

네트워크 중심성을 이용한 건설 사업 리스크 관련 키워드 동향 분석

정성윤*, 김남곤
한국건설기술연구원 미래스마트건설연구본부

Analysis of keyword trends related to construction project risk using network centrality

Seong-Yun Jeong*, Nam-Gon Kim
Department of Future & Smart Construction Research,
The Korea Institute of Civil Engineering and Building Technology

요약 건설 사업에서의 리스크는 사업의 성공을 좌지우지할 만큼 중요한 사업 관리 요소이다. 국내 여러 연구자들은 건설 사업에서 일어날 수 있는 다양한 리스크 관련 주제에 대한 연구 결과를 발표하였다. 이러한 연구 결과에는 발표 시점에서 관심을 갖거나 쟁점이 되는 암묵지 성격의 정보를 내포하고 있다. 본 연구는 네트워크 중심성을 사용하여 건설 사업 리스크에서 중심적 위치에 있는 관심을 갖거나 쟁점이 되는 정보가 시간이 경과하면서 어떻게 변화하였는지를 파악한다면 리스크 관리와 예측에 도움이 되는 자료로 사용할 수 있을 것이라 판단하였다. 이를 위해 온라인으로 원문 수집이 가능한 학술과 학위 논문 및 정기간행물의 제목과 요약문을 수집하고, 전처리 과정을 통해 기초 데이터를 추출하였다. 네트워크 중심성 모형들 중 DC, EC 및 Katz 중심성을 사용하여 네트워크에서 단어들의 중심적 위치를 측정하였다. 이때 중심적 위치에 있는 단어는 이웃 단어들을 대표하는 키워드의 역할을 한다고 생각하였다. 중요도가 높은 키워드들 간의 관계, 건설수주 추이 및 사업 단계와 관련하여 키워드가 시간이 흐르면서 어떻게 변화하였는지를 살펴보았다. 결론으로, 본 연구 결과의 제약과 추가적인 연구에 대해 제시하였다.

Abstract Risk in a construction project is an important process management factor that determines the success of the project. Several domestic researchers have published research results on various risk-related topics that can occur in construction projects. The research results contain information of a tacit nature of interest or issue at the time of publication. It was judged that the information could be used as useful data for risk management and forecasting if we understand how the information of interest or issue in a central position in the construction project risk has changed over time using network centrality. For this purpose, titles and summaries of academic papers, dissertations, and periodicals available online were collected, and basic data were extracted. DC, EC, and Katz centrality were used to measure the central position of words in the network. At this time, we thought that the word in the central position could serve as a keyword that represents the neighboring words. We looked at how keywords changed over time in relation to the relationship between keywords with high importance, the trend of construction orders, and the process stage. The limitations of the results of this study and additional studies are presented.

Keywords : Construction Risk Management, Network Centrality Analysis, Eigenvector Centrality, Katz Centrality, Trend Analysis

이 논문은 과학기술정보통신부의 재원으로 한국건설기술연구원 “(22주요-대1-목적)미래 건설산업 견인 및 신시장 창출을 위한 스마트 건설기술 연구(2/2)” 과제의 지원을 받아 수행되었음.

*Corresponding Author : Seong-Yun Jeong(KICT)

email: syjeong@kict.re.kr

Received May 26, 2022

Accepted August 3, 2022

Revised July 11, 2022

Published August 31, 2022

1. 서론

1.1 연구의 필요성 및 목적

대형 건설 사업은 비용이 공사 초기에 집중적으로 투자되고, 공작물이 완공되기까지 수개월에서 수십 년까지 사업 기간이 상당히 길다. 사업 기간 동안 사업 관계자들 간에 이해 충돌을 비롯하여 설계 변경, 투자예산 확보 및 수익성의 불투명, 안전사고 등 내부적인 요인과 천재지변, 법·제도의 변경, 자재비 인상, 민원 발생 등 외부적인 요인으로 인해 당초 사업 계획대로 진행되는 경우가 거의 없다. 심지어 사업이 중단되는 경우도 있다. 이런 경우에 이미 집행된 비용을 회수하기가 쉽지 않다. 이처럼 건설 사업에서 발생할 수 있는 리스크 관리와 예측에 대한 중요성이 강조되면서 국내 여러 학자들은 건설 사업 관련 리스크 요소, 정성적·정량적 분석 기법 및 예측 방안 등 여러 주제의 연구 결과를 발표하였다. 리스크는 22년 1월에 발생한 광주 화정 아이파크 붕괴 사고와 같이 리스크가 발생하면 그때서야 관심을 갖는다. 이처럼 연구 주제에는 발표 시점에 관심이 되는 건설 사업 리스크에 관한 정보를 내포하고 있다. 하지만 이때의 관심 정보는 개별 자료에 따른 특정 주제에 국한된 것이라 할 수 있다. 이런 단편적 정보를 기초로 하여 리스크 관심 정보를 시기별 동향으로 분석하는 데에는 한계가 있다. 이러한 한계를 보완하기 위한 방법으로 개별 리스크 자료들을 하나의 자료로 취합한 후에 네트워크 이론(Network Theory)을 사용하여 단어들 간의 연관 관계를 분석하고, 시간이 흐름에 따라 리스크 자료에 내재된 유의미적 정보가 어떻게 변화하였는지를 도식화 하여 살펴볼 수 있다. 이렇게 도식화된 변화 추이는 건설 사업에서의 리스크 관리와 예측에 대응할 수 있는 기초 자료에 유용할 수 있을 것으로 판단하였다. 하지만 국내에서 건설 사업 리스크에 대해 텍스트 마이닝을 적용하는 유사 연구 사례는 있었으나 네트워크 중심성 분석과 키워드의 시기별 변화 추이를 연구한 사례는 아직까지 없는 것으로 파악된다. 본 연구는 네트워크에서 중심적 위치에 있는 단어가 이웃 단어를 대표하는 키워드라고 간주하고, 키워드가 기간별로 어떻게 변화하였는지를 파악한다면 리스크 관리와 예측에 도움을 줄 수 있는 자료로 사용할 수 있는지를 보여주고자 하였다.

1.2 연구 흐름

본 연구는 네트워크 중심성을 이용한 건설 사업 리스

크 관련 키워드의 변동 추이를 분석하기 위해 Fig. 1과 같은 절차로 연구를 진행하였다. 먼저, 2000년부터 2021년까지 국내에서 발표된 건설 사업의 리스크를 주제로 한 논문과 정기간행물 자료를 수집하였다. 수집한 자료의 제목과 요약문 및 주제어를 추출하여 하나의 분석용 콘텐츠로 가공하였다. 전처리와 형태소 분석 과정을 통해 단어를 추출하였고, 기간별로 단어의 출현 빈도를 분석하였다. 또한, 네트워크 중심성에 대한 이론과 선행 연구 사례를 살펴보았다. 국지적으로 네트워크의 중심적 위치를 추정하는 연결 중심성과 전면적으로 중심적 위치를 추정하는 고유벡터 중심성과 Katz 중심성을 분석 모형으로 선정하였다. 다음으로, Netminer 분석도구를 이용하여 모형별로 키워드의 중심성 계수를 측정하였다. 중심성 모형별로 중심적 위치에 있는 키워드를 대상으로 기간별로 어떻게 변화하였는지를 도식화 하여 살펴보았다. 끝으로, 본 연구의 한계와 추가적인 연구의 필요성을 제시하였다.

