Journal of Digital Convergence
Vol. 19. No. 2, pp. 1-12, 2021

ISSN 2713-6434 / eISSN 2713-6442
https://doi.org/10.14400/JDC.2021.19.2.001

토픽 모델링 기반의 국내외 공공데이터 연구 동향 비교 분석

박대영[1], 김덕현[2], 김건욱[3*]
[1]영남대학교 경영학과 학부과정, [2]경북대학교 통계학과 학부과정,
[3]대구디지털산업진흥원 빅데이터활용센터 센터장

Topic Modeling-Based Domestic and Foreign Public Data Research Trends Comparative Analysis

Dae-Yeong Park[1], Deok-Hyeon Kim[2], Keun-Wook Kim[3*]
[1]Undergraduate, School of Business, Yeungnam University,
[2]Undergraduate, Department of Statistics, Kyungpook University,
[3]Director, Big Data Center, Daegu Digital Industry Promotion Agency

요 약 최근 4차 산업혁명으로 빅데이터의 성장과 가치는 지속적으로 증가하고 있으며, 정부에서도 공공데이터 개방과 활용에 적극적으로 노력하고 있다. 하지만 여전히 시민들의 공공데이터 활용 요구수준에는 미치지 못하는 상황이며, 현 시점에서 공공데이터 분야의 연구동향 파악과 발전 방향을 모색할 필요가 있다. 이에 본 연구에서는 공공데이터와 관련된 연구 동향을 파악하기 위해서 텍스트 마이닝 기법에서 주로 활용되는 토픽 모델링을 활용하여 분석하였다. 이를 위해 국내외 학술논문 중 '공공데이터', 'Public Data'의 키워드가 포함된 논문(국내 1,437건, 국외 9,607건)을 수집하여 LDA 알고리즘 기반의 토픽 모델링을 수행하였으며, 국내외 공공데이터 연구 동향을 비교 분석하여 정책적 시사점을 제시하였다. 분석 결과 국내의 경우 공공분야 정책 연구가 주를 이루고 있으며, 국외는 의료, 건강 관련 연구가 높게 나타났다. 토픽별 시계열로 살펴보면 국내는 '개인정보보호', '공공데이터 관리', '도시 환경' 분야의 연구가 증가하였으며, 국외는 '도시정책', '세포 생물학', '딥러닝', '클라우드·보안' 분야 연구가 활성화되고 있음을 확인할 수 있었다.

주제어 : 공공데이터, 연구 동향 분석, 텍스트 분석, 토픽 모델링, LDA

Abstract With the recent 4th Industrial Revolution, the growth and value of big data are continuously increasing, and the government is also actively making efforts to open and utilize public data. However, the situation still does not reach the level of demand for public data use by citizens, At this point, it is necessary to identify research trends in the public data field and seek directions for development. In this study, in order to understand the research trends related to public data, the analysis was performed using topic modeling, which is mainly used in text mining techniques. To this end, we collected papers containing keywords of 'Public data' among domestic and foreign research papers (1,437 domestically, 9,607 overseas) and performed topic modeling based on the LDA algorithm, and compared domestic and foreign public data research trends. After analysis, policy implications were presented. Looking at the time series by topic, research in the fields of 'personal information protection', 'public data management', and 'urban environment' has increased in Korea. Overseas, it was confirmed that research in the fields of 'urban policy', 'cell biology', 'deep learning', and 'cloud · security' is active.

Key Words : Public data, Research Trend, Text mining, Topic Modeling, LDA

*Corresponding Author : Keun-Wook Kim(aut7767@dip.or.kr)
Received November 12, 2020 Revised December 10, 2020
Accepted February 20, 2021 Published February 28, 2021

1. 서론

4차 산업혁명으로 빅데이터가 IT 트렌드의 중심으로 부상하고 있다. Gartner는 빅데이터를 '21세기 원유'라 할 정도로 전 세계의 국가들이 빅데이터의 성장과 가치에 주목하고 있다.[1] 이러한 변화 속에서 빅데이터, 인공지능 등 기술적 수요와 중요성은 날로 높아지고 있으며, 특히 빅데이터 분야는 공공과 산업 등 모든 분야에서 걸쳐 혁신적인 변화를 일으키고 있다.

