빅 데이터 분석 기법을 이용한 풍수해 복원탄력성 지표 개발 및 평가: (1) 복원탄력성 지표 개발

Indicator Development and Evaluation of Storm and Flood Resilience Using Big Data Analysis: (1) Development of Resilience Indicators

김연수* · 최창현** · 배영혜*** · 김동현**** · 김덕환***** · 김형수*****

Kim, Yonsoo*, Choi, Changhyun**, Bae, Younghye***, Kim, Donghyun****, Kim, Deokhwan*****,
and Kim, Hung Soo*****

Abstract

In this study, we developed indicators for evaluation of storm and flood resilience using big data analysis. Standard terms were selected to collect data necessary for development of indicators and unstructured data such as papers and reports were collected. Preprocessing such as duplicate document processing, specialized dictionary registration, and stop word removal was performed on the collected data. Classification criteria were selected and indicators of the previous studies were reclassified by the classification criteria. Representative keywords were selected by each classified indicators and the related search words for each representative keywords were extracted by analysis of related search words. The extracted search terms were applied to the collected unstructured data, and 83 resilience indicators were developed by selecting detailed indicators and proxy variables in accordance with the selection criteria. The developed storm and flood resilience indicators can be used as a basic data of evaluation of storm and flood resilience for effective disaster management.

Key words : Big Data, Duplicate Document Processing, Related Search Words, Resilience Indicators, Storm and Flood

요 지

본 연구에서는 빅 데이터 분석 기법을 이용하여 풍수해 복원탄력성 평가에 필요한 지표를 개발하였다. 지표 개발에 필요한 데이터를 수집하기 위해 기준용어를 선정하였고, 논문과 보고서 등의 비정형 데이터를 수집하였다. 수집된 데이터에 중복 문서 처리, 전문용어 사전 등록, 불용어 처리의 전처리 과정을 진행하였고, 분류기준을 선정하여 선행연구들의 지표를 재분류하였다. 분류된 지표별로 대표 키워드를 선정하고, 연관 검색어 분석을 통해 대표 키워드별 연관 검색어를 추출하였다. 추출된 연관 검색어를 수집된 비정형 데이터에 적용하였고, 선정기준에 부합되는 상세지표와 대리변수를 선정하여 총 83개의 복원탄력성 지표를 개발하였다. 개발된 복원탄력성 지표는 효과적인 재난관리를 실시하기 위한 풍수해 복원탄력성 평가의 기초자료로 사용될 수 있을 것이다.

핵심용어 : 빅 데이터, 중복 문서 처리, 연관 검색어, 복원탄력성 지표, 풍수해

*정회원, ㈜LIG 시스템(E-mail: civil.engineer@hanmail.net)

Member, LIG system

**교신저자, 정회원, 인하대학교 토목공학과 박사수료(Tel: +82-32-874-0069, Fax: +82-32-876-9787, E-mail: karesma0cch@naver.com)

Corresponding Author, Member, Ph. D. Candidate, Department of Civil Engineering, Inha University

***정회원, 인하대학교 수자원시스템 연구소 연구원

Member, Researcher, Institute of Water Resources System, Inha University

****정회원, 인하대학교 토목공학과 석사과정

Member, Master's course, Department of Civil Engineering, Inha University

*****정회원, 인하대학교 수자원시스템 연구소 박사후연구원

Member, Post-Doctoral Researcher, Institute of Water Resources System, Inha University

*****정회원, 인하대학교 사회인프라공학과 교수

Member, Professor, Department of Civil Engineering, Inha University

1. 서론

전 세계적으로 재난으로 인한 심각한 경제적 손실과 인명 피해가 발생하고 있고, 기후변화의 영향에 따라 대규모 인명 및 재산피해를 유발하는 자연재난이 지속적으로 증가하고 있다(Choi et al., 2017; Lee et al., 2016). 특히 급격한 도시화로 인해 불투수층 면적과 자산가치가 증대되면서 재산피해가 급증하고 있는 추세이고, 도시의 발전과 더불어 사람들은 재난의 위험으로부터 안전하게 살고자 하는 욕구가 증가하고 있다(Kim et al., 2017). 이에 재난관리 기관들은 자연재난으로 인한 피해를 최소화하기 위한 노력을 지속적으로 기울이고 있으며, 기존의 피해저감 전략과 함께 자연재난 발생의 충격으로부터 신속하게 회복할 수 있도록 지역사회의 복원탄력성(Resilience)을 극대화 하는 전략을 제시하고 있다. 복원탄력성이란 ‘과거 발생한 재난규모에 대한 예방 및 대비가 가능하고, 재난 발생 시 이전 상태를 회복할 수 있는 대응 및 복구 능력’으로 정의할 수 있으며, 지역사회의 복원탄력성을 강화하기 위해선 지역별 복원탄력성 평가를 반영한 효율적인 재난관리가 필요하다(Kim, 2016). 적절한 복원탄력성 평가를 위한 세부적인 지표 개발이 선행되어야 하는데, 국외에서는 복원탄력성 지표 개발에 관한 다양한 연구가 진행되고 있다. De Bruijn (2004)은 Amplitude, Graduality, Recovery rate 측면에서 3가지(Physical, Economic, Social) 차원별로 복원탄력성 지표를 개발하였고, Cutter et al. (2008)은 지역이나 공동체 수준에서 복원탄력성을 평가할 수 있도록 6가지(Ecological, Social, Economic, Institutional, Infrastructure, Community) 차원별로 복원탄력성 지표를 개발하였다. Mayunga (2009)는 재난관리 단계별 복원탄력성을 평가하기 위해 CDRF (Community Disaster Resilience Framework)를 제시하였고, 4가지(Social, Economic, Physical, Human) 차원의 복원탄력성 지표를 개발하여 U.S. Gulf Coast 지역의 복원탄력성을 평가하였다. Razafindrabe et al. (2009)은 복원탄력성을 정량화 할 수 있도록 CDRI (Climate Disaster Resilience Index)를 제시하였고, 5가지(Physical, Social, Economic, Institutional, Natural) 차원별로 복원탄력성 지표를 개발하여 아시아 9개 해안도시의 복원탄력성을 평가하였다. Cutter et al. (2010)은 미국 남동부 지역의 복원탄력성을 평가하기 위해 5가지(Social, Economic, Institutional, Infrastructure, Community) 차원별로 복원탄력성 지표를 개발하였으며, 복원탄력성은 공간적인 변화가 존재하고 특히 대도시의 복원탄력성이 시골 지역보다 더 높은 수준을 가진다고 제시하였다. Parvin and Shaw (2011)도 CDRI를 사용하여 5가지(Physical, Social, Economic, Institutional, Natural) 차원별로 방글라데시 Dhaka의 복원탄력성을 평가하였고, Burton (2012)은 복원탄력성의 차원을 5가지(Social, Economic, Institutional, Infrastructure, Community Capital, Environmental Systems)로 구분하고 각 차원별 복원탄력성 지표를 개발하였다. Cutter

et al. (2016)은 대도시와 시골 지역의 복원탄력성을 구분하여 평가하기 위해 BRIC (Baseline Resilience Indicators for Communities)를 제시하였고, 6가지(Community, Economic, Institutional, Housing/infrastructural, Social, Environmental) 차원별로 복원탄력성 지표를 개발하였다.

