

텍스트 마이닝을 이용한 터널 교통안전 연구동향 분석

김진국^{1*} · 양충헌² · 박수빈³한국건설기술연구원 ¹전임연구원, ²연구위원, ³박사후연구원

A Study on the Research Trends for Tunnel Traffic Safety using Text Mining

Jinguk Kim^{1*} · Choongheon Yang² · Subin Park³¹Research Specialist, ²Research Fellow, ⁴Post-doctoral researcher, Korea Institute of Civil Engineering and Building Technology, Goyang-si, Gyeonggi-do, Korea

[요약]

최근 수도권, 부산 등을 중심으로 지하도로 건설이 지속적으로 증가하고 있다. 국내에서는 터널과 관련된 많은 연구들이 현재까지 활발하게 이어져 오고 있으나, 최근 국내에 첫 도입되어 운영 중인 지하도로의 경우 관련 연구를 거의 찾아볼 수 없는 실정이다. 따라서, 본 논문에서는 과거 약 20년 동안 국내 저널에 게재된 터널의 교통안전과 관련된 연구동향 분석을 통해 향후 지하도로 운영시 교통안전 확보를 위해 필요한 연구주제를 도출하였다. 비정형 자료 분석에 주로 사용되는 텍스트마이닝(Text mining) 기법 중 LDA(Latent Dirichlet Allocation) 토픽 모델링(Topic modeling)을 사용하여 연구동향을 분석하고 도출된 토픽별 키워드를 기반으로 토픽들을 분석하였다. 본 연구는 텍스트 마이닝을 통해 지하도로의 교통흐름을 안전하게 유도할 수 있는 교통안전 관련 연구의 방향성을 제시하였다는 데 의의가 있다. 향후 지하도로 운영에 있어 교통안전 확보를 위한 연구주제의 방향성을 설정하는데 기초자료로 활용될 수 있을 것으로 판단된다.

[Abstract]

Recently, construction of underground roads has been continuously increasing in the metropolitan area and Busan. Many studies related to tunnels have been conducted so far, but there are few related studies on underground roads that were recently introduced and operated for the first time. Therefore, this paper in the past about 20 years, local journals published in the next underground roads through the traffic safety-related research and trend analysis of the tunnel traffic safety during operations to secure the necessary research derived. Among the text mining techniques commonly used in unstructured data analysis, research trends were analyzed using LDA-based topic modeling, and topics were analyzed based on the keywords for each topic derived accordingly. This study is meaningful in that it presented the direction of traffic safety-related research that can safely induce traffic flow to underground roads through text mining. This results imply that research trend analysis results would be useful to establishing traffic operation researches and strategies for traffic safety when constructing and opening underground roads in urban areas in the future.

색인어 : 터널, 교통안전, 텍스트마이닝, 토픽 모델링, 잠재 디리클레 할당**Keyword** : Tunnel, Traffic safety, Text mining, Topic Modeling, Latent Dirichlet Allocation(LDA)<http://dx.doi.org/10.9728/dcs.2022.23.10.2075>

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 28 September 2022; Revised 19 October 2022

Accepted 21 October 2022

*Corresponding Author; Jinguk Kim

Tel: +82-31-910-0064

E-mail: jingukkim@kict.re.kr

I. 서론

올해 초 2021년부터 2025년까지 고속도로 건설에 대한 중장기 투자계획을 담은 “제2차 고속도로 건설계획”이 발표되었다. 본 계획은 경제성장을 지원하고 쾌적하고 편리한 고속도로 건설을 위해 지역균형 발전, 교통혼잡 완화, 물류산업 지원, 남북협력 대비 등 4대 추진과제를 중심으로 수립되었으며, 고속도로 신설 및 확장 등 신규 고속도로 사업을 포함하고 있다[1]. 본 계획에서 주목할 만한 내용은 경부고속도로 양재 IC 남쪽에 위치한 경기 화성과 서울 지역을 연결하는 구간을 지하화하는 것이다. 교통정체가 심한 구간인 만큼 기존 도로에 추가로 지하도로를 건설하여 도로를 확장하여 정체를 완화시킨다는 취지이다.

지하도로는 용지보상비, 주거지역의 소음, 분진, 진동 등의 민원발생, 녹지지역 환경훼손, 공사구간 교통정체, 악기상으로 인한 사고위험 등의 문제를 해결하기 위한 대안으로 수도권 및 대도시를 중심으로 지하도로 건설이 활발하게 추진되고 있다. 국외의 경우 교통혼잡 완화, 도로단절구간 연결, 친환경 공간 조성 등을 목적으로 2000년대에 지하도로가 도입되어 운영되고 있으며, 대표적인 운영사례로 프랑스 파리 A86 지하도로, 미국 보스턴 Big-Dig, 일본 동경 횡단도로 등이 있다. 국내의 경우에는 신월~여의 지하도로와 서부간선도로 지하도로가 개통되어 현재 운영 중에 있으며, 동부간선도로 지하화, 부산 만덕-센텀 지하도로, 사상-해운대 지하도로 등이 건설 또는 추진 중에 있다[2].

최근 수도권 지역에 개통되어 운영 중이거나 건설을 추진 중인 국내 지하도로는 지상도로와 지하도로의 연결을 위해 합류 또는 분류가 이루어지는 형태의 특징을 가지고 있다. 도심지를 통과하는 지하도로의 경우 서로 다른 교통류 흐름을 가지기 때문에 지상-지하 도로가 만나는 접속부와 지하 내(지하-지하 도로)에서 분·합류되어 만나는 연결로 접속부의 경우 서로 다른 교통류간에 상충이 발생하게 되어 도로의 전체적인 교통흐름에 영향을 줄 수 있다. 따라서 지하도로 신규 건설로 인한 분류부와 합류부의 연결로 접속부 구간에서의 교통흐름을 안전하게 유도할 수 있는 교통류 관리 방안에 대한 연구가 필요한 실정이다.

