협업형 재난안전 R&D 탐색을 위한 편집거리 기반 과제 유사도 정량화

Quantification of Similarity Using the Edit-distance Method for Searching Cooperative Programs Related to Disaster and Safety Management

이보람* · 윤상원** · 정재학***

Lee, Boram*, Yoon, Sang Won**, and Chung, Jaehak***

Abstract

In this study, the methodology of evaluating the similarity for R&D programs related to the disaster and safety management (hereafter 'disaster R&D') was proposed to enhance the efficiencies of R&D performances and to encourage the inter-agencies cooperative R&D programs. In the similarity evaluation scheme of NTIS (National Technology Information Service), several parameters should be identified before their applications and the characteristics of disaster-related R&D, so-called 'complexity' could not be taken into considerations. To solve the problem, the methodology based on the edit-distance and Gephi were introduced in this study. The edit-distance is the frequently used for estimating the similarity between interests and can be easily applied by simple preparations. To verify the performances of the methodology proposed in this study, as an example, 91 disaster R&Ds were applied. As the results, the new methodology could classify the similar disaster R&Ds effectively and efficiently with respect to their aims and contents.

Key words : Disaster-related R&D Programs, Edit-distance, Inter-agencies Cooperative R&D Programs, Gephi, Similarity

요 지

본 연구에서는 재난안전과 관련된 연구개발(R&D)의 효율성을 강화하고, 부처간 협력적 재난안전 R&D를 독려하기 위한 방법론으로서 편집거리 기반 재난안전 R&D 과제 유사도 측정모형을 제안하였다. 기존 NTIS 유사도 측정 시스템의 경우 특정 파라미터만을 사용하여 과제 간 유사도를 측정하고 있다. 그러나 해당 방법의 경우 재난유형, 기술분류, 핵심역량 등 재난안전 분야의 특성을 고려할 수 없다는 한계점이 있다. 본 연구에서는 이러한 한계점을 극복하기 위해 편집거리 기반의 유사도 측정 방법론 및 지파이(Gephi)를 활용한 군집화 방법을 제안하였다. 본 연구에서 제안한 방법론의 성능을 평가하기 위해 91개의 재난안전 R&D 사업을 적용하였으며, 각 군집에 포함된 과제들을 비교·분석함으로써 유사·중복 및 협업 가능성을 검토하였다.

핵심용어 : 재난안전 R&D, 편집거리, 다부처 협업 과제, 지파이, 유사도

*정회원, 국립재난안전연구원 R&D기획평가센터 선임연구원(E-mail: boram1024@korea.kr)

Member, Researcher, R&D Innovation Center, National Disaster Management Research Institute

**교신저자, 정회원, 국립재난안전연구원 R&D기획평가센터 연구원(Tel: +82-53-928-8034, Fax: +82-52-928-8149 E-mail: swyoon89@korea.kr)

Corresponding Author, Member, Researcher, R&D Innovation Center, National Disaster Management Research Institute

***국립재난안전연구원 시설연구관(E-mail: blueboat@korea.kr)

Ph.D., Senior Researcher, R&D Innovation Center, National Disaster Management Research Institute

1. 서 론

최근 재난 및 안전사고는 신종재난의 등장, 위험 강도의 증가, 도시화산업화 등 노출 요인의 변화 등 많은 요인들과 함께 발생하고 있다. 과거의 전통적 재난 및 안전사고의 패턴은 점차 ‘다양성’이 커지고, 재난 및 안전사고의 관리 부처 및 이해 당사자들간 관계도 복잡성이 커지고 있는 상황이다. 자연재난과 사회재난의 복합적인 형태인, 이른바 나텍(NATECH, NATural and TECHnical disaster)처럼 하나의 재난은 다른 분야에 영향을 주어 또 다른 재난의 유발요인으로 작용하게 되며, 1개 재난 이벤트에 대한 정의 자체가 어려운 상황에 이르게 된다(Vetere Arellano et al., 2003). 이같은 문제점들은 ‘복잡성’ 또는 ‘복합성’의 문제로 귀결된다. 사회적 문제인 재난안전관리의 이슈는 더 이상 하나의 부처와 관계되어 발생하지 않으며, 다부처(multi-agency)의 관점에서 접근해야만 한다. 아울러 이처럼 복잡성이 높은 문제들을 구조화하고 국민의 안전과 직결되는 재난안전문제를 신속하게 해결하기 위해서는 효율적 방법에 의거 이에 관해 답을 찾아야 한다.