복원탄력성 지표 개발에 관한 국내의 연구를 살펴보면, Korea Maritime Institute (2009)는 연안지역을 대상으로 풍수해의 충격을 최소화하고, 복구능력을 극대화할 수 있도록 복원탄력성 강화 방안을 제시하였다. 연안지역의 복원탄력성 평가를 위해 문헌조사와 전문가 설문조사를 실시하였고, 미국 해양대기청에서 개발한 CRI (Coastal Resiliency Index)를 국내 여건에 맞게 개선하여 8개(사회기반시설 및 공공시설 입지, 교통시설 복구 및 대피, 자연재해 대응계획 및 관리 역량, 자연재해 피해 저감 노력, 민간의 자연재해 대응 노력, 연안 시설물 피해 및 복구, 자연재해 대응 제도적 여건, 자연재해 대응 정책 분야별 상대중요도) 차원별로 복원탄력성 지표를 제시하였다. Kim et al. (2011)은 도시의 물리적·사회적 요소가 재해에 신속하게 적응하고 더 나은 상태로 회복하기 위해 복원탄력성 개념을 적용하였고, 5가지(내구성, 대체성, 자원확보, 신속성, 지역경쟁력)의 차원별 복원탄력성 지표를 제시하였다. Gyeonggi Research Institute (2013)는 국내·외 선행연구들을 참고하여 복원탄력성의 차원을 크게 5가지(사회적 차원, 경제적 차원, 제도적 차원, 인프라/물리적 차원, 환경시스템 회복력/생태적 차원)로 구분하였으며, 차원별 지표를 구축하여 사회적 비용관점에서의 정량적 평가방법을 제시하였다. Lee et al. (2013)은 4가지(대응·복구성, 물리적 차원, 사회적 차원, 경제적 차원) 차원별로 총 12개의 복원탄력성 지표를 제시하였고, Bak et al. (2016)은 국내·외 선행연구에서 도출한 지표의 내용을 종합하여 6가지(인문, 사회, 경제, 제도, 물리, 환경) 차원에서 인천지역의 환경과 지역적 특성을 반영하여 32개의 복원탄력성 지표를 제시하였다.

국내·외의 연구를 살펴보면 주로 복원탄력성 지표를 개발하기 위해 선행연구를 검토하여 기존 연구의 지표를 그대로 활용하거나, 연구자의 주관적 판단에 의해 지표 개발이 이루어져 왔다. 그러나 선행연구의 검토는 새로운 지표를 제시하기에 어려움이 있으며, 연구자의 주관적인 판단에 의한 지표 개발은 고려할 수 있는 지표의 범위가 제한적이다. 특히 최근 이슈가 되고 있는 빅 데이터(Big Data) 분석 기법을 적용하여 복원탄력성 지표를 개발한 사례는 미비한 상황이다. 따라서 본 연구에서는 새로운 지표의 제시가 가능하고, 다양한 분야의 의견을 반영할 수 있는 빅 데이터 분석 기법을 이용하여 풍수해 복원탄력성 지표를 개발하고자 하였다.

이를 위해 기준용어를 선정하여 논문 및 보고서 등의 비정형 데이터(Unstructured Data)를 수집하였고, 수집된 데이터에 중복 문서 처리, 전문용어 사전 등록, 불용어 처리 등의 전처리 과정을 수행하였다. 그리고 선행연구들을 검토하여 분류기준

을 선정하고, 이를 이용하여 선행연구들의 지표를 재분류하였다. 지표별 대표 키워드를 선정하였고, 선정된 대표 키워드를 이용하여 연관 검색어 분석을 실시하였다. 도출된 연관 검색어를 수집된 비정형 데이터에 적용하고, 상세지표 및 대리변수의 선정기준과 비교하여 풍수해 복원탄력성 지표로 활용 가능한 상세지표와 대리변수를 결정하였다. Table 1은 선행연구에서 제시한 복원탄력성 지표의 차원을 정리한 것이고, Fig. 1은 전체적인 연구과정을 도식화한 것이다.

2. 빅 데이터 분석 기법을 이용한 풍수해 복원탄력성 지표 개발 방법론

2.1 빅 데이터

빅 데이터(Big Data)란 무엇인가에 대하여 다양한 정의들

이 존재하고 있지만, 일반적으로 “데이터를 수집, 저장, 관리, 분석하는 기존 데이터베이스 처리방식을 뛰어넘는 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술”을 의미한다(Kim et al., 2016). 빅 데이터는 데이터의 정형화 정도에 따라 크게 정형데이터(Structured data)와 비정형 데이터(Unstructured data)로 구분 되는데, 정형 데이터는 규격화 되어 있는 고정된 필드에 분석이 용이하도록 체계적으로 정리된 데이터를 의미한다. 이러한 관점에서 본다면, 비정형 데이터는 아직 규격화되지 않았으므로 분석이 불가능하고 의미를 읽어낼 수 없는 데이터로 판단할 수 있다(Park and Song, 2014). 본 연구에서는 규격화되지 않은 비정형 데이터의 일종인 웹(Web)상의 전문자료를 이용하여 복원탄력성 지표를 개발하였다.

Table 1. The Dimension of Resilience Indicators Used in Previous Studies

Previous studies	Dimension of resilience indicators
De Bruijn (2004)	Physical, Economic, Social
Cutter et al. (2008)	Ecological, Social, Economic, Institutional, Infrastructure, Community
Mayunga (2009)	Social, Economic, Physical, Human
Razafindrabe et al. (2009)	Physical, Social, Economic, Institutional, Natural
Cutter et al. (2010)	Social, Economic, Institutional, Infrastructure, Community
Parvin and Shaw (2011)	Physical, Social, Economic, Institutional, Natural
Burton (2012)	Social, Economic, Institutional, Infrastructure, Community Capital, Environmental Systems
Cutter et al. (2016)	Community, Economic, Institutional, Housing/infrastructural, Social, Environmental
Korea Maritime Institute (2009)	Locations of infrastructure and public facilities, Restoration and evacuation of transportation facilities, Natural disaster response planning and management capacity, Efforts of natural disaster mitigation, Efforts of natural disaster response by private, Damage and restoration of offshore facilities, System conditions of natural disaster response, Relative importance by area of natural disaster response policy
Kim et al. (2011)	Durability, Substitutability, Resource acquisition, Quickness, Regional competitiveness
Gyeonggi Research Institute (2013)	Social, Economic, Institutional, Infrastructure, Environmental, Ecological
Lee et al. (2013)	Responsibility, Physical, Social, Economic
Bak et al. (2016)	Humanities, Social, Economic, Institutional, Physical, Environmental