일반적으로 지하도로나 터널은 지상도로와는 달리 공간적으로 폐쇄적인 기하구조 특징을 가지고 있다. 이로 인해, 운전자는 주행 중 심리적 부담감이 상당히 높아지게 되므로 안전 확보와 사고예방을 위해서는 교통흐름을 원활하게 유지해야 할 필요가 있다. 지상도로의 경우 교통 사고 및 지·정체를 선제적으로 예방하기 위해 카메라, 영상검지기 등을 이용한 실시간 모니터링 관제시스템을 통해 속도제어가 가능하지만 지하도로의 경우 지상도로와 만나는 부분까지 고려를 해야하므로 지상도로와 차별화된 교통운영전략 수립이 필요하다. 도시 지역 지하도로 설계지침[3]에서는 지하도로 내 기하구조와 운영방안에 대한 내용을 제시하고는 있으나 교통운영 측면에서의 내용은 미흡한 실정이다. 또한, 터널의 경우에도 교통운

영을 위한 관리 지침과 일부 지자체의 교통통제 기준이 있으나 주로 악기상(강풍, 호우, 해무, 강설, 결빙 등)과 관련된 내용이 대부분이다[4].

도시부 도로의 교통문제를 해결하기 위해 지하도로 건설 및 개통이 증가함에 따라 지하도로의 특성을 고려한 교통운영전략 수립을 통해 안전한 교통흐름을 확보할 필요가 있다. 국내에서는 1961년 부산터널 첫 개통 이후 터널과 관련된 많은 연구들이 현재까지 활발하게 이어져 오고 있으나, 최근 국내에 첫 도입되어 운영 중인 지하도로의 경우 교통안전과 관련된 연구는 거의 찾아볼 수 없다. 따라서, 본 연구에서는 비정형 자료 분석 기법 중 하나인 텍스트 마이닝 기법을 이용하여 지하도로와 유사한 기하구조를 가지고 있는 터널을 대상으로 교통안전과 관련된 국내 저널의 연구동향 분석하였다. 이를 통해 과거 터널의 교통안전 측면에서 수행된 연구동향을 파악하여 지하도로의 교통안전 확보를 위해 필요한 연구 주제들을 도출하였다.

II. 선행연구

국내 도로교통 분야에서 텍스트 마이닝을 활용한 다양한 연구들이 수행되어 왔다.

오준석[5]은 동시출현단어 분석과 토픽 모델링을 활용하여 국내 교통 및 ICT 분야의 교통 관련 연구동향을 분석하고, 융합연구 수행을 위해 필요한 분야별 연구들을 제시하였다. 동시출현단어 기반 분석결과와의 경우 다양한 교통분야의 분석 모형들을 ICT와 접목시켜 정보기술 기반 교통 서비스들에 대한 데이터 분석 연구들을 주도적으로 수행할 필요가 있으며, 토픽 모델링 분석결과와 의 경우 교통수단 및 공공교통에 대한 운영 및 안전에 대한 융합연구를 수행할 필요가 있다는 결론을 제시하였다.

박진철 외[6]는 스마트시티에 대한 연구 동향을 파악하고 정책 방향성을 제시하기 위해 스마트시티와 관련된 학술논문 11,527건의 정보(제목과 초록, 발행연도 등)를 수집하여 연구주제, 연구분야 추이 등을 LDA기반 토픽 모델링 기법을 활용하여 분석하였다. 시민중심 스마트시티 추진을 통한 지속가능성의 확보와 관련된 주제가 가장 많이 언급된 것으로 나타났으며, 스마트시티 관련 연구주제로 서비스 및 애플리케이션 분야, 기술 분야, 시민 및 사용자 관점의 8가지 세부주제로 유형화하는 결론을 도출하였다.

이상규[7]는 건설 안전사고에 대한 웹 뉴스 기사를 수집하여 LDA 기반 토픽 모델링을 통해 건설 안전사고 주요 핵심 키워드에 대한 트렌드 분석을 수행하였다. 10가지 토픽과 키워드를 통해 주요 이슈를 도출하고 뉴스 데이터에 대한 시계열 분석을 통해 향후 토픽과 관련된 이슈를 예측하였다. 또한, 토픽별 키워드의 유의수준을 기반으로 통계 검증을 수행하였다.

임옥근[8]은 지난 10년간 국내에서 수행된 수소충전소 관련 논문들을 수집하여 현재까지 연구된 주제들에 대한 동향

을 분석하여 향후 필요한 연구들을 도출하였다. 이를 위해 문서내부의 단어들을 추출하는 텍스트마이닝 기법을 이용해 토픽과 키워드 간의 네트워크를 구성하여 분석하고, 토픽 모델링 기법 중 하나인 LDA 기법을 사용하여 연구 주제들을 분석하였다. 분석결과, 2017년 이후 수소충전소, 연료전지, 전기충전소 등이 핵심키워드로 등장했으며 수소저장, 수소자동차, 수소충전소 안전과 관련된 주제들에 대한 연구 비중이 증가하고 있는 것으로 나타났다.

박정태 외[9]는 토픽 모델링을 활용하여 스마트 모빌리티와 관련된 연구 동향을 파악하기 위해 KCI 학술지로부터 모빌리티 키워드가 포함된 논문을 수집하여 초록을 분석하였다. 국내 학술지에 대한 논문 초록만을 분석자료로 활용하여 결과 해석과 일반화에 제약이 따르며, 해외 연구 동향 또한 반영하지 못한 것을 한계점으로 제시하였다.

이와 같이, 텍스트 마이닝을 활용하여 스마트시티, 모빌리티 등과 관련된 주제의 연구동향을 파악하고 핵심 키워드를 유형화하는 연구는 활발히 진행되어 오고 있으나, 터널과 관련된 연구는 없는 실정이다.