해외 재난안전관리 선진국들은 오래전부터 다부처 차원에서 연구개발 활동을 수행해 오고 있다. 미국의 경우 국가과학기술위원회(NSTC, National Science and Technology Council)의 환경 및 자원위원회(Committee on Environment and Natural Resource)에 연방다부처협의체(Federal Interagency body)로서, 재난저감소위원회(SDR, Subcommittee on Disaster Reduction)를 두고 ‘재난 저감을 위한 대도전(Grand Challenges for Disaster Reduction)’이라 불리는 장기 계획을 2005년 수립하였다. 또한 중점적으로 해결해야 할 문제점을 6개의 전략으로 구성하고, 자연재난, 사회재난 등 15개 주요 분야(유형)에 상술한 6개 전략별 단기(1~2년), 중기(2~5년), 장기(5년~) 실행계획을 세워 재난안전과 관련된 각종 연구개발을 수행하고 있다. 특히 재난안전 문제 해결을 위해 16개 부·처·청 및 주정부 협업을 통해 기술개발을 추진하고 있는 점을 주목할 필요가 있다(NSTC, 2005).

일본 내각부의 전략적 혁신 증진 사업(SIP, Cross-ministerial Strategic Innovation Program)에서는 부처간 협업을 강조하고, 관련 사업 예산으로 2016년 500억엔의 예산을 투입하였다. 본 사업에서는 재난안전과 관련된 2개 분야를 포함하여 사회·경제적 문제점을 유발하는 총 11개 분야에 각 부처의 협업으로 사업 기획을 하였으며, 현재 관련 연구개발을 수행하고 있다(CAO, 2016).

최근 우리나라 과학기술정보통신부(2017)에서도 2018년 정부 투자 방향 및 기준(안)을 통해 재난안전과 관련된 R&D에서 적용될 수 있는 3가지 협업모델을 제안하고 새로운 정책기조를 형성하였다. 또한 제3차 재난 및 안전관리 기술개발 중합계획에서도 ‘범부처 협의체’를 통한 협업과제 발굴을 비중 있는 전략으로 제시하고 있는 상황이다(MOIS,

2017). 하지만 협업 대상이 되는 과제 또는 사업간 조정을 위한 평가도구의 부재 및 각 부처의 참여 의지 부족 등 실질적 협업 과제 추진은 어려운 상황이다.

이와 같은 상황을 해결하기 위해서는 과제 간 비교를 정량화하여 협업 과제 가능성을 모색할 수 있는 연구 및 방법론 개발이 선행되어야 한다. 그러나 현재까지 진행된 대부분의 선행연구들은 과제 간 유사·중복성 탐지 및 정확도 향상에 기반하고 있어 본 연구에서 제시한 문제점 해결에 근본적인 답을 제시하기 어려운 실정이다.

관련 연구들을 간략히 살펴보면 Choi et al.(2003)은 키워드 추출 기반의 전체 문서 마이닝(Full Text Mining) 기법을 적용, 연구과제간 유사도 비교를 수행하였으나, 키워드를 통한 연구과제 내용의 신뢰성 확보에 어려움이 있음을 언급하였다(KISTEP, 2013). 또한 Park et al.(2009)은 K-최근접 문서 기법(K-nn)을 적용하여 유사과제를 파악하는 알고리즘을 제시하였으나 불용어사전, 동의어, 유사어 처리 문제를 해당 연구의 한계점으로 제시하였다.

지속적인 연구개발과제 유사·중복성 이슈 및 선행 연구에 따라 정부는 과제단위의 유사분석시스템을 국가과학기술지식정보서비스(NTIS, National Science & Technology Information Service)에 구축·운영해오고 있다. 또한 국가연구개발사업관리 등에 관한 규정에 의거 과제 기획 시 유사·중복성을 사전에 방지하기 위해 NTIS를 통한 유사성 검토를 의무화 하고 있다. 그러나 규정을 통한 시스템 사용 의무는 사용자 편의성, 유사·중복성 검토의 정확성 등에 관한 이슈를 지속적으로 발생시키고 있으며 키워드 비교를 통한 유사도 측정은 관련 항목들이 변경될 경우 이를 정확히 측정하지 못하는 한계가 있다(Jung et al., 2011).

이에 Jung et al.(2011)은 NTIS를 통한 유사도 측정 시 사용되는 6개의 항목(과제명, 연구책임자, 연구목적, 연구내용, 기대효과, 키워드)에 연구보고서 초록의 키워드 추출을 통한 유사도 측정 개선모형을 제안하였으나 유사정도에 따른 객관적 분류 체계를 제시하지 못하였다. 또한 키워드 추출을 통한 연구주제망 적용 여부에 따라 유사도 정확성의 편차가 존재하므로 이에 대한 추가 분석이 필요한 상황이다.