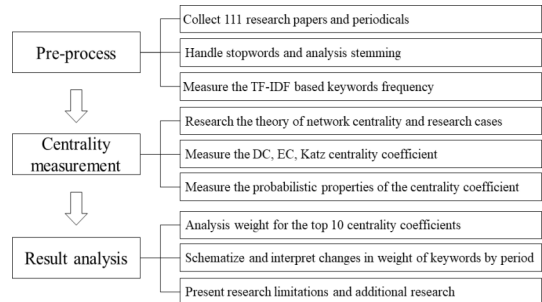


Fig. 1. Research procedure of this study

2. 이론적 고찰

2.1 네트워크 중심성 분석

네트워크 이론에서는 네트워크를 그래프(graph)라는 그림으로 표현하고, 그래프는 개체를 하나의 노드로 표시하고, 노드 간의 관계를 연결선(edge)으로 나타낸다 [1]. 즉, 리스크 네트워크에서 단어를 노드로 보고, 노드와 노드 간의 관계를 연결선으로 표시한다. 이때 특정 노드와 연결된 이웃 노드 간의 연결 방향에 따라 In 또는 Out으로 표시한다. 노드의 중심성 계수(Centrality Coefficient)는 네트워크에서 노드의 중심적 위치를 정량적으로 나타내는 수치로 나타낸 것이다. 중심적 위치는 중요도(importance)나 강도(strength), 영향도(power)

등 의미를 갖는다. 네트워크 이론에서는 PageRank, HITS, Status, Power, Community, Effects, TextRank, Harmony 등 여러 모형의 중심성을 파생하여 사용하고 있다. 이들 모형 중 연결 중심성, 근접 중심성, 매개 중심성, 고유벡터 중심성 등을 대표적으로 사용하고 있다. 이들 중심성은 네트워크에서 차지하는 노드의 위치, 연결 선 및 방향을 반영하여 하나의 행렬로 치환한다. 치환된 노드의 상대적 중요도를 정량적 계수로 표시한다. 이때 중심성 계수가 클수록 네트워크에서 중심에 위치하고, 그만큼 해당 노드의 중요도가 크다고 해석한다. 이때 중요도가 큰 노드는 이웃 노드들의 중계자 또는 허브의 역할을 한다.

2.1.1 연결 중심성(Degree Centrality)

연결 중심성(DC)은 특정 노드와 직접 연결된 노드의 개수를 가지고서 네트워크에서 특정 노드의 중심적 위치를 측정하는 모형이다. 여기서 Degree는 특정 노드와 직접적으로 연결된 이웃 노드의 개수를 의미한다. 하나의 특정 노드에 나머지 노드들이 연결된 스타 형태의 네트워크에서는 특정 노드의 중심성 계수는 최댓값을 갖는다. 반대로, 링 형태의 네트워크는 모든 노드의 계수가 동일한 값을 갖는다. 특정 노드와 이웃 노드 간의 연결 방향에 따라 DC 계수를 달리 계산한다. DC는 계수 계산식과 해석이 간단하기 때문에 거의 모든 네트워크 유형에 적용할 수 있다는 장점을 갖는다. Eq. (1)은 방향성을 고려하지 않는 경우에 DC 계수를 구하는 계산식이다[2].

$$C_d = \frac{d_i}{n-1} \quad (1)$$

여기서 C_d 는 연결 중심성 계수이고, d_i 는 i 노드에 연결되는 정도, 즉 연결선의 수를 말한다. n 은 네트워크에 포함된 모든 노드의 수를 의미한다.

한편, 네트워크의 집중화 정도를 나타내기 위해 Eq. (2)와 같이 연결 집중화 지수(Degree Centralization Index)를 사용할 수 있다[2]. DCI가 1에 가까울수록 스타 형태의 네트워크를 띄운다. 반대로, DCI가 0에 가까울수록 링 형태의 네트워크를 띄운다.

$$C_{id} = \frac{\sum_{i=1}^N [C_{id}(n^*) - C_{id}(n_i)]}{(N-1)(N-2)} \quad (2)$$

여기서 $(N-1)(N-2)$ 는 네트워크에서 중요도가 가장 높은 값이고, $C_{id}(n^*)$ 는 네트워크에서 가장 큰 값을 의미한다.

2.1.2 근접 중심성(Closeness Centrality)

DC는 중요하지 않은 이웃 노드가 많이 연결되는 경우에 자칫 특정 노드가 중요하다고 해석할 수 있다. 이러한 단점을 보완하기 위해 근접 중심성(CC)을 사용할 수 있다. CC는 특정 노드와 연결된 이웃 노드와의 거리가 짧을수록 거리의 역수를 취함으로써 특정 노드의 중심성 계수를 계산한다. 즉, 계수가 높은 노드는 연결된 노드들 간의 거리가 짧고, 짧은 만큼 네트워크에서 중심적 위치에 있다고 해석한다. CC는 특정 노드의 위치가 네트워크 중심에 가까울수록 네트워크에서 중요도는 크고, 반대로 주변으로 갈수록 노드의 계수는 작아진다. CC는 연결선이 끊긴 노드(unreachable node)의 중심성 계수를 측정할 수 없다는 단점을 갖는다. CC는 최단거리 경로(노드 탐색, 전염병 확산, 범죄자의 이동 등 거리적 특성을 갖는 네트워크에 적합하다. CC 계수를 계산하는 식은 아래와 같다[2].

$$C_c = \frac{1}{\sum_{j=1}^n d_{ij}} (n-1) \quad (3)$$

여기서 C_c 는 근접 중심성 계수이고, d_{ij} 는 i 노드와 j 노드 간에 연결된 거리를 말한다. n 은 네트워크에 포함된 모든 노드의 수를 의미한다.

2.1.3 매개 중심성(Between Centrality)

BC는 특정 노드와 이웃 노드 간에 최단 경로의 중심에 위치하는 지를 측정하는 모형이다. BC 계수가 큰 노드일수록 이웃 노드 간의 연결성을 통제할 수 있는 영향력이 크다고 해석한다. 즉, 특정 노드의 연결선을 제거할 경우에 이웃 노드들의 연결 관계에 미치는 영향력을 측정할 수 있다. 스타 형태의 네트워크에서는 중심에 위치하는 노드는 중심성 계수가 가장 크고, 나머지 노드들의 계수는 작은 값을 갖는다. 반대로, 네트워크가 링 형태인 경우에는 중심에 있는 노드는 가장 큰 계수를 가지고, 중심에서 멀어질수록 작은 값을 갖는다. BC 분석은 CC와 비슷하게 전염병 확산, 범죄 확산, 배송 경로 등과 같이 공간적 특성을 갖는 네트워크에 적합하다. Eq. (4)는 BC 계수를 구하는 계산식이다[2].

$$C_b = \frac{\sum_{j,k} g_{jk}(i)}{(n-1)(n-2)/2} \quad (4)$$

(단, $j \neq k = i, j, k, i \in N$)

여기서 C_b 는 매개 중심성 계수이고, g_{jk} 는 j 노드와 k 노드 간에 최단거리 경로의 수를 말한다. $g_{jk}(i)$ 는 j 노드와 k 노드 간에 최단거리 경로 중 i 노드를 통과하는 경우의 수를 의미한다. $(n-1)(n-2)/2$ 는 i 노드를 제외한 모든 노드 쌍의 개수이고, n 은 네트워크에 포함된 모든 노드의 수를 말한다.

