

텍스트 마이닝과 소셜 네트워크 분석을 이용한 재난대응 용어분석

Analyzing Disaster Response Terminologies by Text Mining and Social Network Analysis

강성경 (Seong Kyung Kang) 동국대학교 경영정보학과 석박사 통합과정
유환 (Hwan Yu) 동국대학교 경영정보학과 박사과정
이영재 (Young Jai Lee) 동국대학교 경영정보학과 교수, 교신저자

요약

세월호 침몰사고, 판교 환풍구 붕괴사고 등 재난은 점차 복잡적이고 대형화되고 있다. 따라서 이러한 재난에 신속히 대응하기 위한 기관들의 협업 또한 중요해지고 있다. 다수기관 간 협업과정에서는 다양한 용어를 바탕으로 의사소통이 이루어진다. 의사소통은 ‘용어’를 기반으로 하므로 ‘용어’에 대한 중요성 또한 간과할 수 없다. 따라서 본 연구에서는 재난현장에서 사용하는 용어를 선정하여 텍스트 마이닝 및 소셜 네트워크 분석(SNA: Social Network Analysis)을 이용해 어떤 용어가 대응과정에 있어 핵심적인 용어인지를 파악해보았다. 텍스트 마이닝의 TDM을 이용하여 역문헌 빈도수를 산출해 용어와 문서 간의 관계를 알아보고, SNA를 통해 노드(용어)와 노드 사이의 관계를 파악하였다. 용어분석의 결과 표현은 용어 간의 유기적인 관계를 시각화할 수 있는 마인드맵(Mind Map)을 이용하였다. 용어는 미국의 NIMS, EMR, 그리고 우리나라의 재난 및 안전관리 기본법을 토대로 온톨로지 개념에 따라 계층적(Class, Object, Instance)으로 분류하였으며, 신문기사와 사설, 정책보고서 등의 정부 간행물에서 선정하였다. 이러한 재난대응 핵심용어의 파악은 재난현장에서 사용하는 용어를 표준화하기 위한 기초자료로 활용할 수 있으며, 온톨로지 개념에 따라 용어들을 계층적으로 분류하였기 때문에 재난 대응에 대한 다양한 자료들을 추적하고 검색하는데 용어의 분류체계를 활용할 수 있다. 이 밖에 사고대응 시나리오 작성 시에도 핵심용어를 활용할 수 있을 것으로 판단된다.

키워드 : 재난대응 용어, 텍스트 마이닝, 소셜 네트워크 분석, 근접 중심성, 온톨로지

I. 서론

해마다 크고 작은 재난 및 안전사고가 발생하고 있으며 이러한 재난 및 사고를 대응·수습하는 과정에서 여러 문제점이 제기되고 있다.

재난 현장에서는 항상 불확실성이 존재하기 때문에 재난 대응·수습 중에도 어떤 2차 사고나 피해가 발생할지 예측하기 힘들다. 따라서 재난의 피해가 확산되지 않도록 짧은 시간 내에 구조 및 수습을 해야 한다. 그러기 위해서는 재난 현장에

서 활동하는 재난 실무자의 역량과 대응·수습활동에 참여하는 여러 기관의 협업이 중요하다. 다수기관 간 원활한 협업에서 의사소통은 굉장히 중요한 역할을 한다. 의사소통은 ‘용어’를 기반으로 이루어지기 때문에 ‘용어’에 대한 중요성 또한 간과할 수 없다.

따라서 본 연구에서 재난대응 용어의 분석을 통해 재난대응에 있어 어떤 용어가 핵심적이고 쟁점이 되는지를 파악해본다. 이를 위해 텍스트 마이닝(Text Mining)과 소셜 네트워크 분석(SNA: Social Network Analysis)을 이용한다. 최근 사회적 관계를 바탕으로 형성된 소셜 네트워크 관련 기술이 빠르게 발전하고 있으며, 이에 따라 SNA 또한 급속도로 발전하고 있다. SNA는 텍스트 마이닝 기법과 함께 많이 사용되고 있으며, 확산된 내용과 함께 연결의 맥락을 파악하고 분석하는 기법으로 알려져 있다. 특히 인과성에 대한 선형적 시각에서 벗어나 비선형적이고 창의적 결과를 표현하는 방법론으로 주목받고 있다.

용어분석의 범위는 재난관리의 4단계(예방, 대비, 대응, 복구) 중 ‘재난대응’으로 정했으며, NIMS(미국 FEMA), EMR(ISO22320), 그리고 우리나라 재난 및 안전관리 기본법을 토대로 온톨로지 개념에 따라 용어를 계층적(Class, Object, Instance)으로 분류하였다. 분석결과 재난대응(Class)의 5개 영역(Object) ‘지휘/통제, 상호협업, 정보관리, 공보/커뮤니케이션, 자원관리’ 중 ‘지휘/통제’ 영역이 가장 핵심적인 것으로 나타났으며, ‘지휘/통제’ 영역 내의 용어(Instance) 중에서는 ‘조치’와 ‘책임’이 다른 용어보다 해당 영역과의 근접성이 높은 것으로 나타났다.

용어분석을 통한 핵심용어의 파악은 재난현장에서 사용하는 대응용어를 표준화하기 위한 기초 자료로 활용할 수 있다. 현재 재난현장에서 사용하는 용어가 기관별로 다르기 때문에 기관 간 커뮤니케이션 제대로 이루어지지 않고 있다. 따라서 이러한 문제를 해결하기 위해 핵심용어 기반의 용어 통일이 필요하다. 이 때 우선적으로 어떤

용어를 표준화할 것인지에 대한 기준을 핵심용어를 기반으로 선정할 수 있다. 용어의 통일은 단순히 기관 간 커뮤니케이션의 효율을 증진시키는 것뿐만 아니라, 재난대응에 대한 다양한 자료들을 축적하고 검색할 시 표준화된 용어를 기반으로 일관성 있는 자료를 얻을 수 있도록 해준다. 이 밖에도 사고대응 시나리오 및 매뉴얼 작성 시 핵심영역, 혹은 핵심활동에 대한 다양한 교육/훈련의 기회를 제공할 수 있을 것으로 기대된다.