키워드 기반의 유사도 분석은 결과의 정확성에 관한 문제가 항상 수반되므로 이를 해결하기 위해 다양한 방안들이 고민되었다. Kim et al.(2012)은 연구개발 과제의 과학기술표준분류를 이용한 유사도 분석 기법을 제안하여 정량적인 결과를 제시하였다. 또한 Kim et al.(2014)은 정부 R&D 특허기술동향조사사업의 DB를 활용하여 기존 과제와 신규 과제간의 유효특허 조사·분석에 따른 유사도 측정 모델을 제안하였다. 해당 연구들의 경우 키워드 기반의 유사도 검색 알고리즘에 비해 물리적인 효율성을 나타내고 있으나 결과의 시각화를 통한 검증의 효율성에 한계가 존재한다.

상술한 내용과 같이 기존 연구들은 과제 간 유사·중복성 검토 및 정확도 향상에 집중하고 있어 ‘협업과제 추진’을

위한 방향제시 및 의사결정 지원을 위한 평가 도구로서의 활용 가능성이 낮은 실정이다. 또한 재난안전 분야의 다양한 특성(재난 유형, 성과 유형 등)을 자유롭게 탐색 할 수 있는 방법론의 개발이 시급하다.

따라서 본 연구에서는 재난안전 분야의 다양한 특성값을 고려하고 손쉽게 유사도를 측정하기 위해 편집거리(edit-distance) 기반의 유사도 정량화 기법을 제안하고자 한다. 키워드 기반 유사도 검색의 경우 상술한 정확성 이슈와 함께 처리과정이 비교적 복잡하다는 문제점이 있는 반면 편집거리 기법은 문자열 전체를 비교하여 유사도를 정량화 할 수 있다. 또한 군집화 분석도구인 지파이(Gephi)를 활용하여 과제간 군집화 및 시각적 분석을 통해 협업가능한 과제군(Cluster)을 제시할 수 있는지 평가하고자 한다.

2. 편집거리 기반 R&D 유사도 측정 및 협업대상 군집화 기법 제안

본 연구에서 제안하는 재난안전 R&D 유사도 측정 및 협업대상 군집화 절차는 크게 4단계로 구분할 수 있다. 첫 번째 단계는 유사도 측정을 위한 데이터 구축 단계로, NTIS의 과제 검색 시스템을 활용하여 재난안전 R&D 과제 정보를 마이크로소프트 엑셀(excel) 형식으로 구축한다. 두 번째 단계는 오픈소스 통계패키지인 R을 활용, 문자열 편집거리 기반으로 과제 간 유사도를 계산하고 이를 과제 수(n) × 과제 수(n) 대칭행렬로 표현한다. 세 번째 단계는 Clustering Mapping 툴인 지파이를 사용하여 과제 간 군집화를 수행하고 이를 시각화 한다. 마지막으로 각 군집에 포함되어 있는 과제들 간의 키워드, 연구내용, 기대효과 등을 검토하여 해당 군집의 속성을 분석한다(Fig. 1).

Fig. 1. Flowchart of Similarity Estimating Method Revised in this Study

2.1 입력자료 구축

본 연구는 문자열의 편집거리 계산에 기반하여 재난안전 R&D 과제의 유사도 측정 모델을 구축하였다. 이를 위해, 1) 연구기획자(모델 사용자) 편의를 고려한 입력자료 구축의 용이성, 2) 유사도 측정 시 재난안전 R&D 사업의 특성을 반영할 수 있는 항목 구성의 유연성을 고려하여 유사도 측정 모델을 설계하였다.

2.2 문자열 편집거리 정량화(Edit-distance)

편집거리 알고리즘은 두 개의 문자열이 같아지기 위해 몇 번의 삽입(insertion), 삭제(deletion), 교체(substitution)가 이루어져야 하는지 그 최소 개수를 구하는 알고리즘이다. 두 문자열의 유사도를 정량화 하는데 사용할 수 있는 알고리즘으로 다량의 문서(텍스트)들이 서로 어느 정도의 유사도를 갖는지 확인 할 수 있다(Levenshtein, V.I. 1965).