III. 연구 방법론

3-1 텍스트 마이닝

텍스트 마이닝이란 데이터의 구조가 다양한 비정형 데이터(unstructured data) 중 하나인 텍스트 자료부터 의미 있는 정보를 추출하는 방법이다. 키워드 분석, 토픽 모델링, 감성분석 등 다양한 방법론이 존재하며, 이 중 토픽 모델링은 대량의 텍스트 자료를 이용하여 핵심 주제를 도출하기 위한 텍스트 마이닝의 주요 기법이다[10].

토픽 모델링은 문서를 구성하는 키워드를 기반으로 추출한 주제(Topic)를 기준으로 문서를 분류(Clustering)하고 분석한다. 문서에 내재된 잠재적 주제를 도출하여 새로운 문서의 주제에 대한 예측 및 분석이 가능하기 때문에 다양한 분야에서 비정형 텍스트 분석을 위해 많이 사용되고 있다. 확률적 모형 기반으로 구성된 주제별 키워드 분포를 바탕으로 주어진 문서로부터 발견된 키워드 분포를 분석함으로써 특정 문서에 사용된 주제들에 대한 추정이 가능하다. 토픽 모델링은 어떤 문서의 토픽이 여러 개의 추상적인 주제로 구성되어 있으며, 각각의 토픽은 여러 개의 단어 집합으로 구성되어 있다는 것을 전제하고 있다[6].

LDA는 토픽모델링에 이용되는 대표적인 알고리즘으로 토픽과 단어 간 관계에 대한 추정이 가능하다. 문헌 내의 잠재된 토픽을 추정하는 것으로 어떤 단어들을 어떠한 토픽에서 선택하여 추출할 것인지에 대해 각각의 파라미터(parameter; α, β)를 이용하여 유의미한 결과가 도출되도록 모델링하는 기법이다. LDA는 디리클레 확률 분포(Dirichlet

probability distribution)를 활용하여 문서의 잠재적인 확률을 검토한 후 깃스 샘플링(Gibbs sampling) 알고리즘을 통해 해당 문서의 토픽과 단어에 대한 확률값을 추정한다. LDA는 어떤 토픽과 단어에 대한 사전 정보가 없는 비지도 생성 모델(unsupervised generative model)이기 때문에 각 토픽과 관련된 모든 단어를 쉽게 찾을 수 있다는 장점을 가지고 있다. LDA는 Fig. 1과 같이 각 주제에서 도출할 수 있는 단어들의 확률 분포를 알고 있을 때, 무작위 과정(random process)을 통해 문서가 생성될 수 있으며, 하나의 문서는 여러 주제로 구성되고 문서의 주제 분포에 따라 단어의 분포가 결정된다는 가정을 내포하고 있다. 이를 통해 해당문서에 대한 전체 주제, 각 문서별 주제 비율과 각 주제에 포함될 단어들의 분포를 알아낼 수 있다[6]. 하이퍼 파라미터(Hyper parameter)인 문서내 토픽의 비율(θ)과 문서내 토픽의 단어 비율(β) 및 토픽의 개수(k)를 미리 입력하고 입력된 값을 통해 단어(w)간 비교를 통해 단어별로 적절한 문서집합의 주제(z)가 결정된다. 그리고 사용자가 정하는 반복횟수에 따라 분석을 반복수행하여 θ, β 값의 지속적인 갱신을 통해 모든 z 값 중 가장 높은 z 값을 찾아 해당 문서에 있는 각각의 단어들이 어디에 속하는지를 추정한다[11]. LDA 기법을 활용하여 토픽모델링을 수행할 경우 최종 토픽 모델링 결과는 설정된 α 값과 β 값의 영향을 많이 받는다. 따라서, LDA 기법을 활용한 토픽모델링의 경우 최적의 α 와 β 의 값을 찾아내는 것이 매우 중요하다. LDA는 문서, 단어 집합에서 문장이나 문서의 구조 등 보이지 않는 변수를 추론하는 방법으로 토픽에 따라 문서 집합을 재구성할 수 있으며, 데이터, 이미지 및 소셜 네트워크에서 특정한 패턴을 찾는 데 사용될 뿐 아니라 인터넷 리뷰, 신문기사, 논문 등의 동향을 파악하는 등 잠재적인 주제나 관련 동향을 파악하는데 유용하게 활용되고 있다[12].

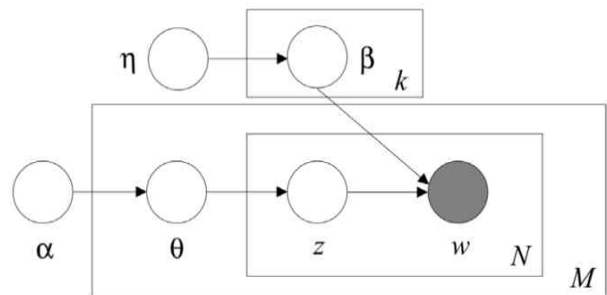


그림 1. LDA 그래프 모형
Fig. 1. Graphical model representation of LDA

- M : 문서의 개수
- N : 문서에 속한 단어의 개수
- W : 단어
- Z : 해당 단어가 속한 토픽번호
- k : 토픽의 개수 (Hyper parameter)
- α : 문서-토픽별 θ 분포값 (Hyper parameter)

- η : 문서-단어별 β 분포값 (Hyper parameter)
- θ : 문서별 토픽의 분포
- β : 토픽의 단어 분포

3-2 자료수집

본 연구에서는 자료수집을 위해 NetMiner에서 제공하는 Biblio Data Collector를 이용하여 논문자료를 수집하였다. Biblio Data Collector는 연구동향 분석에 필수적인 학술논문 서지정보를 온라인으로 부터 실시간 수집이 가능하다. 따라서, API를 이용하여 국내 KCI(한국학술지인용색인)에 등재된 논문 데이터 중 제목과 초록, 발행연도 정보 등의 서지정보를 수집하였다. “터널”과 “교통안전” 키워드만으로는 검색되는 논문 수가 너무 적어 키워드를 “터널”, “교통안전”, “교통류”, “속도”, “사고”, “위험”으로 확장하였다. 2000년부터 현재(2022년 7월말) 까지 약 20년 동안 총 567편의 관련 논문이 수집되었다. 이중 전처리 과정을 통해 키워드와 관련성이 없거나 중복된 논문 453편을 제외하고 총 114편의 논문을 분석에 활용하였다.