II. 문헌 연구

기사, 사설, 논문, 정책보고서 등에서 재난대응에 관련된 문제를 많이 언급하고 있으며, 이러한 문제를 개선하기 위한 정책연구들이 많이 진행되고 있다. 하지만 재난대응과 관련된 용어를 주제로 다루는 논문은 아직 부족한 실정이다. 재난대응 용어분석에 관한 선행연구가 거의 없다는 점에서 본 연구와 유사한 접근방법을 가진 논문 및 타 분야에서 SNA를 이용해 핵심용어를 파악하는 과정을 조사하였다.

이성수(2012)의 연구에서는 과거 절개지 붕괴 사고사례에 텍스트 마이닝 및 사례기반추론 이론을 도입해 규칙모델을 생성하고, 규칙모델의 신뢰도와 사고 사례의 유사도를 정량적으로 측정하여 보여주고 있다. 이러한 과정은 비슷한 재난사고 사례를 분석하여 새로운 재난전조정보에 대한 위험 수준을 결정하여 사전 예방조치를 취하는데 도움이 된다.

이성수의 연구에서는 절개지 붕괴사고 사례에서 언급되는 용어의 빈도를 계산하여 규칙모델을 설계했다. 본 연구에서 핵심용어를 파악하기 위해 20개의 재난사고사례를 사용하고, 문서 속에서 드러나는 용어의 빈도를 계산하는 텍스트 마이닝 기법을 활용한다는 점에서 이성수의 논문과 유사한 분석방법을 일부 사용한다는 것을 알 수 있다. 하지만 이성수의 연구는 재난관리 4단계 범위(예방, 대비, 대응, 복구) 중에서 재난대응

이전에 발생하는 예방단계에 초점을 맞추고 있고, 재난유형도 자연재해로 한정하고 있다. 반면에 본 연구는 재난대응과정의 문제를 담고 있는 사고사례를 대상으로 하며, 선정한 재난유형 또한 사회재난이라는 점에서 차이를 보인다. 또한, 연구목적에서도 이성수의 연구는 새로운 재난전조정보의 위험 수준을 판단하고 유사한 사례를 발견하기 위한 규칙모델을 생성하는 것이지만, 본 연구는 사고사례에서 언급되는 다양한 재난대응 관련 용어 중 핵심용어 혹은 쟁점이 무엇인지를 파악하는 것을 목적으로 한다는 점에서 차이를 보인다.

국립재난안전연구원(2013)의 연구에서는 미래예측에 대한 국내외 방법론과 사례 검토를 통해 재난 분야에 실질적 영향을 줄 수 있는 요인분석과 이슈 도출을 위해 구체적인 미래재난 시나리오의 구축이 가능하도록 하는 미래재난예측 방법론을 정립하였다. 방대한 데이터 분석을 위한 빅데이터 개념과 텍스트 마이닝을 활용한 환경스캐닝 기법의 적용, 전문가 델파이 설문조사를 진행할 수 있는 온라인 시스템을 도입하고 있으며, 미래재난예측기법의 시험적 적용을 위해 세계경제포럼에서 발표한 글로벌 리스크를 대상으로 분석을 시행하였다. 최종적으로는 주요 핵심 키워드를 선정하고 재난에 중요한 영향력을 미칠 수 있는 연관성 높은 토픽들을 분류하여 미래재난 이슈들을 도출하였다.

국립재난안전연구원의 연구에서는 미래재난예측 시나리오를 구성하기 위해 텍스트 마이닝 및 델파이 기법을 이용해 핵심 키워드를 도출·선정하여 주요 Topic 별로 키워드 맵을 구성했지만 본 연구에서는 재난대응과 관련된 용어들을 온톨로지 기법에 따라 3개 계층(Class-Object-Instance)으로 분류하여 TDM(Term Document Matrix)을 구성하고, 역문헌 빈도수(Invert Document Frequency) 및 정규화(Normalization)를 수행하였다. 또한 SNA(Social Network Analysis)를 이용해 용어들의 근접 중심성(Closeness Centrality)을 계산

하고 마인드맵을 통해 3개 계층 간의 관계를 표현하였다.

김유호(2012)의 연구에서는 최근 3년 동안 주요 일간지에 게재된 의료민영화 및 영리병원에 관한 신문 사설을 중심으로 내용분석의 일종인 언어네트워크 분석을 통해 핵심 키워드를 찾아내고, 키워드 간의 연결 중심성 분석을 통해 논란의 핵심이 무엇인지를 파악하고자 했다. SNA 중 연결 중심성(Degree Centrality)을 이용하여 의료민영화 및 영리병원과 관련된 주요 중심키워드를 도출하였다.

연결 중심성의 경우 한 노드에 몇 개의 다른 노드들이 직접 연결되어있는지를 측정하는 방식으로 직접적인 연결 관계를 가지고 있는 노드 간의 관계만을 상정하기 때문에 연결망 내에서 지역적 중심성(local centrality)을 측정하는 개념으로 사용된다. 하지만 단순히 직접적인 연결만으로는 그 노드의 연결망 내 영향력을 파악하기 어려우므로 본 연구에서는 직접적 연결뿐만 아니라 한 노드에 연결된 모든 노드의 거리를 계산하는 방식인 근접 중심성(Closeness Centrality)을 이용한다.

위와 같은 선행연구 외에도 다양한 분야에서 텍스트 마이닝이나 SNA를 활용한 키워드 분석을 하고 있다. 김주호(2015)의 연구에서는 정보통신 분야의 국가 R&D 연구동향에 대한 정량적 분석을 하였다. 국가 R&D 조사, 분석 DB에서 최근 3년간 연구과제의 영문 키워드를 추출하고 정제한 후, 문서출현 빈도를 산출하였다. 또한, 연구 키워드 동시 출현정보를 이용해 세부 기술별 연구동향 네트워크를 구성한 후 네트워크 구조 분석, 3대 중심성 및 구조 등 위성 분석결과를 통해 국가 R&D 연구 특성을 파악하였으며, 김동래(2015)의 연구에서는 최근까지의 경영학 연구를 기반으로 지식구조분석 및 연구동향을 파악하기 위해 논문의 키워드 정보를 추출하고 중심성을 분석하였다.