예를 들어, 'HBCEFG'이라는 문자열을 'ABCDEF'으로 변환시키기 위해 필요한 최소 편집 연산 횟수는 Fig. 2를 통해 설명이 가능하다. 'HBCEFG'을 'ABCDEF'으로 변환할 때 필요한 최소한의 조치는 우선 'HBCEFG' 문자열의 첫 번째 위치에 있는 'H'를 'A'로 교체한다. 다음으로 'D'를 문자열에 추가한 뒤 마지막의 'G'를 삭제함으로써 편집조치가 총 3번 작동함을 확인할 수 있다.

	H B C E F G	Edit operations	Edit result
1st edit operation		Replace and 'H' with 'A'	A B C E F G
2nd edit operation		Insert 'D'	A B C D E F G
3rd edit operation		Delete the last 'G'	A B C D E F
	A B C D E F		

Fig. 2. Conceptual Explanation on the Levenshtein Metric of Edit-distance

편집거리 계산에 따른 문자열 간 유사도 측정방법은 Eq. (1)로 정의할 수 있다.

$$D_{ED}(p, p') = 1 - \frac{\delta}{m + n} \quad (1)$$

Eq. (1)에서 m 과 n 은 각각 p 와 p' 내에 존재하는 단어의 개수를, δ 은 p 와 p' 사이의 편집거리를 의미한다. 편집거리 δ 를 구할 때 편집 연산(삽입, 삭제, 교체)을 사용할 수 있으며, 각 연산의 동작을 1로 설정하였다.

본 연구에서는 편집거리 계산을 위해 오픈소스 통계패키지인 R의, 'Adist' 패키지(Laurikari, V. 2001)를 사용하여 과제 간 편집거리를 계산하였다. 통계패키지를 활용한 편집거리 계산 후 유사도 측정방법에 따라 최종 유사도를 계산하였으며 유사도 값이 1에 가까워질수록 과제 간 유사성이

높도록 설계하였다.

2.3 유사도 계산 구조화

본 연구에서는 재난안전 R&D 과제 간 유사도 정량화 비교 및 향후 군집화 분석을 위해 결과값을 대칭행렬(Symmetric matrix)로 구조화 하였다(Table 1). 과제 수(n)에 따른 $n \times n$ 대칭행렬은 계산결과의 검증 및 군집화 분석을 위한 입력자료로 활용이 가능하다.

Table 1. Structure of Symmetric Matrix for Inter-program Similarity

	p(1)	p(2)	p(3)	p(4)	...	p(n)
p(1)	N/A	N/A	N/A	N/A	N/A	N/A
p(2)	-					
p(3)	-					
p(4)	-					
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
p(n)	-	-	-	-	...	

2.4 군집화 분석

본 연구에서 수행하고자 하는 과제 간 군집화(Clustering)의 목적은 편집거리 기반의 유사도 정량화를 통해 각 사업간 협업 가능성을 파악하고자 함이다. 이를 효율적으로 수행하기 위해서는 군집대상 시각화를 지원할 수 있는 도구의 탐색과 선택이 중요한 요소로 작용한다.

군집화 또는 연결망 분석도구로서 UCInet, JUNG, Netminer, NetDraw 등 현재 다양한 프로그램들이 개발되어 활용되고 있다(Jung and Kim, 2016). 본 연구에서는 효율적인 사용자 인터페이스, 군집망의 상세 특성 분석 능력에 적합한 도구로서 지파이를 사용하였다.

지파이는 데이터 탐색 및 처리를 가능하게 하는 오픈소스 프로그램으로 복잡한 네트워크를 효율적으로 시각화하여 데이터의 의미를 분석하는데 사용된다(Bastian et al., 2009). 지파이는 데이터의 통계적 특성을 노드와 링크로 표현하며, 다양한 편집기능을 통해 정량적 결과의 의미를 시각화하여 효과적으로 전달할 수 있다.

3. 사례 적용

NTIS의 과제 검색 시스템을 통해 2013년부터 2017년까지 수행된 재난안전분야 연구과제 중 1차적으로 120개를 추출하였다. 120개의 과제 중 본 연구에서 사용하고자 하는 입력 정보(연구내용요약)가 상대적으로 잘 구축된 ‘A’연구원 91개의 과제를 선별하여 본 연구에 적용하였다. 본 연구에서는 편집거리 기반의 유사도 계산 시 입력정보의 증가에 따른

정량적 유사도 결과값의 증가를 방지하고자 ‘연구내용요약’만을 사용하여 연구를 진행하였다. 유사도 계산 및 군집화 수행을 위해 91개의 과제 간 비교를 모두 수행하여 91×91의 행렬로 제시하였다. 이를 바탕으로 군집화 분석도구인 지파이를 사용하여 군집분석 및 시각화 결과를 제시하였다. 유사도 결과를 바탕으로 정량적 유사과제 판단, 원자료의 비교·분석을 통해 협업가능성을 검토하였다.