Fig. 1. Flow Chart

2.2 중복 문서 처리

웹상의 수집 문서 중에서 중복되는 문서의 비율은 70~80%에 이른다고 알려져 있다(Manning et al., 2009). 중복되는 문서를 처리하기 위한 방안으로 바이트 비교, TF-IDF 코사인 거리, 단어 거리, 편집 거리, 싱글링(shingling) 등이 사용되는데, 본 연구에서는 정보검색 분야에서 주로 사용하는 싱글링 기법을 사용하였다(Broder et al., 1997; Fetterly et al., 2004; Kwon et al., 2007; Manning et al., 2009). 싱글링 기법에서는 웹 문서를 단어들의 집합으로 나타내며, 이를 이용하여 웹 문서의 중복 정도를 측정한다. Eq. (1)은 문서 A와 B에 대하여 싱글링 기법을 적용하여 중복되는 정도를 측정하는 것이다.

$$Y(A, B) = \frac{X(A) \cap X(B)}{X(A) \cup X(B)} \quad (1)$$

여기서, Y 은 0부터 1까지의 범위를 가지고, 1에 가까울수록 유사한 문서라는 것을 의미한다. 따라서 싱글링 기법을 이용하면 $Y(A, B) = 1$ 인 문서만을 판별하면 되기에 비교적 손쉽게 중복 문서에 대한 제거가 가능하다.

2.3 전문용어 사전 등록

수집된 비정형 데이터에 대하여 연관검색어를 추출하고, 단어의 빈도수를 산정하기 위해서는 사용할 단어들에 대한 기준이 필요하다. 일반적으로 국내에서는 비정형 데이터 안에서 단어의 분류를 국립국어원에서 제공하는 표준국어대사전의 표제어를 기준으로 하고 있다. 그러나 표준국어대사전에 등록되어 있지 않은 표제어는 단어별 빈도수 산정 결과에 영향을 미치게 된다. 예를 들어 수집된 비정형 데이터에서 “하천관리시설”라는 단어가 2번 언급이 되어 있고, 빈도수를 산정하고자 할 경우 수자원 및 재난 관련 용어가 등록 여부에 따라 결과가 다음과 같이 나타난다.

- ① 수자원 및 재난 관련 용어 미등록시: “하천” - 2건, “관리” - 2건, “시설” - 2건
- ② 수자원 및 재난 관련 용어 등록시: “하천관리시설” - 2건

위의 경우와 같이 관련 용어의 등록여부에 따라 “하천관리시설”이라는 용어의 빈도수가 누락이 되거나, 불필요한 용어의 빈도수가 나타나게 된다. 이러한 이유로 비정형 데이터로 일반적인 자료가 아닌 특정 분야의 전문자료를 활용할 경우, 관련 전문용어 사전에 대한 추가 등록이 필요하다(Kim, 2016).

2.4 불용어 처리

수집된 비정형 데이터 중에는 정보로서의 가치나 의미가

없는 단어들이 존재하는데, 이를 불용어(stop word)라고 한다. 불용어는 문서에서 출현 빈도가 높으나 단어 자체는 연구의 목적 및 방향과 맞지 않은 경우가 많고, 연관 검색어 분석 결과와 단어의 빈도수 산정 결과에 영향을 미친다(Kim, 2016). 따라서 불용어의 처리를 위하여 단어가 가지고 있는 동의어와 유의어 기준에 대한 검토가 필요하며, 표준국어대사전과 전문용어 사전을 통해 키워드나 연관검색어로 가치가 없는 단어를 제거해야 한다.

2.5 연관 검색어 분석

연관 검색어 분석이란 이용자의 검색 의도를 파악하여 적합한 관련 검색어를 제공함으로써 이용자의 의도에 맞는 자료들을 제공하는 정보 탐색 방법이다. 연관 검색어 서비스의 활용을 통해 이용자들은 특정 키워드 검색시 자신이 알아보려는 바에 대한 직접적인 정보뿐만 아니라 같은 키워드를 검색한 타인들의 관심사를 동시에 살펴봄으로써 보다 다각적인 차원에서 정보에 대한 접근이 가능하다(Bae, 2013). 예를 들어, ‘홍수량’이라는 단어를 웹상에서 검색하면 홍수, 최대홍수량, 홍수조절, 계획홍수량, 설계홍수량, 침투, 홍수위, 치수계획, 유역면적, 저수용량, 계획홍수위 등 다양한 연관 검색어를 제공해 준다(Kim, 2016). 즉, ‘홍수량’이란 단어 자체보다 이와 연관된 단어를 통해 중요한 정보를 획득할 수 있다. 연관 검색어 분석 결과는 적용 가능한 지표를 검토하고, 지표별 가용 가능한 대리변수의 선정하는데 활용할 수 있다.

2.6 용어 정의

기존의 선행연구에서는 복원탄력성 지표 개발에 있어 지수, 지표, 변수 등 다양한 용어를 특별한 정의나 구분 없이 사용하고 있다. 이에 본 논문에서는 각각의 용어를 다음과 같이 정의하였다.

- ① 차원(Dimensions): 동일한 성격을 가진 요소와 지표를 묶어주는 최상위 개념을 ‘차원’으로 정의하였다. 본 연구에서는 5개의 차원(물리적, 사회적, 경제적, 제도적, 자연적)을 사용하였다.
- ② 요소(Variable): 각 차원을 구성하는데 필요한 성분이나 조건에 해당하며, 중요한 의미를 가지는 부분을 ‘요소’로 정의하였다.
- ③ 개념지표(Concept indicator): 각 요소를 구성하는 개념적인 항목을 나타내는 지표를 ‘개념지표’로 정의하였다.
- ④ 상세지표(Detailed indicator): ‘개념지표’의 구체적인 내용이나 기준을 나타내는 지표를 ‘상세지표’로 정의하였다.
- ⑤ 대리변수(Proxy variable): ‘상세지표’를 측정 및 표현해 줄 수 있는 특정 값을 가지는 통계적 수치를 ‘대리변수’라고 정의하였다.

3. 비정형 데이터의 수집 및 전처리 과정

본 연구에서는 빅 데이터 분석 기법을 이용하여 풍수해 복원탄력성 산정에 필요한 지표를 개발하고자 하였다. 우선, 지표 개발에 필요한 자료를 수집하기 위해 기준용어를 선정하고, 전문자료 제공 사이트에서 재난과 관련된 학술논문 및 전문보고서 등의 비정형 데이터를 수집하였다. 비정형 데이터를 수집한 상태로 바로 사용할 수 없기 때문에, 중복 문서 처리, 전문용어 사전 등록, 불용어 처리 등의 비정형 데이터 전처리 과정을 수행하였다.

3.1 기준용어 선정

검색하는 용어의 선택에 따라 수집되는 비정형 데이터의 정보 및 범위가 달라지며, 이로 인하여 지표의 빈도수도 달라질 수 있다. 따라서 본 연구에서는 비교적 객관적이고, 널리 사용된다고 판단되는 풍수해 관련 법령 및 제도의 단어를 기준용어로 선정하였다.