정부에서도 이러한 시대적 흐름을 반영하기 위해 빅데이터 전문가 충원, 공공행정 사례 발굴, 공공데이터 개방 등의 다양한 정책 수립을 진행하고 있으며, 데이터 개방 확산을 위해 2013년 공공데이터법 제정 이후 공공데이터 포털을 운영하여 적극적으로 공공데이터 개방을 위한 노력을 하고 있다. 그 외 공공데이터 기반의 창업 센터인 오픈스퀘어-D, 데이터 품질 제고 등의 노력으로 2019년 OECD에서 실시한 디지털 정부 평가에서 공공데이터 개방 1위를 차지하는 성과를 이루며 디지털 정부로의 전환에 박차를 가하고 있다.[2]

하지만 공공데이터 개방과 활성화의 높은 정량적 성과에도 불구하고 여전히 시민들의 공공데이터 활용 및 요구수준에는 미치지 못하는 상황이며, 시민들의 눈높이에 맞게 지속적인 공공데이터 이용 활성화를 위해서는 현시점에서 공공데이터에 관한 연구 동향 파악과 발전 방향을 모색할 필요가 있다.

이에 본 연구에서는 국내외 학술지 중 '공공데이터', 'Public Data'의 키워드가 포함된 논문들을 수집하여 이를 텍스트 분석에서 토픽 모델링으로 주로 활용되는 LDA(Latent Dirichlet Allocation) 기법을 적용하여 분석하였다. 이를 통해 그동안의 공공데이터를 활용한 연구 동향을 파악하고 국내외 시계열적 비교 분석으로 공공데이터의 발전 방향과 정책적 시사점을 제시하고자 한다.

2. 선행 연구 분석

2.1 공공데이터를 활용한 선행 연구

국내 공공데이터 포털(KR), 서울 열린데이터 광장, 국외 공공데이터 포털(US, UK)의 자료를 활용하여 이용자 현황을 분석하였고, 포털의 이용자 서비스 유형과 시민의 참여 정보를 고려하여 공공데이터 이용 활성화 개선방안을 제시하였다.[3]

공공데이터 이용 의도에 미치는 영향을 파악하기 위해 설문조사 자료를 통해 회귀 분석하였다. 회귀 분석 결과 공공데이터 이용 의도 영향이 있는 네 개의 유의미한 독립변수(시스템 및 정보 품질, 정보 보안, 혁신성)를 도출하였다.[4]

공공데이터 포털과 국가중점 데이터의 자료를 활용하여 우리나라의 공공데이터 개방실태를 지표를 통해 평가하였다. 가용성, 사용 용이성, 활용도 측면에서 추세 변화와 기관별 현황을 제시하여 공공데이터 개방 정책에 있어 맞춤형 개방전략을 제언하였다.[2]

공공데이터 수요를 정확히 파악하기 위해 공공데이터 제공 신청 데이터를 통해 Held-Out likelihood 지표를 기준으로 토픽 모델링을 실시하였다. 공공데이터 제공 신청의 대규모 텍스트 데이터를 토픽 모델링으로 토픽별 비중, 연도별 추이 분석을 파악하였으며, 이를 통해 공공데이터 개방 정책에 대해 제언하였다.[5]

국외에서도 공공데이터에 대한 중요성이 증가함에 따라 OECD 회원국 중심으로 공공데이터에 관한 연구가 활발하게 진행되고 있다. 대부분 국외 연구는 공공데이터 개방 및 이용 활성화에 초점을 맞추고 있다.

공공데이터 포털에 대한 직접적 평가를 실시하여 전체 데이터 중 소수만이 실제로 활용 가능했기에 개선이 필요하다는 실증 연구 결과를 제시하였으며, 공공데이터를 활성화를 위해서 시민들의 참여가 중요하며, 입소문의 평가가 중요하다는 실증 연구를 실시하였다.[6,7]

이상의 선행 연구들은 공공데이터 활성화 방안 및 기반마련을 위한 정책적 제언을 한 장점은 있으나, 자료의 범위가 설문조사, 공공데이터 포털 데이터로 한정되어 있고, 토픽모델링을 활용한 공공데이터 연구 동향 분석 사례는 없었다. 따라서 본 연구에서는 '07년부터 '20년까지의 국내외 공공데이터 활용 논문을 수집하여 연구 동향을 분석하였으며, LDA 기반의 토픽 모델링을 활용하여 국내외 시계열 비교 분석을 진행한 점이 선행 연구와의 차별성을 지닌다.