연도별로 국내 저널에 게재된 연구논문의 시계열 흐름은 Table 1과 같다. 2000년을 시작으로 현재까지 매년 국내 저널에 게재된 논문이 증가하는 것으로 분석되었으며, 2011년 이후부터 많은 연구들이 수행된 것을 알 수 있다. 이는 실제로 2000년부터 터널 연장 및 개소가 현재까지 지속적으로 증가함에 따라 관련 연구도 활발하게 진행된 것을 보여준다.

3-3 분석방법

본 연구에서는 터널의 교통안전과 관련된 연구주제에 대한 동향을 분석하기 위해 NetMiner 4.5를 이용하여 키워드 빈도 분석과 토픽모델링을 수행하였다. NetMiner는 SNA(Social Network Analysis) 전문 소프트웨어로써 계량통계, 데이터 마이닝 등 분석이 가능하여 다양한 분야에서 많이 활용되고 있다.

표 1. 자료수집

Table 1. Data Collectoin

Data	Contents
Period	2000.01.01.~2022.07.31
Keywords	“Tunnel”, “Traffic safety”, “Traffic flow”, “Speed”, “Accident”, “Danger”
bibliographic Data	Title, Abstract, Keywords, Authors, PubDate, Journal, Publisher
# of Papers	567 Papers

표 2. 연도별 논문게재 현황

Table 2. Paper publications by Year

Year	# of Papers
2000 -2005	10
2006-2010	20
2011-2015	34
2016-2022	50
Total	114

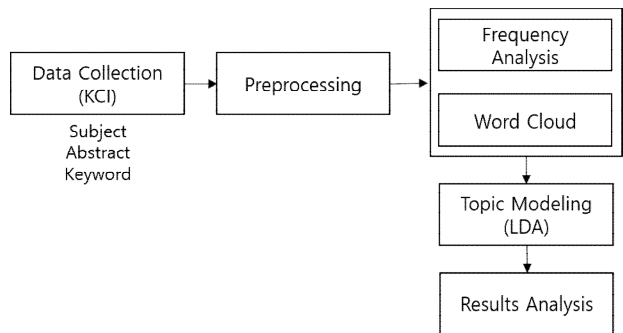


그림 2. 연구 분석절차

Fig. 2. Data Analysis Framework

국내 터널 교통안전 관련 연구 동향을 파악하기 위해 국내 KCI 논문 567편에 대한 제목, 초록, 키워드 등을 수집하여 문서집합을 생성한 후 토픽 모델링의 분석기법 중 LDA를 활용하여 분석하였다.

LDA 분석을 위해서는 수집된 논문의 비정형 데이터로부터 의미 있는 주제를 도출하기 위해 형태소 분석을 통한 전처리 과정이 선행되어야 한다. 형태소 분석을 위해 NetMiner의 BIBLIO Data Collector를 이용하였다. 문서에서 자주 등장하는 어간, 어미, 조사 등 분석에 불필요한 불용어(Stopwords)를 제거하고 동일 의미를 가진 단어를 하나의 단어로 변환하는 어간추출(Stemming) 작업을 수행하였다. 형태소 분석 후 벡터화(Vectorization)를 통해 분석된 문서로부터 명사를 추출하여 문서집합을 생성하였다. 벡터화는 TF-IDF(Term Frequency-Inverse Document Frequency)에 기반한 벡터형식의 코퍼스(Corpus)를 통해 명사를 추출하여 전체 문서집합을 토픽모델링이 가능한 형식으로 변환하는 과정이다[13].

위와 같이 구축된 문서집합을 기반으로 핵심 토픽에 대한 키워드 빈도 분석과 LDA 기반 토픽모델링을 통해 터널 교통안전과 관련된 연구주제들을 도출하였다. 본 논문의 연구 분석절차는 Fig. 2와 같다.

IV. 분석결과

4-1 키워드 빈도 분석

전처리 과정을 통해 도출된 총 1,306개의 명사 노드를 대상으로 키워드 빈도 분석을 수행하였다. 터널 교통안전과 관련하여 가장 많이 등장한 상위 20개의 핵심 키워드는 Fig. 3과 같다. “사고”, “운전자”, “시스템”, “주행”, “차량”, “안전”과 관련된 키워드가 100회 이상 등장하여 이와 관련된 연구가 가장 많이 수행된 것을 알 수 있다. 또한, “위험”, “조명”, “이미지”, “시간”, “설계”, “속도”, “정보”, “시설”, “화재” 등의 키워드도 50회 이상 등장하는 것으로 분석되었으며, “관리”, “운영”, “유고”와 관련된 키워드도 40회 이상 등장하는 것으로 나타났다. 가장 많이 언급된 키워드는 교통사고와 관련된 키워드로 총 183회가 언급되었으며, 차이는 미미하지만 “유고” 관련 키워드가 가장 적게 언급된 것으로 나타났다.

