번호	전조 명사																							출현 횟수				
	위험	전조	가능	정후	조집	위협	불안	위태	우려	불감	걱정	조치	안전	방지	열려	방치	대책	처리	방안	무단	소홀	보수	무방		불편	불만	부실	
	###	30	14	1	4	27	39	10	91	17	11	72	###	38	2	69	44	7	7	5	11	38	7		41	23	##	
1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	3
2	1	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	5
3	1	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	0	1	1	0	7	
4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
5	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	3	
6	1	0	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	5	
7	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	4	
8	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	2	
9	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
10	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	1	0	0	0	0	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	5	
13	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
16	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	
17	1	0	0	0	0	0	0	0	1	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	6	
18	1	0	0	0	0	1	0	0	0	0	0	1	1	1	0	0	0	0	0	1	0	0	0	1	0	0	7	
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
20	0	1	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	5	
21	1	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	1	1	0	6	
22	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	
23	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	

Table 6의 결과를 Table 7처럼 요약하였다.

Table 7. 전조명사 적용 정보 요약

총 데이터	재난 데이터		전조 데이터
	전조명사 출현	전조명사 없음	전조명사 출현
278건	50	11	213

2.5. 재난 전조 데이터 분류

재난 데이터에 대한 그룹핑을 통한 정보의 관리를 쉽게 하기 위하여 인적 재난에 적용한 분류 체계(홍광희, 2010)를 기본으로 전조 데이터를 분류한다. 23개 재난 정보 분류는 다음과 같이 3가지 형태로 재 구성된다. 즉

- (1) 대상과 사건 : 산불, 환경오염, 전기감전, 수난
- (2) 대상 : 도로교통, 철도, 해양, 가스, 유·도선, 공단시설, 광산, 승강기, 보일러, 항공기, 등산, 농기계, 자전거, 레저, 놀이시설
- (3) 사건 : 화재, 폭발, 붕괴, 추락

이를 기반으로 전조 데이터를 분류하기 위한 규칙을 다음과 같이 정의하였다.

- 규칙 1. 우선순위 부여. 위 3가지 형태 중 (1) > (2) > (3)의 형태로 우선순위를 부여한다. 예를 들어, “광산 폭발”은 “폭발”(3번 형태)로 분류하지 않고 “광산”(2번)으로 분류한다.
- 규칙 2. 영역 명사가 포함되어 있으면 해당 분류로 간주한다. 영역 명사란 특정 분류에서 해당 분류에 절대적으로 포함되는 명사로 예를 들어, “지하철”이란 명사가 나타나면, 화재가 발생하든, 무슨 사고나 나든지 “열차”로 분류할 수 있는 단어를 지칭한다.
- 규칙 3. 공통 명사(여러 분류에 속하는 단어) 들만 존재하면 해당 분류에 공통적으로 나타나는 분류로 정의한다.
- 규칙 4. 분류에 가중치 부여. 사고 시 대규모 인적 사고가 야기되는 사항(분류) 또는 기간 산업에 점수를 높이 부여한다.

Table 8에 가중치를 최대 0.5점 만점으로 부여하였다. 본 점수는 소방방재청 재난전조정보 담당자와 회의를 거쳐서 설정한 것이며, 이 점수 또한 피드백 과정을 통하여 재 조정될 것이다.