2.1.4 고유벡터 중심성(Eigenvector Centrality)

DC 분석은 연결된 이웃 노드가 갖는 중요도를 고려하지 않고 단순히 연결선의 개수를 가지고서 특정 노드의 중요도를 추정한다. 만약 특정 노드가 중요도가 큰 이웃 노드들로 연결되었다면 특정 노드의 중요도는 달라질 수 있다. 즉, EC는 특정 노드와 연결된 이웃 노드의 중심성 계수와 최단 거리를 고려한 누적된 연쇄적인 중요도를 반영하여 특정 노드의 중심성 계수를 측정한다. EC 계수는 특정 노드와 연결된 이웃 노드의 DC 계수를 합산한 값을 의미한다. EC는 BC와 유사하게 스타 형태의 네트워크인 경우에 중심에 위치한 노드는 중요도가 크고, 나머지 노드는 낮은 중요도를 갖는다. 네트워크가 링 형태인 경우에는 네트워크의 중심에 위치한 노드의 중요도는 가장 크고, 주변으로 갈수록 중요도는 작아진다. EC는 유명 정치인이나 연예인과 같이 사회적 영향력, 언론 매체나 인용수가 많은 논문의 영향력 또는 특정 주제에서의 키워드가 갖는 영향력 등을 분석하는 네트워크에 적합하다. EC 계수를 구하는 계산식은 아래와 같다[2].

$$C_e(v_i) = \frac{1}{\lambda} \sum_{j=1}^n A_{j,i} c_e(v_j) \quad (5)$$

여기서 C_e 는 고유벡터 중심성 계수이고, A 는 인접 행렬, λ 은 고유 값(Eigenvalue), 최적해 산출을 위한 상수를 의미한다. n 은 네트워크에 포함된 모든 노드의 수를 말한다.

2.1.5 Katz 중심성(Katz Centrality)

EC는 기본적으로 노드의 방향성을 고려하지 않고 연결된 이웃 노드들의 누적된 중요도를 반영하는 반면에 Katz 중심성은 노드의 방향성을 고려하고, 특정 노드와 연결된 노드뿐만 아니라 간접적으로 연결된 이웃 노드의 연결 정도까지 중요도를 반영한다. 여기서 연결 정도는 이웃 노드 간의 거리를 의미한다. 이 중심성 모형은 SNS에서의 영향력이 있는 사람을 찾거나 시간의 흐름에 따라 영향력의 변화를 측정하는 등 노드가 많고 연결 경로는 짧은 네트워크에 적합하다. 식 (6)은 Katz 중심성 계

수를 구하는 계산식이다[2].

$$C_{katz}(i) = \sum_{k=1}^{\infty} \sum_{j=1}^n \alpha^k (A^k)_{ji} \quad (6)$$

여기서 α 는 감쇠요인(attenuation factor)을 의미하고 $0 < \alpha < \frac{1}{\lambda}$ 에 해당하는 값을 갖는다. k 는 네트워크의 모든 노드 간의 거리(walk)를 말한다. 특정 노드가 다른 모든 노드로 연결되는 거리(walk)를 많이 가질수록 특정 노드의 중심성 계수가 크다는 것을 의미한다.

2.2 선행 연구 사례

네이버와 구글의 학술정보 사이트를 이용하여 네트워크 중심성 관련 연구 사례를 조사하였다. 네트워크 중심성을 이용하여 건설 사업의 리스크 관련 키워드의 동향을 분석한 연구 사례는 국내에서 아직까지 없는 것으로 파악된다. 다만, 이원배 외[3]은 텍스트 마이닝을 활용하여 신문, 논문, 사업보고서에 포함된 단어에 대한 TF-IDF 가중치를 가지고서 해외 건설 사업의 리스크 분석을 위한 프레임워크를 제안하였다. 성현곤 외[4]는 학술지에 게재된 국토 및 도시계획 논문 제목과 요약문을 구분하여 시기별 연구 주제어 비중에 대한 변동 추세를 보여주었다. 한편, 네트워크 중심성을 이용한 연구 사례로는 이택균[5], 이정원 외[6], 황재교 외[7], 조성용 외[8], 이정림 외[9], 정근하[10], 및 김하연[11]은 특정 주제와 관련한 키워드의 빈도수와 키워드로 구성된 네트워크에서 노드 간의 중심성과 연구 동향을 파악하기 위해 DC, CC 및 BC에 대한 중심성 계수를 측정하였다. 이처럼 선행연구에 있어 특정 주제와 관련한 노드가 네트워크에서 얼마나 중심적 위치에 있는지를 측정하기 위해 주로 DC, CD 및 BC를 사용하였다. 최정목[12]는 행정학과 정책학 분야의 학술지 간의 주제어의 중요도를 파악하기 위해 DC, 베타 중심성 및 EC 분석을 하였다. 박찬규[13]은 DC, 권위와 허브 중심성 및 권위 중심성 간의 상관관계를 분석하여 지주회사와 같이 기업 지배구조 네트워크를 분석하였다. 베타 중심성, 권위와 허브 중심성과 같이 특정 주제에 적합한 중심성 모형을 적용하였다. 이외에도 임병학 외[14]는 항만물동량에 대한 네트워크 변수의 영향을 분석하기 위해 DC, CD, BC와 함께 EC를 적용하였다. 하지만 EC와 Katz 중심성을 적용한 사례는 상대적으로 많지 않은 것으로 보였다. 이는 특정 노드와 연결된 이웃 노드의 계수를 연쇄적으로 반영해야 하는 등의 상대적으로 복잡한 계산 과정으로 기인한 것으로 사료된다.

3. 분석 결과

3.1 기초 데이터

본 연구는 기초 데이터를 수집하기 위해서 2015년부터 건설기술과 관련한 자료와 정보를 온라인과 오프라인으로 수집하고, 더블린 코어(Dublin Core) 표준으로 가공한 후 웹서비스를 제공하는 건설기술정보시스템(CODIL)[15]을 운영한 경험을 바탕으로 먼저, 네이버와 구글의 학술정보 사이트에서 '건설 사업'과 '리스크' 등 검색어와 2000년 이후 자료 등 조건으로 검색하였다. 검색 결과를 토대로 학술 논문은 DBpia, 한국학술정보(KISS), 한국교육학술정보원(KERIS), 국회전자도서관 및 논문을 게재한 학회 등 정보제공 사이트에서 무료이면서 원문 수집이 가능한 학술 논문을 무작위로 수집하였다. 원문 내용을 확인한 후에 건설 사업의 리스크와 관련한 논문만을 선정하였다. 학위 논문은 같은 방법으로 국회전자도서관 웹사이트 등을 통해 원문을 조회할 수 있는 논문을 수집하였다. 또한, 정기간행물은 동일한 검색어와 검색 조건을 통해 네이버에서 원문 수집이 가능한 자료를 선정하였다. 이렇게 온라인을 통해 Table 1과 같이 건설 사업의 리스크와 관련하여 2000년 이후에 발표된 111건의 학술과 학위 논문 및 정기간행물을 선정하였다. 선정한 자료는 50건이 국내, 61건은 해외 건설사업 리스크와 관련하였다. 선정한 자료는 5년 단위로 분류하였고, 자료 제목과 요약문, 논문의 경우에는 키워드를 추출하였다. 요약문이 없는 정기간행물의 경우에는 서론을 추출하였다. 이렇게 추출한 내용을 하나의 분석용 콘텐츠로 취합하여 정리하였다.