- ① 재난 및 안전관리 기본법: ‘태풍’, ‘홍수’, ‘호우’, ‘대설’
- ② 자연재해대책법: ‘풍수해’, ‘태풍’, ‘홍수’, ‘호우’, ‘대설’, ‘수해’, ‘침수’
- ③ 재난관리 관련 용어: ‘예방’, ‘대비’, ‘대응’, ‘복구’

3.2 비정형 데이터 수집

본 연구에서는 전문자료 제공 웹사이트에서 제공하는 원문자료를 빅 데이터 분석 기법을 이용한 풍수해 복원탄력성 지표 개발에 기초가 되는 빅 데이터(비정형 데이터)로 정의하였다. 전문자료의 경우 원문자료 제공여부가 웹사이트별로 상이하기 때문에, 다양한 전문자료 제공 웹사이트에서 비정형 데이터를 수집할 필요가 있다. 예를 들어 ‘연속적인 극한호우사상의 발생을 가정한 거대홍수모의’ 학술지 논문은 학술연구 정보서비스(Research Information Service System, RISS)와 누리서비스(DataBase Periodical Information Academic, DBpia)에서는 원문자료를 제공하고 있지 않지만, 국가과학기술정보센터(National Digital Science Library, NDSL)와 과학기술학회마을(Science and Technology Society Village, STSV)에서는 원문자료를 제공하고 있다. 따라서,

Table 2에서와 같이 전문자료의 제공 및 활용도가 높은 4개의 웹사이트를 선정하였고, 3.1절에서 선정된 기준용어를 이용하여 각 웹사이트별 문서검색을 통해 학위 논문 및 학술지 논문, 특허, 보고서 및 단행본, 연구 및 기술 동향 등의 원문자료를 수집하였다.

전문자료 제공 웹사이트별 결과를 살펴보면 국가과학기술정보센터(NDSL)에서는 총 629건의 비정형 데이터를 수집하였는데, 다른 웹사이트에서는 제공하지 않고 있는 특허와 연구 및 기술 동향에 대한 정보를 제공하고 있다. 학술연구 정보서비스(RISS)는 주로 학위 논문, 학술지 논문, 보고서, 단행본 등의 정보를 제공해 주고 있으며, 총 3,023건의 비정형 데이터를 수집할 수 있었다. 과학기술학회마을(STSV)은 정보 제공 분야가 학위 논문 및 학술지 논문으로 제한되어 있으며, 총 288건의 비정형 데이터를 수집할 수 있었다. 누리서비스(DBpia) 또한 정보 제공 분야가 학위 논문 및 학술지 논문으로 제한되어 있으며, 기준용어에 대한 검색 결과로 수집한 비정형 데이터는 총 217건으로 가장 적은 정보량을 확보할 수 있었다. 결과적으로 전문자료의 제공 및 활용도가 높은 4개의 전문자료 제공 웹사이트에서 총 4,157건의 비정형 데이터를 확보하여 저자, 제목, 주제어, 초록에 대하여 정리하였다.

3.3 비정형 데이터 전처리

3.3.1 중복 문서 처리 기법 적용

각 웹사이트에서 수집한 비정형 데이터 중에서 저자와 제목이 동일한 문서가 존재한다. 앞에서 언급한 바와 같이 ‘연속적인 극한호우사상의 발생을 가정한 거대홍수모의’의 문서가 국가과학기술정보센터와 과학기술학회마을에서 중복되어 제공되고 있다. 이러한 동일저자의 동일문서를 처리하지 않으면, 복원탄력성 평가시 특정지표의 가중치가 높게 산정되기 때문에 제거할 필요가 있다. 본 연구에서는 중복 문서 처리 기법 중 대표적인 기법으로 알려진 샹글링(shingling) 기법을 적용하여, 중복되는 문서를 제거하였다. 중복 문서 제거 기준을 $r(A, B) = 1$ 로 하여 수집된 원문자료에 대한 검토를 수행하였고, 이를 통해 총 1,093건의 비정형 데이터를 구축하였다.

Table 2. Obtaining Unstructured Data by Professional Resources Web Site

Category	Professional resources web site				Sum
	NDSL	RISS	STSV	DBpia	
Thesis and journals	509	1,982	288	217	2,996
Patent	23	-	-	-	23
Reports and publications	68	1,041	-	-	1,109
Research and technology trends	29	-	-	-	29
Total	629	3,023	288	217	4,157

3.3.2 수자원 용어사전 등록

국내의 정보검색 분야에서는 단어별 빈도수를 산정하기 위해 국립국어원의 ‘표준국어대사전’의 단어를 활용하여 구축된 비정형 데이터의 단어를 분류하고 있다. 그러나 표준국어대사전의 경우 수자원 및 재난 관련 전문용어를 제공하고 있지 않기 때문에, 본 연구와 같이 특정 분야의 단어가 필요한 경우에는 적절한 단어별 빈도수 산정이 어렵게 된다. 따라서, 본 연구에서는 한국수자원공사와 한강홍수통제소 등에서 제공하고 있는 ‘수자원 용어사전’을 추가적으로 분석도구에 등록하여 적절하게 단어별 빈도를 산정하였다.

3.3.3 불용어 처리

빅 데이터 분석 기법을 적용하여 유의미한 결과를 도출하기 위해서는 수집된 비정형 데이터를 검토하여 불용어를 제거해야 한다. 불용어 처리를 위한 기준으로 표준국어대사전과 수자원 용어사전을 이용하였다. 학술지 논문과 보고서의 경우 공통적으로 ‘본 논문에서는 ~’, ‘본 연구에서는 ~’ 등과 같이 의미 없이 반복되는 문장을 제거하였고, 숫자와 기호의 경우 의미 없이 연속적으로 부여된 숫자와 ‘+’, ‘-’, ‘/’, ‘.’ 등의 기호를 제거하였다. 상기 방법에 따라 불용어 처리를 진행하여 비정형 데이터의 전처리 과정을 완료하였다.

4. 빅 데이터 분석 기법을 이용한 풍수해 복원탄력성 지표 개발

선행연구 검토를 통해 풍수해 복원탄력성 지표 개발에 적합한 분류기준을 선정하였고, 이를 통해 선행연구들의 복원탄력성 지표들을 분류하였다. 분류된 지표별로 대표 키워드를 선정하였고, 연관 검색어 분석을 통해 대표 키워드별 연관 검색어를 추출하였다. 추출된 대표 키워드별 연관 검색어를 3장에서 수집 및 전처리 과정을 마친 비정형 데이터에 적용하여 상세지표를 산정하였다. 그리고 상세지표 및 대리변수 선정 기준의 검토를 통해 풍수해 복원탄력성 지표로 활용 가능한 상세지표와 대리변수를 선정하였다.

4.1 선행연구 검토를 통한 지표 분류

빅 데이터 분석 기법을 이용하여 복원탄력성 지표를 개발하기 위해, 우선적으로 선행연구를 바탕으로 복원탄력성 지표 개발에 필요한 차원 및 요소를 검토하였다(Kim, 2000; National Institute for Disaster Prevention, 2000; De Bruijn, 2004; Cutter et al., 2008; Mayunga, 2009; Razafindrabe et al., 2009; Kim, 2010; National Disaster Management Institute, 2010; Jeong, 2010; UNISDR, 2010; Cutter et al., 2010; Esnard et al., 2011; Burton, 2012; Beck et al., 2012; Ko, 2012; Park, 2012; Sung et al., 2012; Gyeonggi Research Institute, 2013; Lee et al., 2013; Park et al., 2014; Lee, 2014; Kang, 2015; Han et al., 2015).