III. 연구 설계

본 연구의 범위는 재난관리 4단계(예방, 대비, 대응, 복구) 중 재난 ‘대응’에 관련된 용어를 분석하는 것으로 한다. 재난대응의 영역은 미국의 NIMS(National Incident Management System) 및 ISO22320 EMR(Emergency Management Requirements) 그리고 우리나라의 재난 및 안전관리 기본법을 참고해 지휘/통제(Command/Control), 상호협업(Cooperation/Coordination), 정보관리(Information Management), 공보/커뮤니케이션(Public Information/Communication), 자원관리(Resource Management) 5개로 분류한다. 또한, 5개 영역과 관련된 용어 선정하고, 신문기사와 사설, 정부 간행물 등에서 그 빈도수를 산출해 용어 간의 근접 중심성을 파악한다.

연구 과정은 3단계로 구분할 수 있다. 첫 번째로, 텍스트 마이닝¹⁾에서 제시한 TDM(Term Document Matrix)을 구성하는 것이다. 텍스트 마이닝을 위한 정보추출 과정에서는 다양한 수학적 알고리즘이 존재하며, 그중 간단하면서도 많이 활용되는 방법 중 하나가 TDM이다. TDM은 여러 문서에서 동시에 출현하는 단어는 범용적 확률이 높다는 전제로부터 역문헌 빈도수(Invert Document Frequency)를 계산하는 방법이다(소방방재청, 2010). TDM은 용어와 문서(재난사고사례)의 관계에서 어떤 용어가 문서에 많이 등장하는지(용어 빈도수)를 파악할 수 있게 해준다. 하지만

특정 용어가 문서에 많이 등장한다고 해서 반드시 중요성이 높다는 것은 아니므로 이러한 문제를 해결하기 위해 빈도수를 산출한 후 정규화(Normalization)를 실시한다.

두 번째로, SNA(Social Network Analysis) 중 근접 중심성(Closeness Centrality)을 이용해 용어 간의 근접성을 파악한다. SNA는 네트워크 내에서의 관계를 여러 가지 수학적 기법들에 따라 분석할 수 있는데, 그중 연결망의 구조적 특성을 가장 기본적으로 분석할 수 있는 방법으로는 연결망의 규모, 밀도, 중심성 등이 있다(김혜진, 2007). 이 중 대표적으로 중심성을 많이 이용하는데, SNA의 중심성 지표로는 연결 중심성(Degree Centrality), 근접 중심성(Closeness Centrality), 매개 중심성(Betweenness Centrality) 등이 있다.

본 연구에서는 재난대응 핵심용어를 파악하기 위해 근접 중심성을 이용한다. 앞서 문헌연구에서도 언급하였듯이 연결 중심성은 직접적으로 연결 관계를 가지고 있는 노드 간의 관계만을 상정하기에 그 노드의 연결망 내 영향력을 파악하기 어렵다. 따라서 재난대응 핵심용어를 파악하기 위해서 직접적 연결뿐만 아니라 한 노드에 연결된 모든 노드의 거리를 계산하는 방식인 근접 중심성을 이용한다.

근접 중심성은 한 노드가 다른 노드들과 얼마나 평균적으로 가까이 위치하는지를 측정하는 전체 중심성을 나타낸다. 근접 중심성은 다른 점들과의 인접성(closeness) 혹은 거리(distance)로 측정할 수 있다. 여기서 말하는 노드 간의 거리는 두 노드를 연결하는 최단 거리를 의미한다.

본 연구에서는 거리(distance)를 이용하여 근접 중심성을 정의하며, 다음 수식을 사용한다(김기환, 2015).

$$C_i = \frac{\sum_{o \neq i} \frac{1}{d(o, i)}}{n - 1}$$

$d(o, i)$ = 노드의 최단거리, n = 전체 노드 수

1) Turban *et al.*(2010)은 Text Mining Process를 3단계로 제시하고 있다. 첫 번째 단계에서는 Corpus를 수집하며, 두 번째 단계에서는 TDM(Term Document Matrix)을 구성한다. TDM은 용어와 문서를 매트릭스 형식으로 구성한 테이블을 말한다. 행(row)은 문서를, 열(column)은 용어를 나타낸다. 용어와 문서 간의 관계는 빈도수라는 지표로 나타낼 수 있다. Text Mining의 마지막 단계는 TDM으로부터 새로운 패턴(Novel Patterns), 즉 지식을 추출하는 것으로 이루어진다.

이 때, 두 노드 간의 거리 계산은 유클리디안 거리(Euclidean Distance)를 이용한다. 유클리디안 거리란 N차원의 공간에서 두 노드 사이의 거리를 계산할 때 사용한다. 두 노드 P와 Q가 각각 $P = (p_1, p_2, p_3, \dots, p_n)$, $Q = (q_1, q_2, q_3, \dots, q_n)$ 의 좌표를 가질 때 두 노드 사이의 거리를 계산하는 유클리디안 거리 공식은 다음과 같다(Wang *et al.*, 2005).

$$d(p, q) = \sqrt{\sum_{k=1}^n (p_k - q_k)^2}$$

$$= \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2},$$

$k = 1, 2, \dots, n$

$d(p, q)$ = p와 q간의 거리

마지막으로 근접 중심성을 바탕으로 용어 간의 거리를 고려한 마인드맵을 그린다. 마인드맵이란 ‘생각의 지도’라는 뜻으로 생각한 것을 지도로 그리듯이 이미지화해 사고력, 창의력, 기억력을 높이는 두뇌 개발 기법이다. 사람은 시간이 지나면 생각한 내용 일부를 잃어버리게 되는데 마인드맵은 유기적으로 연결되는 생각들을 훌륭하게 상기시켜준다는 장점이 있다. 본 연구에서는 마인드맵을 이용해 재난대응과 관련된 용어들의 근접성을 한눈에 알아볼 수 있도록 시각화 한다.

마인드맵의 중심은 ‘재난대응’으로 하며, ‘재난대응’과 5개 영역 간의 거리를 근접 중심성에 따라 나타낸다. 그 후 5개 영역 내의 용어들도 근접 중심성에 따라 순서대로 나타낸다.