Table 8. 재난 정보 분류 위한 분석 데이터

번호	분류	유형	가중치	영역 명사	공통 명사
1	도로교통	2	0.3	도로, 교통, 자동차, 차량, 버스, 전복, 신호등, 과속	충돌(0, 1, 4, 6, 8, 15, 21), 추돌(0, 1, 4, 8, 15, 21), 교량(1, 2, 16, 19), 위반(0, 1, 4, 7, 8, 15), 운전(1, 4, 8, 15, 21)
2	화재	3	0.4		화재(2, 3, 4) 방화(2, 3, 4) 전소(2, 3, 4), 누전(2, 4, 12), 불(2, 3), 연기(2, 3, 4)
3	산불	1	0.3	산불	담뱃불(2, 3, 4, 5), 산(0, 3, 18, 19), 등산객(0, 3, 18, 19)
4	열차	2	0.4	열차, 기차, 탈선, 철로, 선로, 기차역, 지하철	폭발(5, 2, 4, 7, 14, 15)
5	폭발	3	0.5		폭탄(0, 1, 4, 5, 7, 15), 과열(2, 1, 5, 14, 15)
6	해양	2	0.3	어선, 선박, 해양, 표류	기상악화(0, 1, 6, 8, 15, 19) 기름유출(9, 1, 4, 5, 6, 10, 15), 구조(0, 1, 2, 4, 8, 11, 13, 15)
7	가스	2	0.5		가스(7, 2, 5, 10, 11, 14), 누출(0, 6, 7, 9, 15)
8	유도선	2	0.3	유도선, 과승, 과적, 선박	전복(8, 6)
9	환경오염	1	0.3	환경, 오염, 공해, 소음	쓰레기장(0, 2, 9)
10	공단시설	2	0.4	공단	공장(10, 2, 5, 9, 14)
11	광산	2	0.4	광산, 광부	매물(0, 2, 5, 11) 붕괴(16, 2, 5, 11, 23)
12	전기감전	1	0.3	감전	합선(2, 4, 10, 12, 13, 23), 전선(0, 2, 4, 12, 13, 14, 15, 23)
13	승강기	2	0.3	승강기, 엘리베이터	추락(19, 1, 4, 13, 15, 16, 18, 23)
14	보일러	2	0.2	보일러	도시가스(0, 5, 7, 14), 질식(0, 2, 3, 7, 14)
15	항공기	2	0.5	항공기, 비행기, 날개, 불시착, 격추, 헬리콥터	정비(0, 1, 4, 7, 8, 13, 14, 15, 20, 21, 23)
16	붕괴	3	0.5		부실(0, 1, 10, 16, 23), 지반(0, 1, 6, 11, 16)
17	수난	1	0.2	수난, 강, 하천, 물가, 수영, 물살, 물놀이	바다(0, 6, 8, 17, 22)
18	등산	2	0.1	등산, 고립, 산사태, 조난,	산(0, 3, 18, 19), 암벽(0, 18, 19)
19	추락	3	0.3	낙마	미끄러짐(19, 18), 낙하(19, 18)
20	농기계	2	0.1	농기계, 경운기, 농경지, 경작지, 논	
21	자전거	2	0.1	자전거, 바이크	브레이크(1, 4, 13, 15), 고장(0, 1, 4, 6, 8, 13, 14, 15, 21, 23)
22	레저	2	0.2	레저, 낙마	
23	놀이시설	2	0.3	놀이시설	

2.6. 순위 부여

- 규칙 5. 전조 명사를 이용한 문서 점수 부여. 예를 들어, Table 9와 같이 처리한다.

(1) 정규 점수 : (문서 x 합계 점수) / (총 합계). 예를 들어 “문서 1”인 경우 11/23 = 0.48

(2) 조정 점수 : 분류 점수의 평균이 0.3이므로 본 점수도 평균적으로 0.3이 나오게 하는 점수. 조정 점수 = 문서 개수(N) * 0.3 * 정규 점수

* Table 9인 경우, 문서 1의 조정 점수 = 3 * 0.3 * 0.48 = 0.43

- 규칙 6. 최종 점수 = 분류 점수 + 조정 점수. 순위는 최종점수가 가장 많은 것부터 부여.

Table 9. 순위 부여 과정

대상 문서	분류	전조 명사						합계	정규점수	조정 점수	총 합계	순위
		위험	전조	가능성	징후	...	부실					
문서 1	2/0.3	4	1	1	2		3	11	0.48	0.43	0.73	1
문서 2	3/0.5	1	0	1	1		2	5	0.22	0.20	0.70	2
문서 3	1/0.3	2	0	0	1		4	7	0.30	0.27	0.57	3
합계		7	1	2	4		9	23	1	0.9	2.00	

3. 재난전조정보 관리시스템 운영

3.1. 시스템

재난전조정보 관리시스템은 Fig. 3에서 정의한 프로세스를 기반으로 구축된 시스템이다. 2장 연구에서 진행한 재난 전조 정보를 추출하기 위하여 정의한 사전과 규칙에 기반하였고, 특히 사전 내용(단어)은 변경 가능하게 구성하였다.