Table 1. Number of collected material though online

Division		'00~'05	'06~'10	'11~'15	'16~'21	Sum
Total	Academic journal	14	10	25	34	83
	Theses	0	3	1	6	10
	Periodical	2	4	6	6	18
Domestic	Academic journal	13	6	12	10	41
	Theses	0	3	1	1	5
	Periodical	1	2	0	1	4
Overseas	Academic journal	1	4	13	24	42
	Theses	0	0	0	5	5
	Periodical	1	2	6	5	14

다음으로, 서대호[16]가 제시한 정규 표현식과 사전 처리 과정을 준용하여 분석용 콘텐츠에서 0-9까지의 숫자와 ., -, (,), %와 같은 특수 문자 등을 제거하였다. 공

백을 기준으로 하여 콘텐츠에 포함된 단어를 분리(stemming)하였다. 한글과 영문에 따라 품사(Part Of Speech)를 구성하는 방식이 다르고, 분석용 콘텐츠는 대부분 한글 형태로 작성되어 있기 때문에 A-z까지의 영문자로 기술된 단어는 분석 대상에서 제외시켰다. 품사 중 명사에 해당하는 단어는 다른 품사의 단어에 비해 독립적인 의미를 가지고, 대부분의 텍스트 마이닝에서와 같이 명사 단어를 대상으로 분석하였다. 본 연구는 한글 형태소를 분석하는 파이썬의 konlpy 패키지를 사용하여 명사에 해당하는 단어를 추출하였다. 추출한 단어들 중 앞, 뒷, 입, 류, 뜻, 중, 협 등 한 음절 단어와 전체, 각종, 최근, 전체, 진행 등과 같이 단독으로 의미 전달이 부족하거나 중요도가 낮다고 판단되는 명사를 불용어로 지정하여 분석 대상에서 제외시켰다. 다만, 동일한 의미를 갖는 단어는 여러 형태로 표시할 수 있기 때문에 단어마다 동일 또는 유사한 의미를 갖는 단어를 취합하여 정의하는 공인된 시소러스(Thesaurus) 사전이 필요하다. 하지만 건설사업 리스크와 관련한 시소러스 사전을 찾지 못하였다. 따라서 본 연구에서는 동의어 또는 유사어를 별도로 정의하지 않았다. 이러한 전처리 과정을 통해 Table 2와 같이 기간별로 분석용 단어를 추출하였다. Table 2에서 보면 00~05년 동안에 분석용 콘텐츠에서 210개 단어를 추출하였고, 해외 리스크와 관련한 콘텐츠에서 117개 단어를 추출하였다. 국내와 해외 리스크를 합친 콘텐츠에서는 289개 단어를 추출하였다. 이때 국내와 해외에 각각 추출한 단어의 개수와 합친 후 추출한 단어의 개수 간에는 38개가 차이가 있었다. 이러한 차이는 국내와 해외 리스크 콘텐츠에서 동일한 단어의 개수를 의미한다.

Table 2. Number of extracted words for analysis

Division		'00~'05	'06~'10	'11~'15	'16~'21	whole
Construction project	Entire	289	411	616	591	1,177
	Domestic	210	296	306	278	691
	Overseas	117	173	456	592	870
Number of words		306	485	951	896	2,638

3.2 단어의 출현 빈도 분석

보편적으로 문장, 문단 또는 문서에서 특정 단어가 빈번하게 출현(Term Frequency)하는 경우에 해당 단어는 문장, 문단 또는 문서에서 중요하다고 간주한다. 하지만 중요도가 낮은 단어의 출현 빈도가 아무리 많아도 해당 단어의 중요도는 그리 높지 않다. 이를 보완하기 위해 단순

단어출현 빈도(TF)와 역문서 빈도(Inverse Document Frequency)를 곱한 단어빈도-역문서 빈도(TF-IDF) 방식을 사용하여 단어의 출현 빈도를 측정하였다. 이때의 출현 빈도는 가중치를 의미한다. Table 3은 TF-IDF 측정을 통해 출현 빈도가 높은 상위 10개 단어의 가중치와 전체 가중치에서 해당 단어가 차지하는 가중치 비율을 나타낸 것이다.

Table 3. TF-IDF measurement result for the top 10 words

Period	'00~'05	'06~'10	'11~'15	'16~'21	whole
Top 10 key words	government 0.946 (0.355%)	option 1,187 (0.317%)	government building 1.529 (0.156%)	landscape 1.330 (0.138%)	government building 1.556 (0.051%)
	finance 0.946 (0.355%)	exchange rate 0.959 (0.256%)	project costs 1.472 (0.150%)	automatic 1.213 (0.126%)	landscape 1.512 (0.050%)
	civilian 0.946 (0.355%)	real 0.959 (0.256%)	person involved 1.303 (0.133%)	culture 1.213 (0.126%)	person involved 1.378 (0.046%)
	builder 0.946 (0.355%)	small and medium 0.959 (0.256%)	1.303 (0.133%)	Myanmar 1.173 (0.122%)	mimesis 1.378 (0.046%)
	degree of strength 0.860 (0.323%)	currency 0.913 (0.244%)	intellect 1.303 (0.133%)	reserve costs 1.095 (0.114%)	intellect 1.378 (0.046%)
	weight 0.860 (0.323%)	public 0.913 (0.244%)	recycling 1.246 (0.127%)	accident 1.095 (0.114%)	recycling reclamation 1.334 (0.044%)
	portfolio 0.860 (0.323%)	housing 0.913 (0.244%)	reclamation, 1.246 (0.127%)	article 1.095 (0.114%)	agent 1.334 (0.044%)
	order intakes 0.774, (0.290%)	variance 0.913 (0.244%)	agent 1.246 (0.127%)	news 1.095 (0.114%)	real 1.334 (0.044%)
	Korea 0.774, (0.290%)	product 0.868 (0.232%)	finance 1.212 (0.124%)	multiple 1.095 (0.114%)	Myanmar 1.334 (0.044%)
	ranking 0.774, (0.290%)	abandonm ent0.868 (0.232%)	fair 1.212 (0.124%)	blockchain 1.056 (0.109%)	facility 1.289 (0.043%)
continuity 1.189 (0.124%)					
Mean	0.514	0.518	0.624	0.688	0.612
STD. DEV.	0.169	0.177	0.218	0.193	0.263
Min	0.172	0.174	0.201	0.272	0.163
Max	0.946	1.187	1.529	1.330	1.556

Table 3에서 보듯이 기간에 따라 상위 10위 안에 포함된 단어가 달랐다. 11~15년에는 상위 단어와 전체 기간의 단어가 가장 많이 중첩하였다. 다만, 가중치의 차이가 큰 단어는 없는 것으로 보였다.

3.3 중심성 계수 측정

TF-IDF는 출현 빈도를 가지고서 특정 단어의 가중치를 측정할 수 있으나 단어와 단어 간의 연관 관계와 연관 관계에서의 특정 단어의 중요도를 살펴보는 데 한계가 있다. 이를 보완하기 위한 네트워크 중심성 모형을 사용하였다. 2.1절에서 설명하였듯이 네트워크 이론에서는 특정 노드가 네트워크에서 차지하는 중심적 위치 즉, 중요도를 측정하기 위해 다양한 중심성 모형을 사용할 수 있다. CC와 BC는 거리와 공간 개념에서 접근하는 네트워크 유형에 적합한데 본 연구에서 사용하는 키워드 간의 관계를 그린 네트워크 유형을 사용하기 때문에 본 연구의 중심성 모형에는 적합하지 않다고 판단하였다. 한편, DC는 직접적으로 연결된 이웃 노드의 개수를 가지고서 중심성 계수를 계산하기 때문에 대부분 네트워크 유형에서 사용될 수 있는 가장 기본적인 중심성 모형이다. 본 연구에서는 DC가 직접 연결된 이웃 노드의 개수만을 고려하는 제약을 보완하기 위해 이웃 노드의 중요도를 연쇄적으로 반영하는 EC와 Katz 중심성을 DC와 함께 중심성 계수를 측정하는 모형으로 사용하였다. 본 연구에서는 Table 2와 같이 기간별로 리스크 관련 단어들과 단어와 단어 간의 연결 관계를 그린 네트워크를 구성하였다. 네트워크에서 중심적 위치에 있는 단어는 자신과 연결된 이웃 단어의 중요도를 연쇄적으로 반영하였기 때문에 이웃 단어들을 대표할 수 있다. 따라서 중요도가 높은 단어는 이웃 단어를 대표하는 키워드라고 간주하였다. 이 키워드를 통해 관련된 단어들을 찾을 수 있을 것으로 생각하였다.