선행연구를 검토한 결과, Razafindrabe et al. (2009)과 UNISDR (2010)에서 제시하고 있는 CDRI (Climate Disaster Resilience Index)가 본 연구에서 개발하고자 하는 풍수해에 대한 복원탄력성 지표의 성격 및 구성형태와 가장 유사하기 때문에 CDRI의 구성요소(5개 차원, 23개 요소)를 기준으로, 다른 선행연구들의 결과를 정리 및 보완하여 5개 차원, 27개 요소, 98개의 개념지표로 분류하였다. 분류기준으로 사용한 CDRI의 차원과 요소는 Table 3과 같다.

4.2 지표별 대표 키워드 선정

각 요소를 구성하는 개념적인 항목을 나타내는 개념지표의 구체적인 목적이나 기준으로 나타내는 지표를 상세지표라고 하는데, 이를 측정 및 표현해 줄 수 있는 통계적 수치를 대리변수라고 한다. 선행연구 분석을 통해 도출된 개념지표별로 대표 키워드를 선정하기 위해 우선, 선행연구의 지표별 대리변수를 수집하였다. 각 지표별로 분류된 대리변수 중에는 연관 검색어를 추출하는데 적합하지 않은 추상적인 단어들이 있기 때문에, 선행연구의 대리변수가 가지고 있는 내용을 포함하면서도 검색이 가능한 단어로 수정하는 과정을 거쳐서 대표 키워드를 선정하였다.

4.3 연관 검색어 분석

국내에서 연관 검색어 서비스를 제공하고 있는 포털사이트는 구글(<https://www.google.co.kr>), 네이버(<https://www.naver.com>),

Table 3. List of Dimensions and Variables in CDRI (Razafindrabe et al. (2009))

Dimensions	Variable
Physical	Electricity, Water supply, Sanitation, Solid waste disposal, Internal road network, Housing and land use, Community assets, Warning system and evacuation
Social	Health status, Education and awareness, Social capital
Economic	Income, Employment, Households' assets, Access to financial service, Savings and insurance, Budget and subsidy
Institutional	Internal institutions and development plan, Effectiveness of internal institutions, External institutions and networks, Institutional collaboration and coordination
Natural	Hazard intensity, Hazard frequency

다음(<https://www.daum.net>), 네이트(<http://www.nate.com>), 줌(<http://zum.com>) 등이 있다. 최근 7년간(2011년 1월 1일~2017년 12월 31일) 국내의 검색엔진별 사용률을 검토해본 결과, 네이버 78.08%, 다음 15.59%, 구글 3.50%, 네이트 0.90%, 줌 0.65%인 것으로 나타났다(InternetTrend, 2018). 네이버와 다음의 검색엔진 사용률 합은 약 94%로 국내에서 인터넷을 이용하는 사람들의 대부분은 두 검색엔진을 사용한 것을 알 수 있었다. 따라서 4.1.2절에서 선정한 대표 키워드의 연관 검색어를 분석하기 위해 네이버와 다음의 검색엔진을 이용하였다.

4.4 상세지표 및 대리변수 선정

4.4.1 상세지표의 선정 기준

단순히 출현빈도 수가 높은 단어나 개념지표에 부합되지 않은 단어를 무작정 복원탄력성 지표로 선정한다면, 최종적으로 복원탄력성을 평가할 때 문제가 발생할 수 있다. 따라서 4.3절에서 선정된 연관 검색어를 이용하여, 개념지표를 좀 더 구체적으로 표현해 줄 수 있는 상세지표로 제시해 줄 필요가 있다. 개념지표별 대표 키워드의 연관검색어 중에서 다음과 같은 기준을 적용하여 상세지표를 선정하였다.

- ① 대표성(Typicality): 대표 키워드와 연관 검색어 분석 결과를 대표할 수 있어야 한다.
- ② 연관성(Relevance): 요소에 영향을 미치는 지표여야 한다.
- ③ 구체성(Specific): 개념지표의 의미를 비교적 명확하게 구체화 할 수 있어야 한다.
- ④ 이해가능성(Understandability): 쉽게 이해할 수 있도록 명확하고 간단해야 한다.

4.4.2 대리변수의 선정 기준

4.4.1절의 상세지표가 각각의 요소를 구성하는 개념적인 항목인 개념지표를 구체화한 지표였다면, 대리변수는 상세지표를 대신하여 통계적 수치로 표현해 준 것이다. 만약 선정된 상세지표가 대리변수로 활용할 수 있다면 문제가 없지만, 그렇지 않고 직접적인 대리변수로 활용이 어렵다면 상세지표를 참고하여 대체 가능한 대리변수를 선정해야 한다. 따라서 다음과 같은 기준을 통해 대리변수를 선정하였다.

- ① 대표성(Typicality): 상세지표를 대표할 수 있어야 한다.
- ② 가용성(Availability): 활용하기 위한 통계적 자료가 확보되어야 한다.
- ③ 지속성(Durability): 자료가 일정 시간 간격을 두고 지속적으로 갱신되어야 한다.
- ④ 방향성(Directivity): 복원탄력성 평가에 긍정적 혹은 부정적 영향의 일관성을 유지해야 한다.

- ⑤ 비교가능성(Comparability): 시간과 공간 분포의 변화에 따라 비교가 가능해야 한다.

4.5 풍수해 복원탄력성 지표 결정

4.3절에서 선정된 연관 검색어의 대표성, 타당성, 구체성, 이해가능성을 검토하여 상세지표를 선정하였고, 선정된 상세지표의 대표성, 가용성, 지속성, 방향성, 비교가능성을 검토하여 대리변수를 선정하였다. 이를 차원별로 살펴보면 물리적 차원은 ‘전력’, ‘수자원’, ‘위생(하수)시설과 쓰레기 처리’, ‘도로의 접근성’, ‘주택과 토지이용’, ‘가스’, ‘내외수 방재시설’ 등의 7가지 요소로 구성되었다. 사회적 차원은 5개의 요소와 19개의 상세지표로 구성되어 있는데, 5개의 요소는 ‘인구’, ‘건강’, ‘교육 및 인식’, ‘사회적 자본’, ‘재난 중 지역사회 대비책’으로 구분된다. 경제적 차원은 5개의 요소와 13개의 상세지표로 구성되어 있으며, ‘소득’, ‘고용’, ‘가계자산’, ‘금융과 저축’, ‘예산과 보조금’의 요소 구분된다. 제도적 차원은 5개의 요소와 15개의 상세지표로 구성되어 있으며, ‘재해 위험 감소와 기후변화 적응의 주류화’, ‘도시의 위기관리 프레임워크’, ‘재난에 대응하는 도시기관’, ‘다른 조직과 이해당사자와의 협력’, ‘적절한 관리’로 구분된다. 자연적 차원은 5개의 요소와 11개의 상세지표로 구성되어 있으며, 5개의 요소는 ‘자연재난의 강도’, ‘자연재난의 빈도’, ‘생태계 서비스’, ‘자연조건의 토지이용’, ‘환경정책과 식량안보’로 구분된다. 결론적으로 선행연구 분석을 통해 도출된 5개 차원, 27개 요소, 98개의 개념지표를 빅 데이터 분석 기법을 이용하여 풍수해 복원탄력성 지표로 5개의 차원, 27개의 요소, 83개의 상세지표로 제시하였다. Table 4는 본 연구에서 개발된 빅 데이터 분석 기법을 이용한 풍수해 복원탄력성 지표를 차원 및 요소별 상세지표와 대리변수로 구분하여 나타낸 것이다.