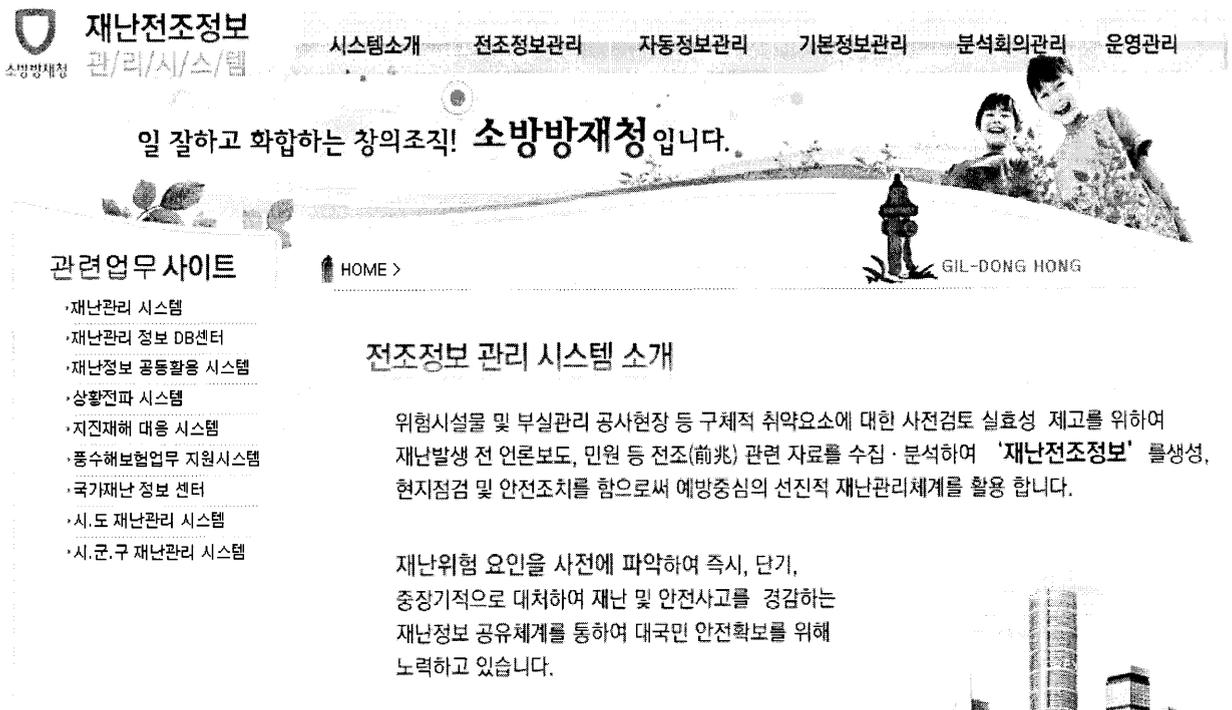


Fig. 3. 재난전조정보관리시스템 첫 화면

3.2. 사전 정보 보완

본 연구로 만들어진, 사건명사, (비-재난) 필터링 명사, 전조명사 등은 웹 데이터 등을 수동으로 분석하고, 전문가의 검토를 거쳐서 만들어진 초기 데이터라 할 수 있다. 이들 데이터는 실제 웹에서 재난전조정보를 도출하는 과정에서 효율성이 떨어지는 단어(명사)를 찾을 수 있고, 또 그 동안 고려되지 않은 새로운 적절한 단어를 도출할 수도 있다. 이러한 피드백 과정을 통하여 재난전조정보 추출의 질적인 보완이 이루어질 것이다. 사전 정보 보완 과정을 Fig. 4에 예시하였다.