IV. 실증 분석

4.1 Text Mining(TDM 구성 및 정규화)

4.1.1 사고사례 수집 및 용어선정

TDM 작성을 위해서는 먼저 사고사례와 용어

를 수집해야 한다. 본 연구에서는 우리나라의 주요 사고사례로 꼽히는 세월호 침몰사고(C1), 구미 불산 유출사고(C2), 삼풍백화점 붕괴사고(C3), 판교환풍구 붕괴사고(C4), 대구 지하철 화재사고(C5)를 선정하였다. 사고사례별로 재난대응 과정의 내용을 많이 포함하고 있는 사설과 전문가 칼럼 위주의 신문기사, 감사보고서 및 재난백서 등의 정부간행물을 데이터로 준비하였으며, 사례별로 신문기사 3종류와 정부간행물 1종류를 분석 데이터셋(set)으로 설정하였다. 5개 사고 사례의 총 문서 수는 20개이며, 이는 TDM에서 행(row)에 배정하였다.

용어는 20개의 사고 사례 문서에서 선정한다. 본 연구에서는 앞서 언급하였듯이 재난대응의 영역을 5개로 분류하였다. 따라서 5개의 영역에 해당하는 용어를 선정하고, 선정한 용어는 TDM에서 열(column)에 배정하였다.

이때 열(column)은 NIMS(미국 FEMA), EMR(ISO22320) 그리고 우리나라 재난 및 안전관리 기본법을 토대로 온톨로지 개념에 따라 계층적으로 분류하였다. 온톨로지란 ‘어떤 관심 분야를 개념화하기 위해 명시적으로 정형화한 명세서’로 주로 인용되고 있다(Gruber, 1993). 재난대응에 관련된 용어들을 계층적으로 표현하기 위해 여기서는 Class, Object, Instance 3계층으로 분류하였다.

Class는 Object를 포함하는 틀로, 본 연구에서는 재난관리의 4단계 중 ‘재난대응’을 Class로 설정하였다. Object는 Class의 하위 항목으로 ‘재난대응’의 5개 영역(지휘/통제, 상호협업, 정보관리, 공보/커뮤니케이션, 자원관리)으로 설정하였다. Instance는 Object의 하위 항목으로 재난대응의 5개 영역에서 사용하는 용어들로 지정하였다(Instance는 재난담당 공무원 및 정책연구 등에서 주요하게 언급되는 용어를 기반으로 선정하였다).

본 연구에서 설정한 Class, Object, Instance를 정리하면 다음 <표 1>과 같다.

〈표 1〉 계층별 용어 분류

분류	용어				
Class	재난대응				
Object	지휘/통제	상호협업	정보관리	공보/커뮤니케이션	자원관리
Instance	책임	재난현장	원인	전파	물자
	권한	현장지휘소	사상자	홍보	지원
	보고	초기대응	신고(접수)	오보	전문가
	지시	구호	정보시스템	미디어	구조인력
	지휘	수색	매뉴얼	국민	시설
	통제	구조	대응계획	불만	자원봉사자
	예/경보	구급	평가	통신	
	훈선	긴급조치		보도	
	대책회의	협업		인터뷰	
	대책본부	유관기관			
	조치	의료			
	예측	복구			
	역할	교통			

4.1.2 TDM 구성 및 정규화

사고사례(Document)와 용어(Term)를 선정 한 후에는 TDM을 구성한다. TDM은 용어-사고사례 간의 관계를 빈도수를 토대로 파악할 수 있도록 해준다. 수집한 사고사례 문서에서 재난대응 5개 영역에 해당하는 용어들이 언급된 빈도수를 계산한다.

이때, 용어 빈도수(term frequency, tf)는 사고사례 문서에서 해당 용어가 몇 번이나 등장하는지를 나타낸다. 예로 구미불산 누출사고(C2) 4번 문서(D4) 내용 중 ‘지휘’라는 용어는 10번 언급되기 때문에 용어 빈도수는 ‘10’이 된다. 문서 빈도수(document frequency, df)는 20개의 사고사례 문서 중 특정 용어가 몇 개의 문서에 등장하는지를 나타낸다. 예로 ‘지휘’라는 용어는 총 20개의 문서 중 8개의 문서에서 언급되기 때문에 문서 빈도수는 ‘8’이 된다.

하지만 어떤 특정 용어가 한 문서에 자주 등장한다고 해도 그 용어가 다른 용어보다 더 중요한 것이라고 단정 지을 수 없으므로 빈도수(frequency)를 계산한 후 다음 식에 따라 정규화(normalization)를 한다(Turban, 2012).

$$idf(i, j) = \begin{cases} 0 & \text{if } wf_{ij} = 0 \\ (1 + \log(wf_{ij})) \log \frac{N}{df_i} & \text{if } wf_{ij} \geq 1 \end{cases}$$

범례

N : 총 사고사례 문서 수

df_i : 문서빈도수 [특정 기본요소(i)를 포함하는 문서의 수]

i : i번째 기본요소(Term, Instance)

j : j번째 사고사례(document)

wf_i : 용어빈도수 [해당 사고사례에서 특정 기본요소(i)가 등장하는 수]

예로 20개의 문서에서 ‘지휘’라는 용어는 30번 등장하므로 wf_(i=지휘) = 30이 된다. 이는 wf_i ≥ 1인 것이 성립하는 것이므로 정규화 공식 idf(i, j) = (1 + log(wf_{ij})) log $\frac{N}{df_i}$ 을 사용할 수 있다. N은 총 문서의 개수로 20이 되고, ‘지휘’라는 용어가 등장하는 문서는 8개이므로 df_(i=지휘) = 8이 된다. 따라서 ‘지휘’라는 단어의 빈도수를 정규화한 값은 idf(지휘, 모든 문서) = (1 + log 30) × log $\frac{20}{8}$ = 0.9857 이 된다(본 연구에서는 j를 한 개의 특정 문서가 아닌 전체문서로 정규화 실시).

다음 <표 2>는 본 연구에서 작성한 TDM 및 정규화 결과의 일부이다. 이는 5개 재난대응 영역 중 ‘지휘/통제’에 관련된 용어와 문서의 빈도수를 산출하고, 정규화한 값을 나타낸다.

4.2 유클리디안 거리와 SNA의 근접 중심성

TDM 구성 및 정규화를 토대로 용어들의 유클리디안 거리를 계산하고 SNA 중 근접 중심성을 계산한다. 이를 통해 Object와 Instance 간의 근접 정도, Class와 Object 간의 근접 정도를 파악할 수 있다.