5. 결 론

본 연구에서는 기존에 고려되지 않았던 전문자료 등의 비정형 데이터와 빅 데이터 분석 기법을 이용하여 효과적인 재난관리를 위한 풍수해 복원탄력성 지표를 개발하고자 하였다. 연구 결과를 요약한 내용은 다음과 같다.

- (1) 재난/재해 관련 법령 및 제도에서 제시하는 풍수해 관련 용어를 기준용어로 선정하였고, 전문자료 제공 웹사이트에서 기준용어를 검색하여 학술지 논문과 보고서 등의 전문자료 4,157건을 비정형 데이터로 수집하였다.
- (2) 수집된 전문자료에 중복 문서 처리 기법인 쉐들링(shingling)기법을 적용하여, 중복되는 문서를 제거하였다. 중복 문서 제거 기준을 $r(A, B) = 1$ 로 하여 수집된 원문자료에 대한 검토를 수행하고, 이를 통해 총 1,093건의 비정형 데이터를 구축하였다.

Table 4. List of storm and flood resilience indicators

Dimensions	Physical												
Variable	Electricity			Water			Sanitation and solid waste disposal				Accessibility of roads		
Concept indicator	Access, Availability	Supply, Usage	Alternative capacity	Access, Availability	Supply	Alternative capacity	Access to sanitation	Toilets, Collection of wastes	Waste treatment	Recycling	Transportation network	Paved roads	Ditch
Detailed indicator	Electrical grid	Electricity, Usage	Renewable energy	Water pipe	Water use	Underground water	Waste pipe	Public sanitary management	Garbage disposal	Recycling	Bus	Network of road	Ditch
Proxy variable	Length of electrical grid	Amount of electricity used	Renewable energy production	Extension of water pipe	Amount of water supply	Amount of groundwater used	Extension of waste pipe	Number of public sanitary management facilities	Garbage emission	Amount of recycling	Number of bus	Extension of paved road	Length of open channel and ditch

Dimensions	Physical												
Variable	Housing and landuse						Gas			Internal and external disaster prevention facilities			
Concept indicator	Residential facilities	Non-permanent building	Housing type	Ownership of house	Year of construction	Steep slope-land	Access, Availability	Supply, Usage	Alternative capacity	Erosion control facilities	Pump facility	Waterproofing facility	
Detailed indicator	Housing	Warehouse	Lowland, Underground, Inundation	Ownership of house	Year of construction	Steep slope-land	Gas facilities	Gas supply	Portable gas	Erosion control facilities	Drainage pumping station, Rain pumping station	Drainage facility	
Proxy variable	Housing distribution ratio	Area of warehouse facilities	Percentage of houses without flood damage	Number of self-owned houses	Number of buildings constructed before 1979 year	Number of steep slope-land	Area of gas facilities	City gas distribution rate	Number of LPG and butane gas	Number of erosion control dam	Capacity of drainage pump	Number of drainage and waterproofing facilities	

Dimensions	Social															
Variable	Population				Health			Education and awareness			Social Capital		Community preparedness during a disaster			
Concept indicator	Population growth rate	Population under 6 and above 65	Population of foreigners	Population density	Gender (male, female)	Access to health facilities	Functional capacity of health facilities	Healthcare and disease control for disaster	Preschooler rate under 6	Internet accessibility	Functionality of schools for disasters	Community and group activity	Community competences	Mixing and interlinking of social classes	Logistics	Support and relief work from NGO
Detailed indicator	Life expectancy	Population by age groups	Population of foreigners	Population density	Gender, Male, Female	Health Insurance	Number of doctors	Prevention, Health behavior	Education status	Access to the internet, Smart phone	School, Education facilities	Community, Community involvement	Voting rate	Social welfare facilities	Mart, Market	NGO, Disaster relief work
Proxy variable	Population growth rate	Population under 6 and above 65	Population of foreigners	Population density	Population of man and woman	Population of health insurance	Number of doctors	Hospital bed retention rate	Number of health education population	Population of preschooler	Current status of smartphone ownership	Number of education facilities	Welfare facilities and public physical facilities	Voting rate	Number of social welfare facilities	Distributor

Dimensions	Economic												
Variable	Income			Employment			Household assets		Finance and savings		Budget and subsidy		
Concept indicator	Population below poverty line	Number of employed	Income level	Businesses	Non-economic participants	Female employment rate	Commuters in other regions	Ownership of phone	Vehicle ownership	Individual finance	Flood and storm insurance	Compensation and national subsidy	Financial independence rate
Detailed indicator	Basic livelihood act recipients	Labor force	Income level, National income	Businesses	Unemployment rate	Female employment	Commuters in other regions	Phone	Private vehicle	Saving, Deposit	Flood and storm insurance	National subsidy	Financial independence rate
Proxy variable	National basic livelihood act recipients	Working force	Local tax burden	Number of businesses	Non-economic participants	Number of female employee	Number of commuters in other regions	Number of telephone facility subscribers	Number of motor registration	Number of financial institution	Flood and storm insured household	Reserve fund	Financial independence rate

Dimensions	Institutional														
Variable	Disaster Risk Reduction and Climate Change Adaptation		City's crisis management framework		City's institutions to respond to a disaster			Institutional collaboration with other organizations and stakeholders					Good governance		
Concept indicator	Land-use and urban plans	Environmental plans	Emergency aid	Human resources	Human resources of emergency rescue	Recovery equipment	Dangerous facilities	Government office and major organizations in jurisdiction	Voluntary organization	NGOs	Religious organization	Press	Community assessment system	Local government integrity	Effectiveness of early warning systems
Detailed indicator	Urban plans	Antipollution facilities	Emergency aid	Public officer	Fire officer	Fire equipment, Recovery equipment	Hazardous material storage	Related organization	Voluntary organization	NGOs	Religious organization	Press	Local safety	Public institution integrity	Flood forecasting, Alarm system
Proxy variable	Area of district unit planning	Number of environmental infrastructure	Number of transportation and rescue	Number of public officer	Number of fire officer	Number of fire equipment and construction equipment	Installation status of dangerous material manufactory	Number of government office and major organizations in jurisdiction	Voluntary registration group	Number of NGOs	Active populations of religious organization	Number of press	Results of local safety assessment	Public institution integrity	Weather observation, Flood forecasting and alarm point