3.1 유사도 정량화

편집거리 기반의 과제 간 유사도 계산을 위해 NTIS에서 제공하는 ‘과제명’, ‘수행기관’, ‘기대효과’ 등 60개의 속성값 중 ‘연구내용요약’을 활용하였으며, 과제 간 유사성의 척도는 유사도 계산값이 1에 가까울수록 유사성이 높아지도록 설계하였다. 91개 재난안전 R&D 과제들의 유사도 정량화 결과 동일 과제 간 비교는 유사도가 1로서 일치하고 있으며 이를 기준으로 대칭을 나타내고 있음이 확인되었다(Fig. 3).

Fig. 3. Symmetric Matrix for Inter-program Similarity Estimated for 91 R&D Programs

유사도 정량화 결과 대칭축을 기준으로 유사도가 1 또는 1에 근사하게 도출된 7개 과제들이 확인되었다. 원인 분석을 위해 원자료의 과제간 비교를 수행한 결과 연차과제 비교에 따른 연구내용의 유사성, 동일한 입력값(연구내용요약)의 비교로 인해 해당결과가 도출된 것으로 확인되었다.

결과값을 바탕으로 정량화된 유사도의 통계적 특성을 파악하고 이에 따른 유사과제 기준의 정량적 가이드라인 제시를 위해 히스토그램을 활용하였다. 유사과제 기준이란 임계값(Threshold)으로 표현할 수 있는데, 이를 토대로 과제 간 유사도 수준을 평가할 수 있다. 즉, 유사과제 식별을 위한 도구로서 활용이 가능하며 향후 과제 간 군집화에 따른 정량적 비교·검증을 가능하게 한다.

앞서 제시된 대칭행렬 중 동일 과제 간 비교값을 축으로 나누어진 두 영역(Area) 중 한 영역을 고려하여 1차 표본을 추출하였다. 영역에 포함되어 있는 총 4,095개의 표본 중 앞서 확인된 유사도 값 1, 0.98(연차과제 및 동일과제)을 제외한 결과 최종 4088개의 표본이 확보되었다. 확보된 표본

을 바탕으로 계급구간을 0부터 0.37까지 0.01로 나누어 히스토그램을 표현하였다(Fig. 4).

Fig. 4. Relative Frequency of Similarity Estimated for 91 R&D Programs

대상 표본의 평균값은 0.18, 중앙값과 최빈값을 나타내는 구간은 0.19로 확인되었다. 해당 결과로부터 본 유사도 정량화에 따른 임계값이 약 0.2 이상일 때 과제 간 유사성에 대한 설명력을 확보할 수 있는 것으로 분석되었다.

본 연구에서는 우선순위 관점에서 ‘평균’보다 높은 유사도를 가지고 있는 것을 대상으로 협업 가능 과제들을 도출하였다. 따라서 본 연구에서 제시한 과제간 유사도의 임계값은 모집단의 변화에 따라 그 값 또한 달라질 수 있다.

3.2 군집 간 유사도 분석 및 협업 가능성 검토

유사도 계산 결과를 바탕으로 지파이를 이용한 군집분석을 실시한 결과 각 과제들은 유사도에 기반하여 21개의 군집으로 분류되었다. 지파이에서 제공하는 다양한 분석방

법 중 본 연구에서는 ‘Force Atlas’ 방법을 활용하여 데이터를 시각화 하였다(Fig. 5).

지파이에 의해 분류된 각 군집들의 유사도를 정량적 유의수준(0.2)과 비교하고 “잠재적 협업가능 과제 군”을 도출하고자 군집별 평균 유사도를 계산하였다(Table 2). 가장 높은 유사도를 보이고 있는 11, 4, 6의 경우 연차과제 간 비교, 동일한 입력값에 따른 군집으로서 ‘협업가능성’에 대한 논의가 어려운 것으로 분석된다. 또한 5, 0 군집은 본 연구에서 제시한 유사도 기준보다 작은 값이 도출되어 ‘협업가능성’ 탐색을 위한 설명력을 확보하기 어렵다. 따라서 본 연구에서는 “잠재적 협업가능 과제 군” 도출을 위해 0, 4, 5, 6, 11을 제외한 나머지 군집들을 대상으로 세부적인 분석을 수행하였다.