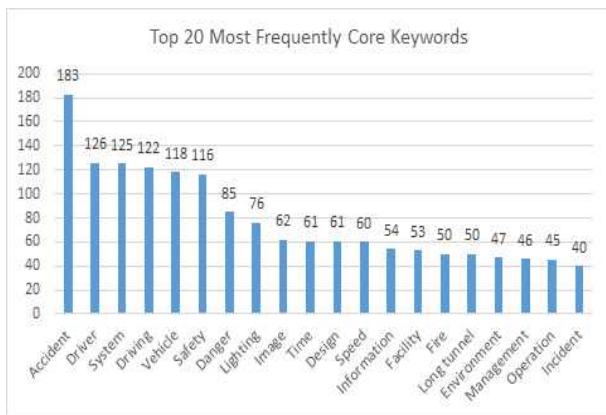


그림 3. 상위 20개 핵심 키워드
Fig. 3. Top 20 Most Frequently Core Keywords

워드 클라우드(word cloud)는 도출된 핵심 키워드를 기반으로 키워드의 중요도나 빈도 등을 잘 구분할 수 있도록 시각화 하는 방법이다. 시각화를 이용하여 중요하거나 자주 등장한 단어는 크게 강조되고, 상대적으로 중요하지 않은 단어는 작게 표시된다. 전처리 과정을 통해 도출된 단어를 기반으로 단어의 중요도를 확인하기 위해 TF-IDF 값으로 대표 키워드를 추출하여 워드 클라우드를 Fig. 4와 같이 나타내었다. TF-IDF는 어떠한 단어가 해당 문서 내에서 얼마나 중요한 위치를 차지하는지에 대한 정도를 나타내는 통계적 지표이다 [14]. TF(Term Frequency)는 특정 단어가 문서 내에서 얼마나 자주 등장하는지에 대한 값으로 단어 빈도를 나타내고, IDF(Inverse Document Frequency)는 특정 단어가 몇 개의 문서에서 등장하는지를 측정하는 방법으로 문서 빈도를 의미한다. 따라서, TF-IDF는 단어 빈도와 문서 빈도를 기반으로 해당 단어가 특정 문서에서 얼마나 중요한지에 대한 파악이 가능하다. Fig. 4에서와 같이 문서 내에서 사고, 운전자, 시스템, 차량, 안전, 주행 등과 같은 키워드의 중요도가 높은 것을 알 수 있다. 또한, 장대터널, 영상, 조명, 환경, 시설 등과 같은 키워드도 높은 비중의 중요도를 차지하는 것으로 나타났다.



* This figure is expressed only in Korean as a result of analysis by NetMiner

그림 4. 핵심키워드의 중요도에 따른 워드클라우드
Fig. 4. Word-cloud for the importance of core keywords

4-2 토픽 모델링

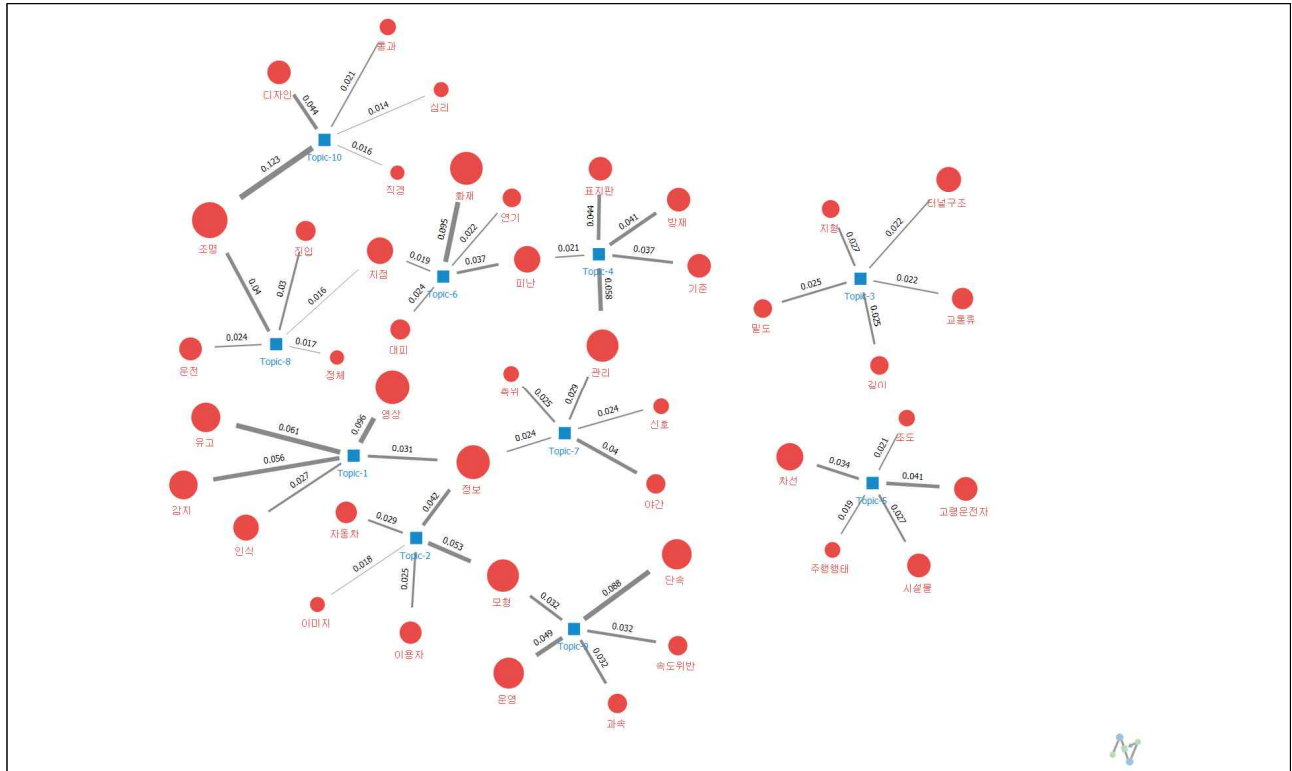
토픽 모델링은 각 단어들이 통계적으로 특정 토픽에 포함될 확률을 파악하여 문서의 핵심 주제를 추정하는 기법으로 텍스트 마이닝 분석에서 많이 사용되는 방법이다[14]. 본 연구에서는 LDA(Latent Dirichlet Allocation) 토픽 모델링 기법을 통해 핵심 토픽을 추출하였다. LDA 분석은 문서나 문헌 텍스트의 토픽과 토픽별 주제어의 생성 확률에 기초한 분석 기법이기 때문에 텍스트가 잠재적으로 가지고 있는 토픽의 개수와 그 토픽을 대표할 수 있는 토픽의 개수를 연구자가 임의로 지정하여 분석해야 한다. 토픽을 지정할 때는 여러 번의 사전 테스트를 통해 적절한 토픽의 개수를 설정하는 것이 중요하다[15]. 앞서 설명한 바와 같이 연구자가 직접 토픽의 수, 알파(alpha), 베타(beta) 값 등과 같은 하이퍼 파라미터(Hyper Parameter)값들에 대한 설정이 필요하다. W. Zhao(2015)은 선행 연구를 통해 알파 값은 0.1~0.01, 베타 값은 0.01을 최적값으로 제시하였다[16]. 따라서 본 연구에서는 토픽 수를 5개~10개까지 적용하고 알파 값을 0.03, 베타 값을 0.01로 설정한 후 샘플링 분석을 1,000회 반복 수행하여 결과를 도출하였다.