<표 2> TDM(지휘통제 영역)

Term		재난대응												
Document		지휘통제												
		책임	권한	보고	지시	지휘	통제	예/경보	혼선	대책회의	대책본부	조치	예측	역할
C1	D1	1		2		1		4			2	1		1
	D2													
	D3				1							1		
	D4	2	1	4	2				3		5			1
C2	D1	1			2		1	1		1		1		1
	D2											1		
	D3											1		
	D4	4	1	8	5	10	14	13	1	1	6	13	3	10
C3	D1				1							5		
	D2					1			1	2	9	5		
	D3			2	7	11	2	2	6		5	2		
	D4					4			7	1	3	2	2	
C4	D1	8			1	1	3					1		
	D2	11		2		1	1		1	1	3	1	1	
	D3	2			2									
	D4			1			2			1	4	5		
C5	D1	2										1		
	D2													
	D3	2												
	D4	2				1	1		2					
총 용어 빈도수		35	2	19	21	30	24	20	21	8	37	50	8	15
정규화		0.7658	1.3010	1.1915	0.9241	0.9857	1.0852	1.6084	1.0588	0.8677	1.0220	0.3372	1.3302	1.3101

주) C: 사고사례, D: 문서(문서 명은 참고문헌 참조).

4.2.1 유클리디안 거리

용어 간의 거리 계산은 앞서 언급한 유클리디안 거리공식을 이용한다. 본 연구에서는 2차원에서의 유클리디안 거리를 구하기 때문에 $n = 2$ 이다. 유클리디안 거리공식에 대입할 값은 각 용어의 정규화 값(X축)과 문서 빈도수(Y축)이다.

본 연구에서는 Object-Instance 간의 거리, Class-Object 간의 거리를 파악한다. 먼저, 5개의 Object와 각 영역에 있는 Instance 사이의 유클리디안 거리를 모두 구한다. 이때 비교 용어 중 더 상위 계층에 있는 용어를 ‘기준’으로 보고 정규화 값과 문서 빈도수를 모두 0으로 대입한다. 예로, ‘지휘/통제’ 영역과 ‘지휘’ 사이의 거리를 구할 때, object인 ‘지휘/통제’가 instance인 ‘지휘’보다 더 상위 계층이기 때문에 기준으로 보고 X, Y값에 모두 0을 대입한다. 두 용어간의 거리를 계산해보면 $d(\text{지휘/통제}, \text{지휘}) = \sqrt{(0 - 0.9857)^2 + (0 - 8)^2} = 8.0605$ 이 된다. 즉, Object와 모든 Instance 사이의 거리를 구할 때, 상위 계층인 Object의 값은 항상(0, 0)이 되며, 나머지 Instance에만 정규화 값과 문서 빈도수를 대입하게 된다.

Class와 Object 간의 거리 비교 시에는 각 Object별 유클리디안 거리를 해당 영역 모든 Instance 유클리디안 거리의 합으로 한다. Object와 Instance 사이의 유클리디안 거리는 <표 3>에 나타내었다.

4.2.2 SNA의 근접 중심성

$$\text{Instance의 근접 중심성은 공식 } C_i = \frac{\sum_{o \neq i} \frac{1}{d(o, i)}}{n-1}$$

에 따라 계산한다. $d(o, i)$ 는 object와 instance의 최단 거리를 의미하므로, 앞서 계산한 유클리디안 거리를 대입한다. N은 전체 노드 수로, 각 5개 영역에 해당하는 총 용어의 개수가 된다.

단, 본 연구에서는 instance 각각에 대한 근접 중심성을 계산하므로, $C_i = \frac{1}{n-1} \frac{d(o, i)}$ 로 계산한다.

예로, ‘지휘’라는 용어의 근접 중심성은 $C_{\text{지휘}} = \frac{1}{\frac{8.0605}{13-1}} = 0.0103$ 이다. 값을 구하려는 Instance가 어느 object의 하위에 속하는지에 따라 n은 달라진다. Object의 근접 중심성을 구할 때는 $C_o = \frac{\sum_{o \neq i} \frac{1}{d(o, i)}}{5-1}$ 을 사용한다(Object는 총 5개 영역을 나타내므로, $n = 5$ 가 된다).

4.2.3 마인드맵(Mind Map)

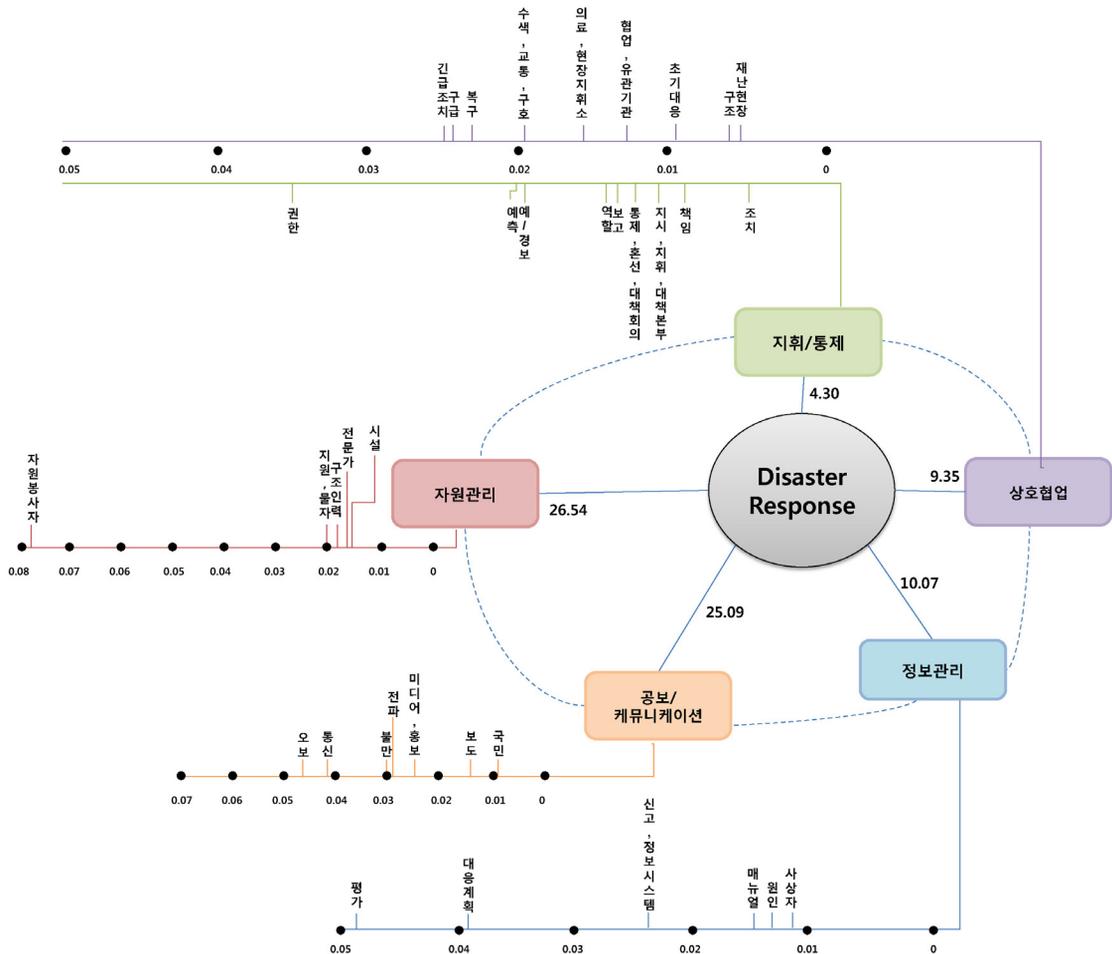
용어의 근접 중심성을 계산한 후 계층별 용어 간의 거리를 한눈에 파악할 수 있도록 마인드맵으로 표현한다. 마인드맵의 중심(시작점)은 Class인 ‘재난대응’이 된다. 본 연구에서 마인드맵은 2가지 측면(Class-Object, Object-Instance)으로 비교할 수 있도록 표현하였다(<그림 1> 참조).