3.3.1 DC 계수 측정

본 연구에서는 Table 2와 같이 기간별로 추출한 단어들을 대상으로 하여 단어별로 DC 계수를 계산하였다. Table 4는 DC 분석을 통해 얻은 상위 10개의 키워드에 대한 DC 계수를 나타낸 것이다. Table 4에서 보듯이 전체 기간에서 '시장' 키워드의 계수가 0.048로서 가장 높은 것으로 측정되었다. 이 계수는 평균이 0.004에 비해 12배 이상 높은 수치이다. 이만큼 네트워크에서 중심적 위치에 있다고 해석할 수 있다. 기간별로 볼 때 11~15년에 최상위의 키워드의 계수가 평균에 비해 10배 이상 높았다. 반대로, 06~10년에는 5배보다 낮은 계수를 가진 것으로 보였다. 다만, 전체적으로 표준 편차가 0.005~0.006 사이에 있기 때문에 중요도가 매우 큰 키워드는 많지 않은 것으로 보였다. 한편, DC index도 5%

미만으로 측정되었기 때문에 스타 형태와 같이 중앙집중식 네트워크보다는 링 형태의 네트워크 모양을 갖는다고 추정할 수 있다. TF-IDF를 통해 얻은 상위 키워드의 가중치와 비교할 경우에 일치하는 키워드가 하나도 없는 것으로 파악되었다. TF-IDF에서 측정된 가중치의 표준편차가 0.169~0.263인데 비해 DC의 표준 편차가 상대적으로 낮은 것으로 보였다.

Table 4. Top 10 keywords and distribution of DC

Period	'00~'05	'06~'10	'11~'15	'16~'21	whole	
Top 10 words	hierarchy (0.038), decision (0.035), contract (0.035), system (0.031), market (0.031), factor (0.031), intention (0.024), build (0.021), model (0.021), classification(0.021)	development(0.029), configuration(0.029), revenue (0.027), form (0.027), bid (0.024), response (0.022), market (0.022), company(0.022), prediction(0.022), uses (0.022)	order intakes(0.052), market (0.042), hierarchy(0.039), development(0.034), uses (0.028), industry (0.026), perform (0.026), continuing (0.026), system (0.024)	contract (0.042), market (0.037), debouchment (0.036), result (0.029), information(0.027), development(0.025), deduction(0.025), revenue (0.025), system expert (0.025)	market (0.048), intakes(0.040), hierarchy (0.038), contract (0.037), development(0.037), uses (0.036), model (0.033), system (0.032), effect (0.031), result (0.028)	market (0.048), intakes(0.040), hierarchy (0.038), contract (0.037), development(0.037), uses (0.036), model (0.033), system (0.032), effect (0.031), result (0.028)
	Mean	0.007	0.006	0.005	0.005	0.004
	STD. DEV.	0.006	0.005	0.006	0.005	0.005
	Min	0.003	0.002	0.002	0.002	0.001
	Max	0.038	0.029	0.052	0.042	0.048
	Degree Centrality Index(%)	3.106	2.363	4.717	3.731	4.38

3.3.2 EC 계수 측정

Table 5는 EC 분석으로 통해 얻은 상위 10개의 키워드와 네트워크에서 영향을 미치는 계수를 나타낸 것이다. 전체 기간에서 '결정' 키워드의 계수가 0.697로서 가장 높았고, 상위 10개의 키워드의 평균 계수는 전체 키워드의 평균치인 0.003에 비해 68배 이상 높은 것으로 측정되었다. 이러한 수치는 상위 10개의 키워드가 네트워크에서 중요도가 큰 이웃 노드들과 연속적으로 연결된 결과라고 해석할 수 있다. 또한, EC 계수 간의 표준 편차가 0.029로서 DC의 표준 편차에 비해 편차의 폭이 큰 것으로 나타났다. 한편, 0.0005 이하의 계수를 갖는 키워드가 390개로서 이 수치는 전체 키워드에서 33.2%를 차지하였다. 기간별로 볼 때 최상위 키워드의 계수는 평

균 계수에 비해 약 167.5배 이상 높은 것으로 측정되었다. 16~21년에서 '중소중견' 키워드의 계수는 차 순위의 '조경' 키워드에 비해 2.8배 이상의 높은 영향력을 갖는 것으로 추정되었다. TF-IDF의 상위 키워드의 가중치와 비교하면 06~10년에서 '옵션', '실물' 등의 키워드와 11~15년에서 '재정', 16~21년에서 '조경' 키워드가 상위 순위를 차지하였다. 표준 편차는 TF-IDF에서 측정된 가중치보다 낮았으나 DC보다는 높았다.

Table 5. Top 10 keywords and distribution of EC

Period	'00~'05	'06~'10	'11~'15	'16~'21	whole	
Top 10 words	decision (0.686), intention (0.627), model (0.274), proposal (0.076), procedure(0.061), chance (0.061), function (0.061), propel (0.060), added (0.020), value (0.060), tree (0.060)	development(0.399), housing (0.398), option (0.391), real (0.288), prediction(0.286), model (0.283), public (0.275), abandonm (0.202), investment (0.165), uses (0.161)	decision (0.670), intention (0.633), support (0.309), finance (0.193), system (0.059), strategy (0.046), hierarchy (0.043), solid (0.043), development(0.041), model (0.024)	small and medium (0.700), landscape (0.249), engineering(0.127), System (0.074), information(0.064), deduction (0.054), result (0.054), lack (0.054), company (0.44), change (0.037)	decision (0.679), intention (0.665), support (0.192), finance (0.120), model (0.114), development(0.069), decision maker (0.067), strategy (0.059), system (0.044), hierarchy (0.040)	decision (0.679), intention (0.665), support (0.192), finance (0.120), model (0.114), development(0.069), decision maker (0.067), strategy (0.059), system (0.044), hierarchy (0.040)
	Mean	0.01	0.014	0.004	0.005	0.003
	STD. DEV.	0.058	0.047	0.04	0.041	0.029
	Min	0	0	0	0	0
	Max	0.686	0.399	0.67	0.7	0.679

3.3.3 Katz 중심성 계수 측정

Table 6은 Katz 중심성 분석을 통해 특정 키워드가 직간접적으로 연결된 이웃 키워드의 중요도를 반영한 상위 10개의 키워드와 네트워크에서 차지하는 계수를 나타낸 것이다. EC 분석과 유사하게 전체 기간에서 '결정' 키워드의 계수가 1.871로서 가장 높았다. 상위 10개 키워드의 평균 계수는 전체 키워드의 평균치인 0.091에 비해 15배 이상 높은 것으로 측정되었다. 이러한 수치는 EC 분석에 비해 중요도는 낮으나, 그래도 네트워크에서 영향력을 갖는다고 해석할 수 있다. 표준 편차에 있어서 Katz 중심성은 0.173으로서 EC의 0.029보다 편차의 폭이 큰 것으로 측정되었다. 평균 중심성 계수인 0.091보다 작은 키워드가 891개이다. 이는 전체 키워드에서 75.6% 이상을 차지하는 것으로, 상위를 차지하는 키워드

의 중요도가 크다고 유추할 수 있다. EC와 비교하면 Katz 중심성을 통해 얻은 상위 10개 키워드 중 '시장', '진출', '구축' 키워드만 상이하고, 나머지 키워드는 동일한 것으로 나타났다. 끝으로, '시스템' 키워드가 모든 기간에서 중심적 위치를 차지하였다. 00~05년과 11~15년에서 동일한 키워드가 많았다. TF-IDF를 통해 얻은 상위 키워드의 가중치와 비교하면 06~10년에서 '옵션', '실물' 등 키워드와 일치하였다. 표준 편차는 TF-IDF에서 측정된 가중치와 유사하였으나, DC와 EC 보다는 높은 것으로 측정되었다.