Dimensions	Natural											
Variable	Intensity/severity of natural hazards	Frequency of natural hazards	Ecosystem services			Land-use in natural terms			Environmental policies and food security			
Concept indicator	Volume of extreme event	Date of extreme event	Conservation District	Air pollution	Water pollution	Disaster Risk Zone	Urbanization rate	Greens of park	Environmental pollution control and measures	Environmental Performance	Food supply during disasters	
Detailed indicator	Rainfall	Rainfall	Biodiversity, Habitat protection	Air pollution	Water pollution	Disaster Risk Zone	Urbanization rate	Greens of park	Environmental pollution control	Environmental management, Environmental inspection	Food supply	
Proxy variable	RX1day, RX5day, SDII, R95p, R99p, PRCPTOT	R10, R20, R25, CWD	Number of ecosystem and watershed	Air pollution ratio	Water pollution ratio	Number of disaster risk zone	Urbanization rate	Number of park	Number of crackdown on businesses environmental pollutant emissions	Number of environmental inspection	Number of supermarket	

- (3) 구축된 비정형 데이터의 불용어 처리 및 연관 검색어 분석을 위해 전문용어 사전으로 표준국어대사전과 수자원 용어사전을 등록하였고, 불용어 처리 과정을 거쳐 비정형 데이터의 전처리 과정을 완료하였다.
- (4) 국내·외 23개의 복원탄력성 관련 선행연구를 검토하여 분류기준을 선정하였고, 이를 반영하여 선행연구의 지표들을 재분류하였다. 재분류된 개념지표별로 대표 키워드를 선정하였고, 국내 검색엔진 사용률의 대부분을 차지하는 네이버와 다음 검색엔진을 이용하여 대표 키워드별 연관 검색어를 추출하였다.
- (5) 추출된 개념지표의 연관 검색어를 전처리 과정을 거친 비정형 데이터에 적용하여 대표성, 연관성, 구체성, 이해가능성 검토를 통해 총 5개 차원, 27개 요소의 83개 상세지표를 선정하였고, 대표성, 가용성, 지속성, 방향성, 비교가능성을 검토하여 상세지표를 정량적으로 표현해 줄 수 있는 대리변수를 선정하였다.

본 연구에는 여러 가지 제약조건들이 있었는데, 한계점 및 보완점을 정리하면 아래와 같다.

- (1) 비정형 데이터를 정리하는데 있어 중복 문서 처리와 동일 저자들이 비슷한 제목으로 제출한 유사 문서 처리를 고려해야 하는데, 유사 문서 처리방법은 연구자의 주관적인 판단이 상당 부분 반영되기 때문에 본 연구에서는 중복 문서만을 제거하였다.
- (2) 적절하게 불용어를 처리하는 것은 검토해야 하는 비정형 데이터의 양을 축소시켜 결과를 도출하는데 필요한 시간을 감소시키는 장점이 있지만, 적절하지 않게 제거되는 경우 결과의 신뢰성이 감소되는 단점이 있다. 그러나 불용어 처리를 위한 다양한 방법들이 제시되고 있지만, 아직까지 어느 정도 수준으로 검토 및 제거해야 한다는 명확한 기준은 없기 때문에 보다 객관적이고 타당한 비정형 데이터 처리방법에 대한 연구가 추가적으로 필요하다.
- (3) 국내 검색엔진 사용률을 살펴보면 대부분의 경우 네이버와 다음의 검색엔진을 사용하지만, 전문자료의 경우 구글 검색을 통해서 자료를 많이 얻을 수 있다. 따라서 구글을 연관검색어 사이트에서 제외한 것은 본 연구의 한계점으로 볼 수 있으며, 특히 한국어 단어로만 검색했기 때문에 해외의 비정형 데이터를 수집하는데 한계가 있었다.
- (4) 전문용어 사전을 수자원 용어사전만 사용하였는데, 최근 방재학회에서 출간한 재난안전용어사전을 추가 고려한다면 조금 더 재난분야에 적합한 용어들이 선정될 수 있을 것으로 기대된다.
- (5) 선정된 상세지표들 사이에는 유사한 특성을 지닌 지표들이 존재하는데, 이러한 높은 상관성을 가진 지표들

을 통합 및 제거할 필요가 있다. 따라서 적정 지표를 선정하기 위해선 요인 분석과 신뢰도 분석 등이 추가적으로 필요할 것으로 보인다.

본 연구에서는 재난관리 분야에서 다루어지지 않았던 빅 데이터(Big Data) 분석 기법을 이용하여 풍수해 복원탄력성 지표를 개발하였다. 빅 데이터 분석기법을 통해 기존의 복원탄력성 지표 개발 방법의 고질적인 문제로 지적되었던 새로운 지표를 제시하기에 어려운 문제와 다양한 차원의 의견을 반영하지 못하여 개발 가능한 지표의 범위가 제한적인 문제를 일정 부분 극복할 수 있었다. 특히 기존의 복원탄력성 지표 개발 연구에서는 고려되지 않았던 비정형 데이터와 빅 데이터 분석 기법의 적용 가능성을 제시하였다는 점에서 큰 의미가 있다. 또한 풍수해 복원탄력성을 단순히 물리적 차원만을 고려하여 평가하는 기존의 1차원적인 접근 방식(One-dimensional approach)에서 벗어나서, 사회적, 경제적, 제도적, 자연적 차원에 영향을 미치는 다양한 요소들을 빅 데이터 분석기법을 통해 검토하는 복합적 접근 방식(Multi-dimensional approach)을 시도하였다. 본 연구에서 개발된 지표는 보다 객관적이고, 다양한 차원의 의견을 반영하여 기후재난의 복원탄력성을 평가하는데 기초자료로 사용될 수 있을 것이다.

감사의 글

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음(과제번호 18AWMP-B083066-05).

References

- Bae, Y. (2013) Additional Policy Decision Reviews for Related and Autocomplete Queries. *Journal of Korea Internet Self-governance Organization*, Vol. 10, pp. 1-4.
- Bak J.U., Kim D.H., and Kim H.S. (2016) Assessment of Resilience in Incheon City: (1) Estimation Model Development and Assessment of Resilience. *J. Korean Soc. Hazard Mitig.*, Vol. 16, No. 6, pp. 109-118.
- Beck, M., Shepard, C., Birkmann, J., Rhyner, J., Welle, T., Witting, M., Wolfertz, J., Martens, J., et al. (2012) *World Risk Report 2012*. Alliance Development Works, Berlin.
- Broder, A.Z., Glassman, S.C., Manasse, M.S., and Zweig, G. (1997) Syntactic Clustering of the Web. *Computer Networks and ISDN Systems*, Vol. 29, No. 8-13, pp. 1157-1166.