먼저, Class-Object 간의 비교다. ‘재난대응’이라는 용어와 5개의 Object 용어 간의 비교가 가능하다. 지휘/통제, 상호협업, 정보관리, 공보/커뮤니케이션, 자원관리 중에서 재난대응과 가장 가까운 영역은 근접성 수치가 가장 작은 ‘지휘/통제’ 영역이다. 도출한 값이 작을수록 용어 간의 거리가 가깝다고 할 수 있다. 따라서 재난대응과 가장 가까운 영역은 지휘/통제(4.3), 상호협업(9.35), 정보관리(10.07), 공보/커뮤니케이션(25.09), 자원관리(26.54) 순이 된다. 본 연구에서는 Class와 5개의 Object 간의 거리를 쉽게 비교할 수 있도록 비교 수치를 α 로 치환하였다($\alpha = \text{근접 중심성} \times 10^4$).

이뿐만 아니라 Object와 Instance 간의 거리 비교도 가능하다. ‘지휘/통제’ 영역에 해당하는 13개의 용어 중 가장 ‘지휘/통제’ 영역과 근접한 용어는 근접 중심성수치가 제일 낮은 ‘조치(0.0056)’이다. ‘지휘/통제’ 영역 내 용어들의 근접 중심성 수치 비교를 통해 어떤 용어가 ‘지휘/통제’ 영역과 가장 근접한지를 파악할 수 있다(Instance의 근접 중심성 수치는 <표 3> 참조).

〈표 3〉 유클러리다인 거리(U.D) 및 근점 중심성(C.C) 계산

Object	지휘/통제	Total U.D:		상호협업	Total U.D:		정보관리	Total U.D:		공보/커뮤니케이션	Total U.D:		자원관리	Total U.D:		
		U.D:	C.C:		U.D:	C.C:		U.D:	C.C:		U.D:	C.C:		U.D:	C.C:	
Instance		92.6271	0.002698994		58.9100	0.004243761		58.6774	0.004260587		57.3388	0.004360053		49.8239	0.005017671	
	채임	U.D: 10.0293	C.C: 0.0083	재난현장	U.D: 13.0122	C.C: 0.0064	원인	U.D: 12.0134	C.C: 0.0139	진과	U.D: 4.2469	C.C: 0.0294	물자	U.D: 10.0412	C.C: 0.0199	
	권한	U.D: 2.3859	C.C: 0.0349	현장지휘소	U.D: 5.1296	C.C: 0.0162	사상자	U.D: 14.0069	C.C: 0.0119	홍보	U.D: 5.1543	C.C: 0.0243	지원	U.D: 10.0202	C.C: 0.0200	
	보고	U.D: 6.1172	C.C: 0.0136	초기대응	U.D: 9.0339	C.C: 0.0092	신고(접수)	U.D: 7.1006	C.C: 0.0235	오보	U.D: 2.6762	C.C: 0.0467	전문가	U.D: 12.0141	C.C: 0.0166	
	지시	U.D: 8.0532	C.C: 0.0103	구호	U.D: 4.2372	C.C: 0.0197	정보시스템	U.D: 7.0810	C.C: 0.0235	미디어	U.D: 5.1543	C.C: 0.0243	구조인력	U.D: 11.0281	C.C: 0.0181	
	지휘	U.D: 8.0605	C.C: 0.103	수색	U.D: 4.2372	C.C: 0.0197	매뉴얼	U.D: 11.0266	C.C: 0.0151	국민	U.D: 13.0098	C.C: 0.0096	시설	U.D: 13.0113	C.C: 0.0154	
	통제	U.D: 7.0836	C.C: 0.0118	구조	U.D: 12.0194	C.C: 0.0069	대응계획	U.D: 4.2268	C.C: 0.0394	불만	U.D: 4.1726	C.C: 0.0300	자원봉사자	U.D: 2.5625	C.C: 0.3780	
	예/경보	U.D: 4.3112	C.C: 0.0193	구급	U.D: 3.3850	C.C: 0.0246	평가	U.D: 3.4546	C.C: 0.0482	통신	U.D: 3.0648	C.C: 0.0408				
	혼란(혼선)	U.D: 7.0796	C.C: 0.0118	긴급조치	U.D: 3.3386	C.C: 0.0250				보도	U.D: 9.0346	C.C: 0.0138				
	대책회의	U.D: 7.0536	C.C: 0.0118	협업	U.D: 11.0207	C.C: 0.0135				인터뷰	U.D: 3.3105	C.C: 0.0378				
	대책본부	U.D: 8.0650	C.C: 0.0103	유관기관	U.D: 6.1687	C.C: 0.0135										
	조치	U.D: 15.0038	C.C: 0.0056	의료	U.D: 5.1772	C.C: 0.0161										
	예측	U.D: 4.2154	C.C: 0.0198	복구	U.D: 3.5181	C.C: 0.0237										
	역할	U.D: 5.1688	C.C: 0.0139	교통	U.D: 4.2268	C.C: 0.0197										