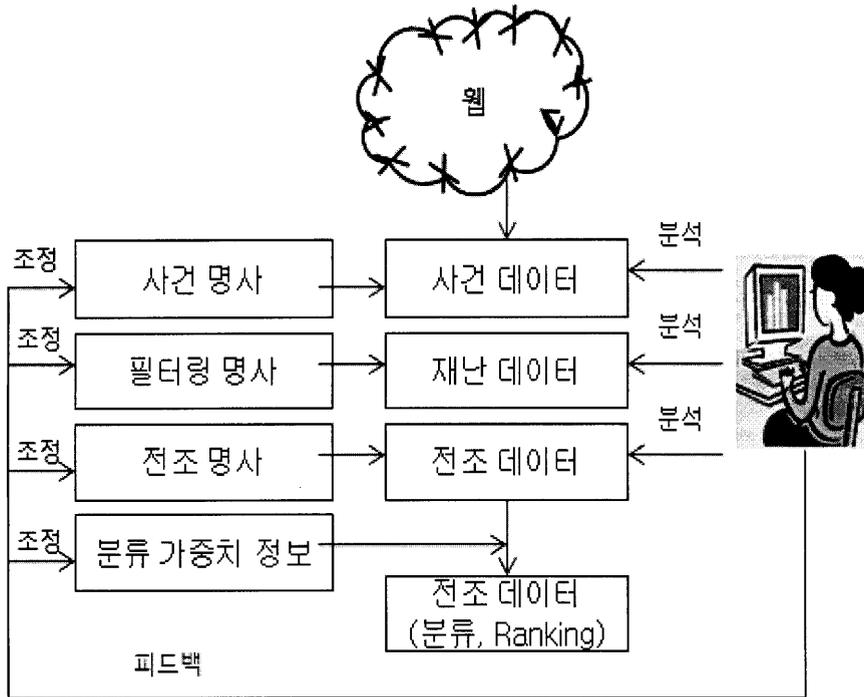


Fig. 4. 피드백을 통한 사전 정보 보완

4. 결론

본 논문은 웹 데이터로부터 재난전조정보를 추출하여 발생할 재난에 대하여 미리 대비 조치를 취하기 위한 것이다. 재난전조정보를 추출하는 방법은 재난전조정보를 예측할 수 있는 적절한 단어(명사)를 정의하고, 해당 단어가 포함된 데이터를 추출하는 것이다.

물론 단어(명사) 한 개로부터 재난전조정보를 판단할 수 있는 것은 아니다, 어떤 데이터는 동사를 분석해야 하고, 어떤 데이터는 여러 단어의 복합적 상황을 판단해야하지만, 본 연구에서는 한 단어(특히 명사)에 기반한 재난전조정보를 추출하는 연구를 진행하였다. 그러므로 얼마나 정확히 관련 단어를 모으는 것이 본 연구의 이슈사항이었다. 이러한 단어를 모으는 과정은 많은 웹 데이터에 대한 수동 분석과, 전문적 판단에 기반하였다. 그러므로 이런 수집된 단어에 대한 지속적보완이 필요하며, 이러한 보완은 재난전조정보관리시스템을 운영하면서 이루어질 수 있다. 본 시스템은 잘못된 결과(예를 들어 재난 데이터 속에 실제적으로 재난이 아닌 데이터가 있음)를 분석할 수 있는 모듈(Fig. 2에서도 표시하듯이 결과물에 대한 수동 확인 프로세스 있음)을 제공하고 있다.

본 시스템을 운영하면서 보완되는 사전에 대한 추가적 연구가 필요할 것이다.

감사의 글

본 연구는 소방방재청이 출연하고 인적재난안전기술개발사업단이 시행하는 2010년도 안전관리기술개발사업의 지원으로 이루어졌습니다.

참고문헌

- [1] 기술과 가치 (2006). 재해 사전 예방 강화 및 투자확대 방안. 소방방재청, 11-1660000-000075-01.
- [2] 김민주 (2008). 하인리히법칙. 토네이도..
- [3] 남지순 (2010), 재난전조정보 분석을 위한 사전 정보 구축. 메타라이츠(주).
- [4] 소방방재청 (2010). 제20차 재난전조정보 분석회의 결과보고..
- [5] 윤남이, 최용규 (2011). “재난전조정보 수집, 분석 및 의사결정체계 연구.” 2010 한국재난관리표준학회 학술 발표회.
- [6] 이영재 (2010). 재난전조정보 관리체계 구축 등을 통한 인적재난 피해저감 방안 연구보고서. 소방방재청.
- [7] 정예모 (2005). 자연재해의 경제적 손실과 대응. 중앙일보..
- [8] 홍광희 (2010). 재난취약분야 발굴 및 예방안전체계구축. 소방방재청.

- ▶ 논문접수일 : 2011년 06월 08일
- ▶ 심사의뢰일 : 2011년 06월 09일
- ▶ 심사완료일 : 2011년 06월 23일