Table 6. Top 10 keywords and distribution of Katz centrality

Period	'00~'05	'06~'10	'11~'15	'16~'21	whole	
Top 10 words	decision (1.873), intention (1.563), hierarchy (1.501), classification(1.331), model (0.971), factor (0.648), contract (0.601), system (0.529), proposal (0.481), discrimination(0.481)	development(1.965), prediction (1.912), option (1.834), model (1.649), housing (1.459), configuration(1.405), profit (1391), real (1.285), system (1.208), uses (1.084)	decision (1.534), intention (1.479), system (1.023), development(0.995), support (0.922), market (0.898), order intakes(0.869), system (0.607), industry (0.547), classification(0.516)	debutment(1.710), small and medium (1.645), contract (1.524), information(1.494), market (1.448), system (1.142), profit (1.049), model (1.048), extraction (1.012), deduction (0.971)	decision (1.871), intention (1.736), development(1.638), model (1.588), hierarchy (1.421), market (1.362), system (1.154), debutment(1.040), support (0.968), build (0.899)	
	Mean	0.147	0.216	0.092	0.151	0.091
	STD. DEV.	0.201	0.271	0.147	0.221	0.173
	Min	0.046	0.053	0.017	0.026	0.01
	Max	1.873	1.965	1.534	1.710	1.871

한편, Katz 중심성을 통해 측정된 키워드의 중요도 크기와 키워드 간의 연결선을 그래프 형태로 나타낼 수 있다. 하지만 Table 2와 같이 기간에 따라 최소 306개에서 최대 2,638개의 단어들을 하나의 그래프로 표현할 경우에 마치 하나의 큰 점으로 표현될 수 있어 키워드 간의 관계를 살펴보기가 쉽지 않다. 따라서 본 연구에서는 Fig. 2와 같이 일정 수준 이상의 중심성 계수를 갖는 키워드만을 추출하여 키워드 간의 연결 관계를 그래프로 나타내었다. Fig. 2에서 보면 00~05년에는 '제안' 키워드는 영향력이 있는 이웃 키워드의 통계를 받는 것으로 보인다. 06~10년에는 '개발'과 '예측' 키워드는 이웃 키

워드에 영향력을 미쳤고, '추정' 키워드는 이웃 키워드로부터 영향을 받았다. 11~15년에는 '개발', '체계', '성능' 키워드는 이웃 키워드에게 영향을 주고받는 중계자 역할을 하는 것으로 보였다. 16~21년에는 '중견', '엔지니어링' 키워드가 이웃 키워드에 영향을 미쳤고, '회사', '계약', '정보', '결과' 등의 이웃 키워드와 상호 연결된 것으로 보였다. '시장', '추정' 키워드는 주로 이웃 키워드로부터 영향을 받는 것으로 보인다.

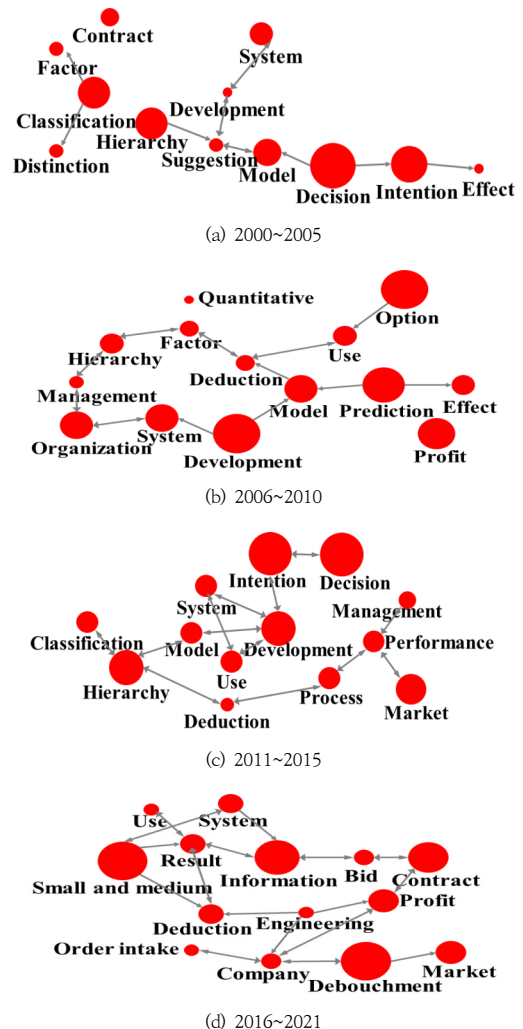


Fig. 2. Katz graph of association between top keywords

3.4 키워드 변화 추이

앞에서 측정된 DC, EC 및 Katz 중심성 계수를 기초로 시간이 경과하면서 네트워크에서 중심성이 어떻게 변

하였는지를 살펴보았다. Table 4~6에서 보듯이 중심성 모형별, 기간별로 상위 10개의 키워드가 서로 상이하였다. 그래서 본 연구에서는 키워드의 변화 추이를 일관되게 분석하기 위해 먼저, 전체 기간을 기준으로 Katz 중심성을 통해 중요도가 높은 순으로 키워드를 우선적으로 선정하였다. 선정된 키워드를 기준으로 기간마다 모두 존재하는지를 확인하였으며, 모든 기간에서 존재하는 10개의 키워드를 선별하였다. 선별한 키워드를 기준으로 DC와 EC에서 동일한 키워드를 선정하였다. 선정한 키워드를 기초로 다음과 같이 DC, EC 및 Katz 중심성별로 키워드의 변환 추이를 도식화 하였다. 먼저, Fig. 3은 DC 분석을 통해 얻은 추정된 10개의 키워드에 대한 변화 추이를 나타낸 것이다. '모델' 키워드는 모든 기간에서 거의 변화가 없었다. '결정', '의사', '시장', '체계' 등의 키워드는 감소-증가-감소하는 변화 추이를 보였다. 이러한 추이는 토목부문의 건설 수주가 06년까지 감소하다가 09~10년에는 증가하였고, 14년까지 감소한 후에 현재까지 증가하는 추이와 유사한 것으로 보인다[17]. 따라서 DC 중심성의 관점에 볼 때 건설 수주가 증가한 시기에는 리스크에서 '결정', '의사', '시장', '체계'가 관심이 높아진다고 유추할 수 있다.

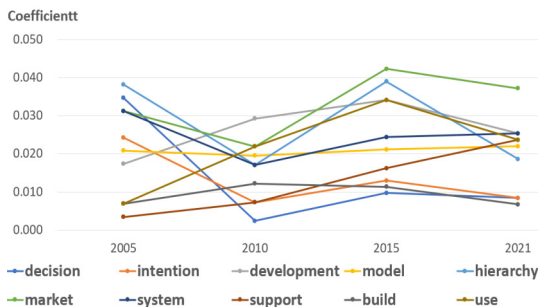


Fig. 3. Trends of DC-based keyword by periods

Fig. 4는 EC 분석 결과로, '결정'과 '의사' 키워드는 DC와 같은 변화 추이를 가졌으며 DC에 비해 변화의 폭은 큰 것으로 보였다. 이러한 변화의 폭은 중요도가 큰 이웃 키워드들과 많이 연결되었다고 추정할 수 있다. '시장', '체계', '시스템', '시공' 및 '활용' 키워드는 증가-감소-증가하는 추이를 가졌다. '시장'과 '체계' 키워드는 DC와 반대되는 추이를 보였다. 이런 현상은 자신의 중요도보다 이웃 키워드의 중요도에 더 큰 영향을 받아서 나타난 것으로 유추할 수 있다.