〈그림 1〉 근접 중심성 기반의 재난대응 용어 마인드맵

V. 결론

본 연구의 목적은 예방, 대비, 대응, 복구단계로 구성되는 재난관리 중에서 ‘재난대응’에 관련된 용어를 선정하고 용어 간의 근접 중심성을 계산해 어떤 용어가 재난대응 영역에서 핵심용어인지를 파악하는 것이다. 이를 위해 대응과정의 내용을 담고 있는 다양한 사고사례 문서에서 텍스트 마이닝 개념 중 하나인 TDM을 바탕으로 용어의 빈도수를 계산한 후 SNA의 근접 중심성을 분석하였다.

Class(재난대응)와 5개의 Object 간 비교에서는 ‘지휘/통제’ 영역이 재난대응 영역 중에서도 핵심 영역으로 나타났다. 실제로 재난대응 시 가장 많이 언급되는 문제 중 하나가 ‘지휘/통제’와 관련된 내용이다. 현재 우리나라는 재난이 발생하면 현장을 진두지휘할 명확한 지휘체계가 없어서 대응 주체에 혼선이 생기고 있으며, 재난대응에 참여하는 유관기관의 역할도 바로잡혀있지 않아 문제가 되고 있다. 지휘/통제 영역의 경우 재난현장을 총괄하는 중요한 부분이고, 얼마나 지휘/통제가 잘 이루어지냐에 따라 상호협업 등 타 영역

에 연쇄적인 영향을 미치기 때문에 무엇보다 재난대응에서 핵심적인 부분이라고 할 수 있다(국민안전처, 2014, 2015). 지휘/통제 영역의 뒤를 이어서 상호협업, 정보관리, 공보/커뮤니케이션, 자원관리의 순으로 재난대응과 밀접하게 연관되어 있음을 확인할 수 있었다.

5개 Object와 Instance 간의 비교에서는 지휘/통제, 상호협업 영역의 용어들이 나머지 3개 영역의 용어보다 해당 영역과 근접하게 연결된 것을 확인할 수 있었다. 특히 지휘/통제 영역에서는 ‘조치’ 및 ‘책임’이 가장 낮은 근접 중심성을 나타내고 있어 해당 영역의 핵심용어로 볼 수 있으며, 상호협업 영역에서는 ‘재난현장’과 ‘구조’를 핵심용어로 파악할 수 있다. 나머지 영역 또한 Instance와의 근접 중심성을 비교하여 어떤 용어가 해당 영역의 핵심용어가 되는지를 파악할 수 있다.

본 연구에서는 용어의 근접 중심성 계산을 통해 어떤 영역이 재난대응과 가장 밀접하게 연관되어있는지, 또 재난대응 각 영역 내에서 어떤 용어가 가장 핵심적인지를 파악하였다. 재난대응 핵심용어는 다양한 측면에서 활용할 수 있다. 먼저, 재난현장에서 사용하는 대응용어를 표준화하기 위한 기초자료로 이용할 수 있다. 현재 부처/기관별로 사용하는 법, 제도 등이 달라 재난현장에서 사용하는 용어 또한 상이한 실정이며, 재난대응 매뉴얼에서도 용어의 의미는 같으나 서로 다른 용어로 표현하고 있다. 이러한 문제는 재난대응 협업과정에서 기관 간 의사소통을 어렵게 하므로 용어 통일이 필요하다. 이때 우선적으로 어떤 용어를 표준화할 것인지에 대한 기준을 핵심용어 기반으로 선정할 수 있다. 본 연구에서는 재난대응 과정을 중점적으로 다루고 있는 사설 등의 신문기사, 정부 간행물로 데이터 셋을 구성하였다. 하지만 인명피해 및 사회적 파급 정도를 고려하여 사례를 선정하였기 때문에 데이터 수가 적다는 한계가 있었으므로 향후에는 지침, 매뉴얼, 법령 등 더 넓은 범위에서 문서를 수집하여 용어를 분석하면 핵심용어 도출 결과가 좀 더 명

확할 것으로 판단된다. 또한, 온톨로지 개념에 따라 용어들을 계층적으로 분류하였기 때문에 재난대응에 대한 다양한 자료들을 추적하고 검색하는데 용어의 분류체계를 활용할 수 있다. 정보시스템 및 소프트웨어 개발 분야에서 온톨로지는 정보 또는 지식을 조직하는 체계적인 방법이다. 온톨로지를 기본으로 한 시스템은 여러 도메인에서 지식의 배열을 명료하게 만든다. 이러한 시스템은 여러 전문가들 간의 소통이 원활하게 이루어지게 한다(소방방재청, 2010). 이 밖에도 재난대응 핵심용어는 사고 대응 시나리오 작성에 활용할 수 있다. 핵심용어를 통해 만들어진 재난대응 시나리오는 재난 담당자에게 다양한 교육/훈련의 기회가 될 수 있으며, 추가적인 문제를 발견해 피드백 할 수 있도록 해준다. 이로써 향후 발생할 수 있는 유사 사고를 대비하기 위한 실무자의 역량을 강화할 수 있을 것으로 기대된다.