터널의 교통안전과 관련된 114편의 논문에 대한 LDA 토픽 모델링 분석결과는 Table 3과 같다. 토픽별로 도출된 주요 핵심 키워드를 살펴보면 토픽 1은 영상, 유고, 감지 관련 주제, 토픽 2는 모형, 정보, 자동차 관련 주제, 토픽 3은 지형, 밀도, 교통류 관련 주제, 토픽 4는 관리, 표지판, 방재 관련 주제, 토픽 5는 고령운전자, 차선, 주행행태 관련 주제, 토픽 6은 화재, 피난, 대피 관련 주제, 토픽 7은 야간, 측위, 신호 관련 주제, 토픽 8은 조명, 진입부, 운전 관련 주제, 토픽 9는 단속, 속도위반, 과속 관련 주제, 토픽 10은 조명, 디자인, 심리 관련 주제로 도출되었다. 따라서 위의 토픽 모델링 결과를 기반으로 도출된 10개의 주제에 대해 토픽 1 - 유고, 토픽 2 - 도로정보, 토픽 3 - 교통 지정체, 토픽 4 - 방재, 토픽 5 - 고령운전자, 토픽 6 - 화재, 토픽 7 - 측위, 토픽 8 - 조명, 토픽 9 - 과속, 토픽 10 - 운전자 심리와 같이 토픽명을 유형화 하였다.

표 3. LDA 토픽 모델링 분석 결과

Table 3. LDA topic modeling result

Topic	1 st		2 st		3 st		4 st		5 st	
	Keyword	Prob.	Keyword	Prob.	Keyword	Prob.	Keyword	Prob.	Keyword	Prob.
Topic 1 (Incident)	Video	0.096	Incident	0.061	Detecting	0.056	Information	0.031	Recognition	0.027
Topic 2 (Information)	Lighting	0.053	Information	0.042	Vehicle	0.029	Driver	0.025	Image	0.018
Topic 3 (Delay)	Terrain	0.027	Length	0.025	Density	0.025	Tunnel structure	0.022	Traffic flow	0.022
Topic 4 (Disaster prevention)	Management	0.058	Traffic sign	0.044	Disaster prevention	0.041	Manual	0.037	Evacuation	0.021
Topic 5 (Older driver)	Older driver	0.041	Lane	0.034	Facility	0.027	Illuminance	0.021	Driving behavior	0.019
Topic 6 (Fire)	Fire	0.095	Evacuation	0.037	Evacuation	0.024	Smoking	0.022	Point	0.019
Topic 7 (Indoor positioning)	Night	0.040	Management	0.029	Indoor positioning	0.025	Information	0.024	Signal	0.024
Topic 8 (Lighting)	Lighting	0.040	Tunnel entrance	0.030	Driving	0.024	Delay	0.017	Point	0.016
Topic 9 (Speeding)	Assessment	0.088	Operation	0.049	Speed enforcement	0.032	Model	0.032	Speeding	0.032
Topic 10 (Psychology)	Lighting	0.123	Design	0.044	Through Traffic	0.021	Diameter	0.016	Psychology	0.014



* This figure is expressed only in Korean as a result of analysis by NetMiner

그림 5. 토픽별 키워드 네트워크

Fig. 5. keyword network by topic

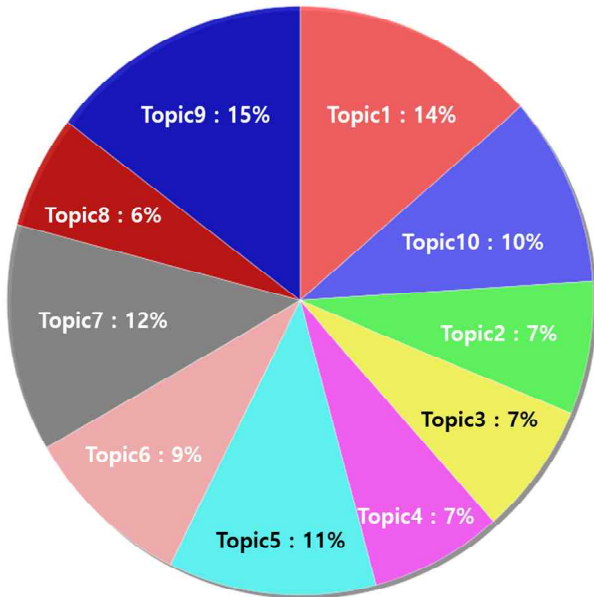


그림 6. 토픽별 점유율
Fig. 6. Proportion of topics

Table 3에서 단어 출현 확률(Probability)은 문서 내에서 해당 단어가 출현할 확률을 의미한다.