Table 2. Average Similarity of Each Cluster

Cluster No.	Ave.	Cluster No.	Ave.
Cluster 11	1.00	Cluster 16	0.27
Cluster 4	0.99	Cluster 14	0.26
Cluster 6	0.98	Cluster 1	0.25
Cluster 17	0.53	Cluster 12	0.24
Cluster 8	0.53	Cluster 9	0.23
Cluster 7	0.30	Cluster 18	0.23
Cluster 20	0.30	Cluster 10	0.23
Cluster 19	0.29	Cluster 13	0.22
Cluster 15	0.28	Cluster 5	0.19
Cluster 3	0.27	Cluster 0	0.18
Cluster 2	0.27

Fig. 5. Results of Inter-programs Clustering using Gephi Application

“잠재적 협업가능 과제 군” 탐색을 위해 분석대상 군집 속 각 과제의 키워드 및 내용을 비교·분석한 결과, 군집별 (Cluster No.) 과제들(Projects)의 특성에 따라 협업 가능성 여부를 구분하였다(Table 3). 우선 협업 가능성이 낮은 것으로 분류된 군집들은 유사과제 및 연차과제 가운데 연구내용, 목적이 상이한 과제가 포함되어 있어 협업의 가능성 예측이 어렵다는 특징을 갖고 있다.

13번 군집을 예로써 살펴보면, ‘국제협력’, ‘ODA’ 등 방재 기술의 ‘글로벌 역량강화 및 국제공동연구’와 관련된 3과제 가운데 ‘보행안전기술’에 관한 과제가 군집을 이루고 있어 협업에 대한 직관적 판단을 내릴 수 없다.

반대로 협업 가능성에 대한 판단이 상대적으로 명확한 군집(잠재적 협업 가능 과제 군) 또한 확인되었다. 1번 군집에 포함되어있는 과제들의 키워드 및 내용분석을 통해 ‘예경보’, ‘재난상황관리’, ‘인공위성’, ‘지능형 CCTV’ 등의 특성을 확인하였다. 따라서 해당 군집은 “자료 구축을 통한 통합 모니터링 시스템 개발”로의 협업을 모색할 수 있다. 이와

유사한 협업의 가능성은 7번 군집에서도 확인되었다. 해당 군집의 경우 두 과제의 특성은 ‘인공위성’, ‘가뭄’, ‘원격탐사’, ‘재난정보’, ‘공간정보’ 등으로 확인되었으며 연구내용 요약 검토 결과 연구개발 추진 방향이 같은 맥락에 놓인 것으로 확인되었다. 나머지 과제의 키워드는 ‘신종복합재난’, ‘사회재난’, ‘피해영향예측’으로 구성되어 있으며 재난 위험성 분석기술 개발에 목적을 두고 있다. 이로부터 “위성 자료 분석을 통한 피해영향 예측”으로의 협업 가능성을 전망할 수 있으며, 해당 군집은 1번 군집과의 협업 가능성 또한 포함하고 있는 것으로 판단된다.

2번 군집의 경우 ‘재난원인과학조사’, ‘재난프로파일링’, ‘포렌식 조사’, ‘침단장비’ 등의 키워드가 3과제 간 일치하였다. 연구내용, 기대효과 등 해당 과제들의 속성을 비교한 결과 유사한 연구내용으로 인해 유사도가 높게 측정되어 하나의 군집을 이룬 것으로 확인되었다. 해당 군집의 경우 과제 간 유사증복성에 대한 검토, 더 나아가 향후 추진과제의 차별성 확보를 위한 검토 자료로 활용 가능성을 기대할 수 있다.

Table 3. Characterization of Each Clusters for Estimating the Potential Cooperative R&D Programs