참 고 문 헌

- [1] 국립재난안전연구원, 복합적 미래예측방법론 분석을 통한 미래재난예측기법 개발, 2013.
- [2] 국민안전처, 안전혁신마스터플랜, 2015.
- [3] 국민안전처, 재난사례 분석모델 연구 및 사례분석, 2014.
- [4] 국민안전처, 한국형 재난현장 대응·수습 표준체계 개발연구, 2015.
- [5] 김동래, 권기석, 정석봉, “키워드 네트워크 분석을 통한 경영학 연구의 지식구조 분석”, *의사결정학연구*, 제23권, 제1호, 2015, pp. 111-125.
- [6] 김인재, 최재원, 김기환, 민금영, “전문가 그룹의 소셜 네트워크 분석: 국내 학술지 공저자 및 심사자 네트워크를 중심으로”, *한국 IT 서비스학회지*, 제13권, 제1호, 2014, pp. 181-196.
- [7] 김주호, *키워드 네트워크 분석을 통한 IT분야 국가연구개발 동향 분석에 관한 연구* (박

- 사학위논문), 숭실대학교, 2015.
- [8] 김현희, 안태경, “온톨로지를 이용한 인터넷 웹 검색에 관한 실험적 연구”, *정보관리학회지*, 제20권, 제1호, 2003, pp. 417-455.
- [9] 김혜진, “사회연결망(Social Network Analysis: SNA)을 이용한 스포츠 경기분석”, *한국체육 측정평가학회지*, 제9권, 제1호, 2007, pp. 99-112.
- [10] 소방방재청, 재난전조정보 관리체계 구축 등을 통한 인적재난 피해저감 방안 연구 보고서, 2010.
- [11] 이영재, 이성수, “텍스트 마이닝 기반의 인적재난사고사례 신뢰도 측정연구”, *정보시스템연구*, 제20권, 제3호, 2011, pp. 63-79.
- [12] Gruber, T. R., “A translation approach to portable ontology specifications”, *Knowledge Acquisition*, Vol.5, No.2, 1993, pp. 199-220.
- [13] Turban, E., *Decision Support and Business Intelligence Systems*, 9th Edition, Pearson Education, 2010, pp. 302-308.
- [14] Wang, L., Y. Zhang, and J. Feng, “On the euclidean distance of images”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.27, No.8, 2005, pp. 1334-1339.

〈부 록〉

*사고사례 문서

세월호 침몰사고	D1	진도 여객선 침몰...해수부, 해양사고 '심각' 경보 발령(sbs, 2014. 4. 16)
	D2	여객선 침몰사고...왜 이렇게 인명피해 큰가(연합뉴스, 2014. 4. 16)
	D3	초기 우왕좌왕·구명조끼 착용 늦어...사고도 대응도 후진국형(국민일보, 2014. 4. 17)
	D4	세월호 사건에서의 정부대응의 문제와 개선방안(한국행정연구원, 2014)
구미불산 누출사고	D1	초동대응 총체적 부실 '화 키웠다'(조선일보, 2012. 10. 05)
	D2	구미 불산 누출 사고, '화학강국'이 부끄럽다(조선일보, 2012. 10. 17)
	D3	구미 불산 유출 사고 '제2의 페놀 사태'로 다뤄야(경향신문, 2012. 10. 04)
	D4	유해화학물질 누출 사고시 대응방안(2013, 김제소방서연구반)
삼풍백화점 붕괴사고	D1	삼풍백화점 대참사(조선일보, 1995. 6. 30)
	D2	삼풍백화점 붕괴(중앙일보, 1995. 7. 1)
	D3	백화점 붕괴참사, 구명 난 지휘체계(한겨레, 1995. 7. 4)
	D4	삼풍백화점 붕괴(국방재난백서, 2012. 6. 25)
판교환풍구 붕괴사고	D1	건축법에 시설물 안전규제 강화(데일리한국, 2014. 10. 19)
	D2	판교환풍구 붕괴사고 집중분석(주간현대, 2014. 10. 27)
	D3	판교환풍구 붕괴사고 재구성(서울대건설환경연구소, 2014. 11. 11)
	D4	집객현장 안전대책 추진방안(경기개발연구원, 2014)
대구 지하철 화재사고	D1	어떻게 이런 참사가(국민일보, 2003. 02. 18)
	D2	국내 지하철 대형재난 무방비(연합뉴스, 2003. 02. 18)
	D3	지하철 참사로 본 자화상(문화일보, 2003. 02. 19)
	D4	대구 지하철 참사로 본 재난관리의 문제점 및 개선방안(경남발전연구원, 2013)

Analyzing Disaster Response Terminologies by Text Mining and Social Network Analysis

Seong Kyung Kang* · Hwan Yu** · Young Jai Lee***

Abstract

This study identified terminologies related to the proximity and frequency of disaster by social network analysis (SNA) and text mining, and then expressed the outcome into a mind map. The term-document matrix of text mining was utilized for the terminology proximity analysis, and the SNA closeness centrality was calculated to organically express the relationship of the terminologies through a mind map.

By analyzing terminology proximity and selecting disaster response-related terminologies, this study identified the closest field among all the disaster response fields to disaster response and the core terms in each disaster response field. This disaster response terminology analysis could be utilized in future core term-based terminology standardization, disaster-related knowledge accumulation and research, and application of various response scenario compositions, among others.

Keywords: *Disaster Response Terminologies, Text Mining, Social Network Analysis, Closeness Centrality, Ontology*

* Department of MIS, Business College, Dongguk University

** Department of MIS, Business College, Dongguk University

*** Corresponding Author, Department of MIS, Business College, Dongguk University

◎ 저 자 소 개 ◎



강 성 경 (hshs4123@naver.com)

동국대학교 대학원 경영정보학과 석박사 통합과정에 재학 중이며, 동국대학교 경영정보학과를 졸업했다. 주요 관심분야는 재난관리, 의사결정, Data Mining 등이다.



유 환 (hani@ibk.co.kr)

동국대학교 대학원 경영정보학과 박사과정에 재학 중이며, 현재 IBK기업은행 부지점장(PB)으로 재직 중이다. 연세대학교에서 법학을 전공, 석사학위를 취득하였으며, 주요 관심분야는 의사결정, Data Mining, 금융자산관리 등이다.



이 영 재 (yjlee@dongguk.edu)

동국대학교 경영대학 경영정보학과 교수로 재직 중이다. 미국 George Washington University에서 정보관리 전공으로 이학박사학위를 취득했으며, (사)한국데이터베이스학회, (사)한국재난안전학회, (특)기업재해경감협회 회장 및 세계재난관리학회(TIEMS) 부회장을 역임하였다. 주요 관심분야는 의사결정, 재난관리, 산업설비 자산관리 등이다.

논문접수일 : 2015년 10월 06일

1차 수정일 : 2016년 01월 08일

게재확정일 : 2016년 02월 25일

2차 수정일 : 2016년 02월 24일