10개 토픽에 대한 비중을 살펴보면 Fig. 6과 같이 토픽 9 - 과속 (15%), 토픽 1 - 유고 (14%, 토픽 7 - 측위 (12%), 토픽 5 - 고령운전자 (11%), 토픽 10 - 운전자 심리 (10%), 토픽 6 - 화재 (9%), 토픽 2 - 도로정보 (7%), 토픽 3 - 교통지정체 (7%), 토픽 4 - 방재 (7%), 토픽 8 - 조명 (6%) 순으로 비중이 높은 것으로 나타났다.

Fig. 5는 10개의 토픽과 그에 따른 키워드들 간의 연관성을 Kamada & Kawai 알고리즘을 이용하여 네트워크로 시각화한 그림이다. 사전에 모든 단어노드 사이의 최단 경로거리를 구해야 하므로 다소 시간이 많이 걸리는 단점을 가지고 있지만, 노드들 간의 중복이나 링크들 간의 교차가 적어 가독성이 좋은 특징이 있다.

V. 결 론

서울과 수도권, 부산 등을 중심으로 지하도로 건설이 지속적으로 증가하고 있다. 현재 기존 도로를 확장할 수 있는 공간은 한계가 있기 때문에 공간의 효율성, 도로의 친환경을 위해서 지하도로는 지상도로의 대안이 될 수 밖에 없을 것이다.

본 논문에서는 과거 약 20년 동안 국내 저널에 게재된 터널의 교통안전과 관련된 연구동향 분석을 통해 향후 지하도로 운영시 교통안전 확보를 위해 필요한 연구주제를 도출하였다. 비정형 자료 분석에 주로 사용되는 텍스트마이닝 기법 중 토픽 모델링을 사용하여 연구동향을 분석하고 도출된 토픽별 키워드를 기반으로 토픽명을 부여하였다.

터널의 교통안전 관련 연구는 과속, 유고, 측위, 고령운전자, 운전자 심리와 연관성이 있는 주제가 상위 10% 이상 높은 비중을 차지하는 것으로 분석되었다. 그 뒤로 화재, 도로정보, 교통지정체, 방재, 조명과 관련해서도 많은 연구가 수행된 것으로 나타났다.

10개의 토픽별 주제 중심의 연관어에 대한 상관관계 분석 결과를 살펴보면, 과속 토픽의 경우 단속, 운영, 속도위반, 모형, 과속과 관련된 주제의 연구가 많이 수행된 것으로 나타났다. 유고 토픽은 영상, 유고, 감지, 정보, 인신과 관련된 주제의 연구가 많이 수행된 것으로 나타났으며, 측위 토픽은 야간, 관리, 측위, 정보, 신호와 관련된 주제의 연구가 많이 수행된 것으로 나타났다. 고령운전자 토픽의 경우 고령운전자, 차선, 시설물, 속도, 주행행태와 관련된 주제의 연구가 많이 수행되었으며, 운전자 심리 토픽의 경우 조명, 디자인, 심리와 관련된 주제의 연구가 많이 수행된 것으로 분석되었다. 또한, 화재 토픽의 경우 화재, 피난, 대피, 연기, 지점과 관련된 주제의 연구가 주로 수행되었으며, 도로정보 토픽의 경우 모형, 정보, 자동차, 이용자, 이미지와 관련된 연구가 많이 수행되었다. 교통지정체 토픽의 경우 지형, 길이, 밀도, 터널구조, 교통류와 관련된 주제의 연구가 주로 수행되었고, 방재 토픽의 경우 관리, 표지판, 방재, 기준, 피난과 관련된 주제의 연구가 많이 수행된 것으로 나타났다. 마지막으로 조명 토픽은 조명, 진입, 운전, 정체, 지점과 관련된 주제의 연구가 많이 이루어진 것으로 분석되었다.

분석결과를 기반으로 향후 교통측면에서 지하도로의 교통안전 확보를 위해서는 다음과 같은 연구가 필요한 것으로 판단된다. 첫째, 지하도로 내에서 안전속도를 유도할 수 있는 속도저감 및 교통지정체 관련 연구가 필요할 것이다. 둘째, 일반운전자와 고령운전자의 안전 주행 확보를 위한 주행행태 및 운전자 심리와 관련된 연구가 필요할 것이다. 셋째, GPS가 원활하지 않은 지하도로 내에서 돌발상황(화재, 고장, 사고 등) 발생시 이에 대한 정보를 신속하게 운전자 및 도로관리자에게 제공할 수 있는 측위기술, 도로정보 및 방재에 대한 연구가 필요할 것이다. 마지막으로, 지하도로 환경에서 운전자가 심리적 부담감을 가지지 않고 편안하게 주행할 수 있는 환경을 조성하기 위한 조명 관련 연구 또한 필요할 것이다.

향후 지하도로가 지속적으로 확장되어 운영될 경우 진출입부나 지하도로와 연결되는 주변 도로 등에서 발생할 수 있는 교통지정체와 다양한 돌발상황을 선제적으로 해결하여 교통사고 위험을 감소시킬 수 있는 다양한 관련 연구가 필요할 것이다. 따라서, 본 연구는 텍스트 마이닝을 통해 지하도로의 교통흐름을 안전하게 유도할 수 있는 교통안전 관련 연구의 방향성을 제시하였다는 데 의의가 있다. 향후 지하도로 운영에 있어 교통안전 확보를 위한 연구주제의 방향성을 설정하는데 기초자료로 활용될 수 있을 것으로 판단된다.

하지만 본 연구는 국내 KCI 학술논문 DB만을 수집하여 분석을 하였다는 한계점을 가지고 있다. 향후 국외 학술논문 DB 까지 확장하여 분석을 수행한다면 더 유의미한 결과를 도출할 수 있을 것으로 생각된다.