Cluster No.	Projects	Characteristics	Cooperation Possibility
Cluster 1	p8, p9, p31	Forecasting-Warning Techniques, Smart Disaster Management System, SBB, Disaster Situation Management, AWS, Weather Information, Long-term Road-map, Satellite, UAV, Intelligent CCTV	○
Cluster 2	p13, p19, p27	Disaster Scientific Investigation, Disaster Profiling, Forensic Investigation of Disaster, State-of-the-art Technologies	○
Cluster 3	p1, p11, p20	Real-time Disaster Prediction, Tidal-Storm Surge-wave	○
Cluster 7	p7, p18, p28	Satellite, Drought, Remote Sensing, Disaster Information, GIS, Disaster	○
Cluster 8	p14, p29, p63	Foresight, Text-mining, Horizon Scanning, Risk Scanning, Damage Loss Function, Disaster Estimation	×
Cluster 9	p38, p43, p55	Disaster Information Management, Smart Big Board, Disaster Information, SBB, Smart Business Continuity, Decision Making, Evacuation, Relief, Disaster	○
Cluster 10	p37, p41, p44, p49	Foresight, Text-mining, Horizon Scanning, Risk Scanning, Personal Information Leak, Avian, Influenza, Indirect Damage, Disaster, Identification, ICT	×
Cluster 12	p50, p53, p81	Critical Infrastructure, Local Safety Features, Safety Index, Safety Indicator, Interdependency, Integrated Safety Information Management System, Disaster and Safety, Mobile Contents Application	×
Cluster 13	p36, p39, p51, p85	Walk Safety Technique, Typhoon, Disaster Information, ODA, International Cooperation, Disaster Prevention Technology Transfer, International Cooperation Research	×
Cluster 14	p45, p57, p58	Safety community, Safety Village, Pedestrian Safety, Community Mapping	○
Cluster 15	p6, p30, p65	Safety Support Technology, Electronic Manual, Crisis Management Manual, CBR, Virtual training Scenario	×
Cluster 16	p17, p24, p62, p70, p73, p74, p89	Foresight, Text-mining, Horizon Scanning, Risk Scanning, Social Big Data, Inflow Equation, Safety Diagnosis, Flood Risk Simulator, Risk Detection, Voluntary Program	×
Cluster 17	p67, p69, p71	Drought, NDVI, Satellite, Horizon Scanning System, Foresight, Emerging Issues	×
Cluster 18	p76, p78	Horizon Scanning System, Foresight, Emerging Issues, Rainfall Hazard Map	×
Cluster 19	p16, p82, p84, p88	Empirical Experiment, Safety Codes, SAR, Subsidence, Coastal Inundation Predicting, Tsunami	×
Cluster 20	p3, p86, p87, p90	Cause Analysis, Critical Infrastructure, Risk Assessment, Vulnerability Assessment	×

3번 군집은 ‘피해예측’이 3과제 간 공통의 키워드로 포함되어 있어 1번 군집과 유사하게 과제 간 협업의 가능성을 포함하고 있음을 확인할 수 있다. 9번 군집의 과제 간 키워드 및 연구내용을 비교·분석한 결과 유사한 두 과제와 나머지 과제로 구성되어 있음을 확인하였다. ‘재난정보관리’, ‘의사결정’, ‘대피’, ‘구호’ 등의 키워드, 각 과제의 연구내용을 통해 “재난정보를 활용한 대피, 구호 의사결정”으로의 협업 가능성을 검토할 수 있다.

4. 결 론

본 연구는 재난안전과 관련된 연구개발(R&D)의 효율성을 강화하고, 부처 간 협력적 재난안전 R&D를 독려하기 위해 편집거리 기반의 과제 유사도 측정 모델을 구축하였다. 구축된 모델을 통해 91개 재난안전 R&D 과제에 대한 유사도를 계산하여 지파이를 통한 군집화 및 데이터 시각화를 수행하였다. 정량화된 유사도 및 군집화 결과를 바탕으로 군집에 포함된 과제들의 정보를 비교·분석함으로써 유사·중복성 및 협업 가능성을 검토하였다.

편집거리 기반의 유사도 계산에 의한 과제 간 군집화 결과, 입력대상 91개의 과제가 21개의 군집으로 분류되었다. 분류된 군집 중, 정량적 유사기준 및 연차과제, 동일과제 등을 고려하여 “잠재적 협업 가능 과제 군”을 도출하였다. 해당 군집들의 과제별 키워드, 연구내용 등을 토대로 각 군집의 특성을 추출하였으며 이를 통해 ‘협업가능 과제 군’을 정리하였다.

21개의 군집 중, 연차과제 및 동일과제 간 군집으로 확인된 3개의 군집을 통해 신뢰도 높은 원자료(raw data)의 확보가 중요할 것으로 판단된다. 이와는 다른 측면에서 본 연구 결과는 원자료의 신뢰성을 검증할 수 있는 도구로서도 사용이 가능할 것으로 판단된다.

본 연구에서 제안된 방법론은 과제간 유사·중복성 탐지보다 향후 협업 가능한 과제들을 발굴하는 것을 목적으로 한다. 즉, 향후 추진 예정인 과제들을 DB화하여 본 방법론에 적용시킨다면 보다 향상된 결과를 도출할 수 있을 것이다. 다만, 모든 연구분야에서 제기되는 자료의 신뢰성에 관한 문제 해결이 선행되어야 할 것이며 향후 본 연구 결과 활용 시 자료의 충분한 검토 및 보완이 이루어져야 할 것이다.