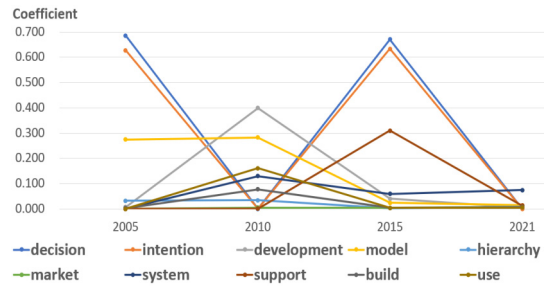


Fig. 4. Trends of EC-based keyword by periods

끝으로, Fig. 5는 Katz 분석을 통해 중심적 위치에 있는 키워드의 변화 추이를 도식화한 것이다. '시장' 키워드에 있어서 시간이 흐를수록 중요도는 지속적으로 증가하였다. 이러한 증가 추세는 모든 기간에서 리스크에서 중요한 관심 또는 쟁점이 되는 키워드라고 할 수 있다. '결정', '의사' 및 '체계' 키워드는 감소-증가-감소 등 추이를 가졌고, 이런 추이는 DC와 같이 토목부문의 건설 수주 변화에 상응하는 추이를 보였다. '결정' 및 '의사' 키워드는 건설사업의 타당성 등 기획 단계와 관련되어 있다. 반대의 추이를 갖는 '시공', '활용', '개발', '시스템' 키워드는 시공 단계와 관련된 것으로 유추할 수 있다. Katz 중심성 관점에서 볼 때 시공 단계와 관련된 키워드는 건설 수주 변화 추이와 반대되었을 때 리스크에 대한 관심은 높은 것으로 사료된다.

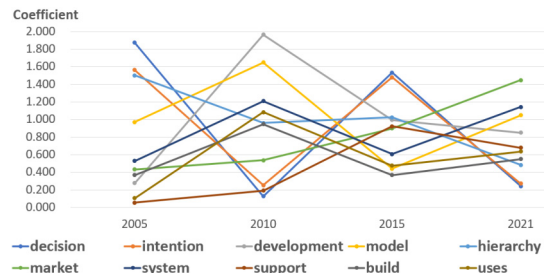


Fig. 5. Trends of Katz-based keyword by periods

이처럼 DC, EC 및 Katz 중심성 분석을 통해 중요도가 큰 키워드들의 변화 추이를 살펴보았다. 이러한 추이는 다음과 같이 정리할 수 있다. 모든 중심성 모형에서 '결정'과 '의사' 키워드는 동일한 변화 추이를 가졌다. 이러한 추이는 토목부문의 건설 수주 추이와 유사하였다. 또한, '결정'과 '의사'는 투자 계획 또는 타당성 평가 등과 같이 기획 단계에 연관될 수 있다. 따라서 '결정'과 '의사' 키워드가 네트워크에서 중심적 위치에 가깝다면 사업의 기획 단계에서 리스크에 관심이 커질 수 있다고

유추할 수 있다. EC 및 Katz 중심성에서는 '시공', '시스템', '활용' 등의 키워드는 동일한 추이를 가졌다. 이러한 추이는 '결정'과 '의사' 키워드와 반대되는 현상이다. 국내 건설 수주의 추이와 반대될 때 시공이나 운용 단계에서 관심을 갖는 리스크 관리 요소라고 유추할 수 있다. '시장' 키워드는 중요도가 높은 이웃 단어들과 연결되었고, 시간이 흐를수록 이웃 단어들의 중요도가 높아지는 것으로 추정되었다. 이는 건설 시장과 관련한 리스크에 대한 관심이 커지고 있다고 유추할 수 있다. 한편, DC와 EC•Katz 중심성 간의 변화 추이를 비교하면 '결정', '의사' '개발' 및 '체계' 키워드는 동일하거나 유사한 추이를 가졌으나 나머지 키워드는 서로 다른 추이를 보였다. 이러한 현상은 이웃 단어의 중요도를 연쇄적으로 반영하였기 때문에 나타난 결과로 보였다. 이러한 결과를 볼 때 어떤 중심성 모형을 사용하였는지에 따라 네트워크에서의 중심적 위치가 바뀔 수 있다. 따라서 국지적 성격을 갖는 DC와 전면적 성격을 갖는 EC와 Katz 중심성의 위치 변화를 상호보완적으로 살펴볼 필요가 있는 것으로 사료된다.

4. 결론

건설 사업에서의 리스크는 사업의 성공을 좌지우지할 만큼 중요한 사업 관리 요소이다. 국내 여러 연구자들은 건설 사업에서 일어날 수 있는 리스크 요인, 분류, 분석, 대응, 예측 등 다양한 주제의 연구 결과를 발표하였다. 이러한 연구 결과에는 발표 시점에서 주요 관심 또는 쟁점에 대한 정보를 포함한다. 이때의 정보는 개별 자료에 국한된 단편적 관심 또는 쟁점이라 할 수 있다. 이런 단편적 정보를 가지고서 기간별로 리스크 관련 정보에 대한 변화를 분석하는데 한계가 있다고 생각하였다. 이러한 한계를 보완하기 위해 네트워크 이론을 사용하여 리스크의 관심 또는 쟁점이 되는 정보가 시간이 흐름에 따라 어떻게 변화하였는지를 분석할 수 있다고 판단하였다. 하지만 국내에서 건설 사업 리스크에 대해 네트워크 중심성 분석과 리스크 정보의 시기별 변화 추이를 연구한 사례는 아직까지 없는 것으로 파악되었다. 본 연구는 네트워크 중심성을 사용하여 건설 사업 리스크에서 중심적 위치에 있는 키워드가 시간이 경과하면서 어떻게 변화하였는지를 파악한다면 리스크 관리와 예측에 도움이 되는 자료로 사용할 수 있는지를 보여주고자 하였다. 이를 위해 온라인으로 원문 수집이 가능한 학술과 학위 논문