- Burton, C.G. (2012) *The Development of Metrics for Community Resilience to Natural Disasters*. Ph.D. dissertation, University of South Carolina.
- Choi, C.H., Kim, J.S., Kim, J.H., Kim, H.Y., Lee, W.J., and Kim, H.S. (2017) Development of Heavy Rain Damage Prediction Function Using Statistical Methodology. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 3, pp. 331-338.
- Cutter, S.L., Ash, K.D., and Emrich, C.T. (2016) Urban-Rural Differences in Disaster Resilience. *Annals of the American Association of Geographers*, Vol. 106, No. 6, pp. 1236-1252.
- Cutter, S.L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., and Webb, J. (2008) A Place-based Model for Understanding Community Resilience to Natural Disasters. *Global Environmental Change*, Vol. 18, No. 4, pp. 598-606.
- Cutter, S.L., Burton, C.G., and Emrich, C.T. (2010) Disaster Resilience Indicators for Benchmarking Baseline Conditions. *Journal of Homeland Security and Emergency Management*, Vol. 7, No. 1, Article No. 51, pp. 1-22.
- De Bruijn, K.M. (2004) Resilience Indicators for Flood Risk Management Systems of Lowland Rivers. *International Journal of River Basin Management*, Vol. 2, No. 3, pp. 199-210.
- Esnard, A., Sapat, A., and Mitsova, D. (2011) An Index of Relative Displacement Risk to Hurricanes. *Natural Hazards*, Vol. 59, No. 2, pp. 833-859.
- Fetterly, D., Manasse, M., Najork, M., and Wiener, J.L. (2004) A Large-scale Study of the Evolution of Web Pages. *Software: Practice and Experience*, Vol. 34, No. 2, pp. 213-237.
- Gyeonggi Research Institute (2013) *A Policy Implication for Community Resilience from Natural Disasters*.
- Han, S.R., Kang, N.R., and Lee, C.S. (2015) Disaster Risk Evaluation for Urban Areas Under Composite Hazard Factors. *J. Korean Soc. Hazard Mitig.*, Vol. 15, No. 3, pp. 33-43.
- InternetTrend (2018) *Logger Trend Data* (February 20, 2018). Retrieved from <http://www.internettrend.co.kr>
- Jeong, S.H. (2010) *Analysis on Determinants of Efficiency in Disaster Management System in Regional Governments*. Ph.D. dissertation, Kyungnam University.
- Kang, H.W. (2015) *A Study on the Urban Flood Vulnerability Assessment for Urban Area Considering Climate Change*. Master's thesis, University of Seoul.
- Kim, J.D. (2010) An Analysis on the Order of Priority in Disaster Management Policy. *J. Korean Soc. Hazard Mitig.*, Vol. 10, No. 2, pp. 61-68.
- Kim, J.S., Choi, C.H., Lee, J.S., and Kim, H.S. (2017) Damage Prediction Using Heavy Rain Risk Assessment: (1) Estimation of Heavy Rain Damage Risk Index. *J. Korean Soc. Hazard Mitig.*, Vol. 17, No. 2, 361-370.
- Kim, T.H., Kim H.J., and Lee K.J. (2011) The Concept and Functional Objectives of the Urban Resilience for Disaster Management. *Journal of the Korean Society of Safety*, Vol. 26, No. 1, pp. 65-70.
- Kim, T.Y. (2000) *National Disaster Management System Establishment Plan*. Korea Institute of Public Administration.
- Kim, Y.S. (2016) *Development of Resilience Indicator Based on Big Data Analysis Under Climate Change*. Ph.D. dissertation, Inha University.
- Kim, Y.S., Kang, N.R., Jung, J.W., and Kim, H.S. (2016) A Review on the Management of Water Resources Information based on Big Data and Cloud Computing. *Journal of Wetlands Research*, Vol. 18, No. 1, pp. 100-112.
- Ko, C.S. (2012) *A Study for the Improvement of Disaster Management Systems in Korea*. Ph.D. dissertation, Kyunghee University.
- Korea Maritime Institute (2009) *Measures to Strengthen the Resilience of Coastal Areas in Response to Climate Change*.
- Kwon, S.Y., Kim, S.J., and Lee, S.H. (2007) An Effective Metric for Measuring the Degree of Web Page Changes. *Journal of Korea Information Science Society: Databases*, Vol. 34, No. 5, pp. 437-447.
- Lee, J.S., Eo, G., Choi, C.H., Jung, J.W., and Kim, H.S. (2016) Development of Rainfall-Flood Damage Estimation Function using Nonlinear Regression Equation. *Journal of The Korean Society of Disaster Information*, Vol. 12, No. 1, pp. 74-88.
- Lee, J.Y. (2014) *A Study on the Regional Differences in Flood Damage Restoration Project*. Master's thesis, Chung-ang University.
- Lee, L.Y., Kim S.J., Kim Y.S., and Kim H.S. (2013) Evaluation of Resilience in Anyang River Basin for Flood Damage Mitigation. *J. Korean Soc. Hazard Mitig.*, Vol. 13, No.2, pp. 291-298.
- Manning, C.D., Raghavan, P., and Schütze, H. (2009) *An Introduction to Information Retrieval*. Cambridge University Press.

- Mayunga, J.S. (2009) *Measuring the Measure: A Multi-dimensional Scale Model to Measure Community Disaster Resilience in the U.S. Gulf Coast Region*. Ph.D. dissertation, Texas A&M University.
- National Disaster Management Institute (2010) *Development of the Indicators and Checklists for Diagnosis of the Urban Resilience*.
- National Institute for Disaster Prevention (2000) *A Study of Mitigation Efforts for the Repetitive Flooded Areas*.
- Park, D.H., and Song, D.H. (2014) *Political, Economic and Cultural Implications of Atypical Data Activation*. Internet & Security Focus: February 2014, Focus 1, pp. 4-20.
- Park, J.H. (2012) *A Study on City Risk Assessment Method Considering the Typhoon*. Master's thesis, University of Seoul.
- Park, Y.K., Jeong, S.M., and Kim, S.D. (2014) Natural Disaster Vulnerability Assessment at Boroughs and Census Output Areas in Seoul Focusing on Socio-economic Perspective. *J. Korean Soc. Hazard Mitig.*, Vol. 14, No. 6, pp. 439-449.
- Parvin, G.A., and Shaw, R. (2011) Climate Disaster Resilience of Dhaka City Corporation: An Empirical Assessment at Zone Level. *Risk, Hazards & Crisis in Public Policy*, Vol. 2, No. 2, pp. 1-30.
- Razafindrabe, B.H., Parvin, G.A., Surjan, A., Takeuchi, Y., and Shaw, R. (2009) Climate Disaster Resilience: Focus on Coastal Urban Cities in Asia. *Asian Journal of Environment and Disaster Management*, Vol. 1, No. 1, pp. 101-116.
- Sung, J.H., Baek, H.J., Kang, H.S. and Kim, Y.O. (2012) The Assessment of Future Flood Vulnerability for Seoul Region. *Journal of Wetlands Research*, Vol. 14, No. 3, pp. 341-352.
- UNISDR(United Nations International Strategy for Disaster Reduction) (2010) *Climate and Disaster Resilience Initiative Capacity-building Program*.

Received	■	April 4, 2018
Revised	■	April 9, 2018
Accepted	■	April 18, 2018