본 연구에서 제시한 결과는 협업을 위한 절대적인 근거로 작용하기에는 한계점이 존재한다. 예를 들어 연구결과에서 제시된 과제 간 키워드의 넓은 범위는 실질적인 협업에 또다른 문제점으로 작용할 가능성이 있다. 향후 보다 좁은 범위의 키워드를 도출하고 이에 기반한 협업군을 구성한다면 협업의 가능성의 탐지 및 평가를 효율적으로 지원할 수 있을 것이다.

앞서 언급한 바와 같이 본 연구는 ‘협업가능한 재난안전 R&D’ 탐색에 그 목적을 두고 있다. 그러나 지속적인 연구를

통한 최종목표는 ‘협업방안 제시’에 있으므로 향후 ‘제3차 재난 및 안전관리 기술개발 종합계획’내 ‘범부처 협의체’ 운영 시 협력적 재난안전 R&D 추진을 위한 가이드라인으로서 활용이 가능할 것으로 판단된다. 또한 연구개발 추진 시 예산당국의 문제 제기에 따른 대응역량, 연구 기획력 강화에 효율성을 높일 수 있을 것이다.

또한 본 연구에서는 재난안전 R&D의 다양한 특성을 반영할 수 있도록 모델을 고안하였으므로 향후 입력값의 확장을 통해 다각적 분석을 수행하는데 기여할 수 있다. 이와 함께 ‘협업 가능 과제 군’ 도출의 신뢰성을 확보할 수 있도록 관련 전문가 검증이 수반되어야 할 것이다.

감사의 글

본 연구는 2017년 국립재난안전연구원 열린프로젝트에 의해 수행되었습니다.

References

- Bastian, M., Heymann, S., and Jacomy, M. (2009) Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the Third International ICWSM Conference*, pp. 361-362.
- CAO (2016) *What is the Cross-ministerial Strategic Innovation Promotion Program (SIP)*. Homepage of Cabinet Office of Japan.
- Choi, C.U., Kim, S.U., Kim, H.R., and Park, Y.T. (2003) On the Design of R&D Proposal Screening System. *Proceedings of the Technology Innovation Conference*, Korea Society for Innovation Management & Economics, pp. 3-11.
- Jung, O.N., Rhew, S.Y., and Kim, J.B. (2011) An Empirical Study on Improvement Model for Measuring of Project Similarity. *Journal of Digital Contents Society*, Vol. 12, pp. 457-465.
- Jung, S.W., and Kim M.D. (2016) A Study on the Possibility to Use Christopher Alexander’s Pattern Language by Using Network Analysis Tool. *Journal of the Korean Institute of Interior Design*, Vol. 25, No. 3, pp.31-39.
- Kim, J.B., Byun, J.W., Sun, D.J., Kim, T.G., and Kim, Y. (2014) A Model for Measuring the R&D Project Similarity using Patent Information. *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 18, No. 5, pp. 1013-1021.
- Kim, J.H., Kim, Y.J., and Kim, J.B. (2012) A Study on Similarity Analysis of National R&D Programs using

R&D Project's Technical Classification. *Journal of Digital Contents Society*, Vol. 13, pp. 317-324.

KISTEP (2013) *An Empirical Study on Pilot Scale System Development for Detection of Similar or Redundant R&D Program*. Korea Institute of S&T Evaluation and Planning.

Laurikari, V. (2001) *Efficient Submatch Addressing for Regular Expressions*. Master's thesis, Helsinki University of Technology.

Levenshtein, V.I. (1966) Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics Doklady*, Vol 10, No. 8, pp. 707-710.

MOIS (2017) The 3rd Comprehensive Plan for Technology Development of Disaster and Safety Management. *Proceeding on the Public Hearing of the 3rd Comprehensive Plan*, Seoul.

NSTC (National Science and Technology Council) (2005)

Grand Challenges for Disaster Reduction: A Report for the Subcommittee on Disaster Reduction. A Report for the subcommittee on Disaster Reduction.

Park, D.J., Choi, K.S., Lee, M.S., and Lee, S.T. (2009) A Study on the Development of Search Algorithm for Identifying the Similar and Redundant Research. *Journal of the Korea Contents Association*, Vol. 9, No. 11, pp. 54-62.

Vetere Arellano, A.L., Cruz, A.M., Nordvik, J.P., and Pisano, F. (2003) Analysis of Natech (Natural Hazard Triggering Technical Disasters) Disaster Management. *NEDIES Workshop Proceedings*, Italy.

Received	December 15, 2017
Revised	December 19, 2017
Accepted	January 4, 2018