감사의 글

본 연구는 2022년도 정부(국토교통부) 재원으로 국토교통과학기술진흥원의 지원을 받아 수행되었습니다. (22UUTI-C157786-03, 도심 지하 교통 인프라 건설 및 운영 기술 고도화 연구)

참고문헌

- [1] Ministry of Land, Infrastructure and Transport(MOLIT), "2st Highway Construction Plan," *Press Release*, 2022
- [2] J. G. Kim, G. W. Lee, C. H. Yang, "Prioritizing efficient Traffic Control Strategies in the On-off Ramp Junction of Urban Deep Underground Tunnels," *International Journal of Highway Engineering*, Vol. 23, No. 4, pp. 125-132, August 2021. <https://doi.org/10.7855/IJHE.2021.23.4.125>
- [3] Ministry of Land, Infrastructure and Transport(MOLIT), "Urban Underground Design Guideline," 2016.
- [4] S. B. Park, J. P. Moon, J. G. Kim, C. H. Yang, "Determining possibility of the occurrence of traffic Congestion on underground rodas," *International Journal of Highway Engineering*, Vol. 24, No. 4, pp. 55-65, August 2022. <https://doi.org/10.7855/IJHE.2022.24.4.055>
- [5] J. S. Oh, "Identifying Research Opportunities in the Convergence of Transportation and ICT Using Text Mining Techniques," *Journal of Transport Research*, Vol. 22, No. 4, pp. 93-110, December 2015. <https://doi.org/10.34143/jtr.2015.22.4.93>
- [6] K. C. Park, C. H. Lee, "A Study on the Research Trends for Smart City using Topic Modeling," *Journal of Internet Computing and Services*, Vol. 20, No. 3, pp. 119-128, June 2019. <https://doi.org/10.7472/jksii.2019.20.3.119>
- [7] S. G. Lee, "A Study on the Trends of Construction Safety Accident in Unstructured Text Using Topic Modeling," *Journal of the Korea Academia-Industrial cooperation Society*, Vol. 19, No. 10, pp. 176-182, October 2018. <https://doi.org/10.5762/KAIS.2018.19.10.176>
- [8] O. K. Lim, "Keyword Network Analysis of Domestic Research Articles for Determining Recent Trends of Hydrogen Refueling Stations," *Fire Science and Engineering*, Vol. 35, No. 4, pp. 83-90, July 2021. <https://doi.org/10.7731/KIFSE.0565d609>
- [9] J. T. Park, C. Y. Kim, T. J. Kim, "An Analysis of the Research Trend on Smart Mobility : Topic Modeling Approach," *Journal of Korea Institute of Intelligent Transport Systems*, Vol. 21, No. 2, pp. 85-100, April 2022. <https://doi.org/10.12815/kits.2022.21.2.85>
- [10] S. M. Park, H. P. Lee, J. H. So, I. S. Yun, "Study of Analysis for Autonomous Vehicle Collision Using Text Embedding," *Journal of Korea Institute of Intelligent Transport Systems*, Vol. 20, No. 1, pp. 160-173, February 2021. <https://doi.org/10.12815/kits.2021.20.1.160>
- [11] C. W. Woo, J. Y. Lee, "Investigation of Research Topic and Trends of National ICT Research-Development Using the LDA Model," *Journal of the Korea Convergence Society*, Vol. 11, No. 7, pp. 9-18, July 2020. <https://doi.org/10.15207/JKCS.2020.11.7.009>
- [12] H. C. Yang, "Topic Modeling Analysis of Franchise Research Trends Using LDA Algorithm," *Korean Journal of Franchise Management*, Vol. 12, No. 4, pp. 13-23, December 2021. <http://dx.doi.org/10.21871/KJFM.2021.12.5.13>
- [13] J. H. Park, H. J. Oh, "Comparison of Topic Modeling Methods for Analyzing Research Trends of Archives Management in Korea: focused on LDA and HDP," *Journal of Korean Library and Information Science Society*, Vol. 48, No. 4, pp. 235-258, December 2017. <http://dx.doi.org/10.16981/kliss.48.4.201712.235>
- [14] CYRAM, "NetMiner Semantic Network Analysis," *NetMiner 4 Manual*, 2022
- [15] S. I. Hwang, M. K. Kim, "An Analysis of Artificial Intelligence(A.I.) related Studies' Trends in Korea Focused on Topic Modeling and Semantic Network Analysis," *Journal of Digital Contents Society*, Vol. 20, No. 9, pp. 1847-1855, September 2019. <http://dx.doi.org/10.9728/dcs.2019.20.9.1847>
- [16] W. Zhao, J.Chen, W. Zen, "Best Practices in Building Topic Models with LDA for Mining Regulatory Textual Documents," *CDER 9th*, November 2015.



김진국(Jinguk Kim)

2006년 : 관동대학교 교통공학 석사
2004년 : 강릉대학교 컴퓨터공학 학사
2021년 : 한양대학교 교통·물류공학 박사수료

2015~현재 : 한국건설기술연구원 도로교통연구본부 전임연구원
※ 관심분야 : 도로안전, 도로관리, 도로정보, 지하도로



양충헌(Choongheon Yang)

2008년 : Ph.D. in Civil Engineering at Univ. of California, Irvine
2002년 : 명지대학교 교통공학 석사
2000년 : 명지대학교 교통공학 학사

2009년~현재 : 한국건설기술연구원 도로교통연구본부 연구위원
※ 관심분야 : 도로안전, 도로관리, 지하도로



박수빈(Subin Park)

2022년 : 한양대학교 교통공학 박사
2016년 : 한양대학교 교통공학 학사

2022년~현재 : 한국건설기술연구회 도로교통연구본부 박사후연구원
※ 관심분야 : 도로안전, 지하도로, 교통운영, 스마트모빌리티