및 정기간행물의 제목과 요약문에서 불용어 제거와 형태소 분석 등 전처리 과정을 거쳤다. 이 과정을 통해 추출한 단어를 기초 데이터로 하여 네트워크 중심성 모형들 중 국지적으로 네트워크에서 단어의 중요도를 추정하는 DC와 전면적으로 단어의 중요도를 추정하는 EC와 Katz 중심성을 사용하여 네트워크에서 단어들의 중심적 위치를 측정하였다. 이때 중심적 위치에 있는 단어는 이웃 단어를 대표하고, 이웃 단어들을 찾을 수 있는 키워드의 역할을 할 수 있다고 생각하였다. DC, EC 및 Katz 중심성 모형별로 시간이 경과하면서 키워드가 어떻게 변화하였는지를 도식화 하였다. 그리고 '시장' 키워드는 건설 시장과 관련한 리스크에 대한 관심이 커지고 있는 것을 알 수 있었다. 사업의 기획 단계와 관련성이 있는 '결정'과 '의사' 키워드는 모든 중심성 모형에서 동일한 변화 추이를 가졌고, 이런 추이는 토목부문 건설 수주의 변화 추이와 유사한 것으로 보였다. 반대로, 시공 단계와 관련성이 있는 '시공', '시스템' 및 '활용' 등의 키워드는 건설 수주의 추이와 반대될 때 중요도가 높은 것을 알 수 있었다. 이처럼 네트워크 중심성을 통해 키워드의 중요도에 대한 변화 추이를 분석한다면 포괄적 관점에서 리스크 관련 관심 또는 쟁점이 되는 정보 제공이 가능하다는 것을 보여주었다. 이러한 변화 추이는 리스크 관리와 예측에 대응할 수 있는 기초 자료에 유용하게 사용할 수 있는 것으로 사료된다. 게다가, DC와 EC•Katz 중심성 간의 변화 추이가 상이하였다. 이는 국지적 성격을 갖는 DC와 전면적 성격을 갖는 EC와 Katz 중심성의 위치 변화를 상호보완적으로 분석할 필요가 있다고 판단되었다. 한편, 본 연구는 다음과 같은 제약과 추가적인 연구가 필요하다. 첫 번째, 본 연구에서는 온라인에서 원문을 수집할 수 있는 학술과 학위 논문 및 정기간행물을 기초 자료로 사용하였다. 하지만 수집 자료의 분량과 분야별 편향에 따라 분석 결과는 달라질 수 있다. 따라서 본 연구 결과가 객관적인 리스크의 키워드 동향 분석이라 하기에는 부족함이 있다. 보다 객관적인 연구 결과로 인정받기 위해서는 모든 건설 사업관련 자료를 대상으로 분석하여야 하나, 현실적으로 쉽지 않다. 그래서 국내와 해외별, 가설, 골조, 마감, 설비, 조경 등 공종별, 기획, 설계, 시공, 준공 등 공정별, 도로•교량•터널•항만•철도 등 시설물별, 대형•중건•중소 등 사업 규모별 등 세분화된 기초 자료를 수집하여 키워드 동향을 분석할 필요가 있다. 두 번째로, 건설사업 리스크는 국내의 건설 수주, 사회적 관심, 재난재해, 안전사고 등 주변 상황에 따라 특정 리스크에 대한 관심이나 쟁점이 부각된다. 따라서 주변 상황

에 따라 중요도 높은 키워드가 무엇인지, 키워드가 어떻게 변화하였는지를 지속적으로 분석한다면 건설 사업에서의 리스크 관리와 예측에 대응할 수 있는 기초 자료로서 유용한 가치가 있을 것으로 사료된다. 세 번째로 동일한 의미를 갖는 단어는 여러 형태로 표시될 수 있다. 따라서 명확한 의미 전달을 위해서는 연결된 공공데이터(Linked Open Data)와 비슷한 유사어 사전(Thesaurus) 구축이 선행될 필요가 있다. 끝으로, 본 연구는 이론적 접근보다는 범용적인 DC, EC 및 Katz 중심성 모형을 사용하여 키워드 동향을 분석하였다. 박찬규[13]의 사례와 같이 건설 사업 리스크 분석에 부합한 중심성 모형에 대한 이론과 실증에 대한 연구가 필요할 것으로 판단된다.

References

- [1] W. J. Kim, Network Centrality Theory, P.112, Chaos Book publishers, 2015, pp.3-30.
- [2] T. Y. Kim, Network Centrality using Netminer, Online training materials, Korea, 2022.
- [3] W. B. Lee, M. J. Kim, M. K. An, Y. S. Kim, "Risk Analysis Framework for Overseas Construction Projects by using Text-Mining", *Annual Conference of KICEM*, pp.71-73, 2019. ISSN2287-6065.
- [4] H. G. Sung, et al., "The Analysis on Research Trend of Territorial and Urban Planning for Mack Joong Choi Through Text Mining: Focused on His Academic Papers", *The Korea Spatial Planning Review*, Vol.103, pp.3-26, 2019.
DOI: <http://dx.doi.org/10.15793/kspr.2019.103..001>
- [5] T. K. Lee, "Domestic Research Trend of Internet of Things based on Keyword Frequency and Centrality Analysis", *The Journal of the Korea Contents Association*, Vol.20, No.12, pp.23-35, 2020.
DOI: <https://doi.org/10.5392/JKCA.2020.20.12.023>
- [6] J. W. Lee, K. W. Lee, "Analysis of Seoul Metropolitan Subway Network Characteristics Using Network Centrality Measures", *Journal of The Korean Society for Railway*, Vol.20, No.3, pp.413-422, 2017. ISSN 2288-2235(Online)
- [7] J. G. Hwang, et al., "Analysis Results in Technical Trends of 2018 Farnborough International Airshow via Centrality Analysis", *Journal of the Korea Academia- Industrial Cooperation Society*, Vol.20, No.8, pp.164-173, 2019.
DOI: <https://doi.org/10.5762/KAIS.2019.20.8.164>
- [8] S. Y. Cho, K. S. Byun, "Research Trends of Microplastic in Food via Centrality Analysis Method", *Journal of the Korea Academia-Industrial cooperation Society*, Vol.21, No.5, pp.508-515, 2020.
DOI: <https://doi.org/10.5762/KAIS.2020.21.5.508>
- [9] J. L. Lee, Y. G. Kim, E. J. Kwak, S. G. Park, "A study on research trends for gestational diabetes mellitus and breastfeeding: Focusing on text network analysis and topic modeling", *The Journal of Korean Academic Society of Nursing Education*, Vol.27, No.2, pp.175-185, 2021.
DOI: <https://doi.org/10.5977/jkasne.2021.27.2.175>
- [10] G. H. Jung, A Study of foresight method based on textmining and complexity network analysis, Research Report, Korea Institute of Science & Technology Evaluation and Planning, Korea, pp.25-60, 2010.
- [11] H. E. Kim, *Trend Analysis of Green Logistic Study Using Text Mining*, Master's thesis, Chung-Ang Univ., Korea.
- [12] J. M. Choe, "Investigating Journal Citation Network with Centrality Measures in the Public Administration and Policy Field", *The Society of Digital Policy and Management*, Vol.14, No.9, pp.301-308, 2016.
DOI: <https://doi.org/10.14400/JDC.2016.14.9.301>
- [13] C. K. Park, "A Study on Applying Social Network Centrality Metrics to the Ownership Networks of Large Business Groups", *Korean Management Science Review*, Vol.32, No.2, pp.15-35, 2015.
DOI: <https://dx.doi.org/10.7737/KMSR.2015.32.2.015>
- [14] B. H. Leem, H. J. Chun, "A Study on the Impact of Social Network at Ports on Throughput with a focus on Port Cooperation Index", *POSRI Business and Economic Review*, Vol.11, No.3, pp.289-307, 2011.
UCI: [1410-ECN-0102-2012-320-002976807](https://nrs.ri.ac.kr/ucid/1410-ECN-0102-2012-320-002976807)
- [15] Construction Technology Digital Library, "<https://www.codil.or.kr/>" (accessed June, 30, 2022)
- [16] D. H., Seo, Catch Text Mining with Python, P.203, BL Public, 2019, pp.36-57.
- [17] Statistics KOREA Government Official Work Conference, Construction order trend, https://www.index.go.kr/potal/main/EachDtlPageDetail.do?idx_cd=1220 (accessed June, 30, 2022)

정 성 윤(Seong-Yun Jeong)

[정회원]



- 1992년 2월 : 한양대학교 전자계산학과 (공학학사)
- 1994년 2월 : 숭실대학교 컴퓨터공학과 (공학석사)
- 2018년 2월 : 서울과학기술대학교 산업정보시스템전공 (공학박사)

- 1994년 3월 ~ 현재 : 한국건설기술연구원 미래스마트건설연구본부 연구위원

<관심분야>

건설정보화, 텍스트 마이닝, 투자공학

김 남 곤(Nam-Gon Kim)

[정회원]



- 1989년 2월 : 울산대학교 전자계산학과 (공학학사)
- 2001년 2월 : 공주대학교 컴퓨터공학과 (공학석사)
- 1991년 10월 ~ 현재 : 한국건설기술연구원 미래스마트건설연구본부 연구위원

<관심분야>

건설정보화, 텍스트 마이닝, 빅데이터