Table 1. Public data previous study

Author (Year)	Purpose of study	Analysis materials		Method of analysis
		material	period	
D.G. Kim (2014)	Classification System Design for Public Sector Information	Open Data Portal (KR US, UK),	2013~2014	-
Y.I. Cha (2017)	The Influence on public data usage in private business sectors	Question investigation	-	multiple regression analysis

S.O. Yun (2019)	Activating the Public Data Sector in Korea by Comparing Public Data Opening Status by Country	Open Data Portal	2019.03	descriptive statistical analysis
S.B. Cho (2020)	A Study on the Direction of Open Policy through the Analysis of Public Data Demand through Structural Topic Modeling	Open Data Portal	2017~ 2019.03	Topic modeling (STM)
Wang (2017)	A Study on the Status of Public Data Portal in the UK and Suggestions for Improvement	UK Open Data Portal	2012~ 2017	multiple regression analysis
Bernd W (2017)	An Empirical Study on the Revitalization of Public Data Portal in Germany	Question investigation	2017	multiple regression analysis

2.2 연구 동향 분석 선행 연구

텍스트 분석에서 연구 동향 분석을 수행하기 위한 대표적인 알고리즘으로는 LDA가 있다. 이는 해당 문헌에서 잠재적 의미가 있는 키워드를 추출하기 위한 확률 분포 모델이다.[5]

공간 데이터 관련 연구 동향을 파악하기 위해 Scopus DB에서 1971년부터 2014년까지 관련 논문 1,621건을 수집하여 토픽 모델링을 실시하였으며, 해당 논문들의 출현 빈도 추세, 게재된 저널, 연구 분야, 연구기관, 국가, 피인용 상위 논문을 통해 공간 데이터 연구 동향을 분석하였다.[8]

창업 관련 연구 동향을 파악하기 위해 RISS에 등재된 석·박사 논문 중 1974년부터 2017년까지 관련 논문 1,010건을 수집하여 토픽 모델링을 실시하였으며, 로그우도(Log-Likelihood)기법을 통해 적정 토픽 수를 34개로 설정하였고, 토픽별 키워드 분석을 시행하였다.[9]

인공지능(A.I) 기술 관련 연구 활동 및 동향 분석을 파악하기 위해 SCI 저널에서 2000년부터 2017년까지 인공지능 분야와 관련된 논문 7,978건에 대해 토픽 모델링 기법을 활용하고 군집화를 실시하여, 인공지능 관련 기술 동향 및 주요 연구 활동 추이를 파악하였다.[10]

스마트 시티에 대한 연구 동향을 파악하기 위해 Scopus DB 및 Springer DB에서 2008년부터 2019년까지의 스마트시트와 관련된 학술논문 11,527건을 토픽 모델링 기법을 활용하고 기술통계와 다차원척도법을 활용하였다. 스마트 시티 연구현황에 대한 정량적 분석을 수행하여, 그 결과 스마트 시티 개념에 대한 발전 방향을 이해하고 정책 시사점을 제언하였다.[11]

4차 산업 관련 연구 동향 분석을 파악하기 위해 KCI에서 2016년부터 2019년까지 관련 논문 685건을 수집한 후 Perplexity[1] 비교를 통해 최적의 토픽 개수를 결정하여 토픽 모델링을 수행하였다. 또한 IDM 분석과 트렌드 분석을 통해 4차 산업 관련 연구 동향을 파악하였다.[12]

이상의 선행 연구들은 각기 다른 키워드를 활용하여 LDA 기반의 토픽 모델링을 실시하였으며, 적정 토픽의 수를 결정하는 데 있어 Perplexity, Log-Likelihood와 같은 정량적인 기법을 적용하여 토픽의 수를 7~19개로 분류하여 분석하였다. 이는 비록 통일된 분석 방법론을 사용하여 연구 동향을 분석하여 정책적 시사점을 도출한 장점은 있으나, 분석 알고리즘이 LDA로 제한적이고 토픽의 수를 결정하는 데 있어 해당 분야의 전문가의 의견이 반영된 선행 연구는 활성화되지 않았다. 이에 본 연구에서는 LDA 기반의 토픽 모델링과 해당 분야 전문가의 조언을 통해 연구를 수행한 점이 선행 연구와 분석 방법론에 있어 차별성을 가진다.

Table 2. Topic Modeling previous study

Author (Year)	Purpose of study	Analysis materials		Method of analysis
		material	period	
M.S.Chung (2014)	Analysis of research activities and trends related to artificial intelligence technology	11,085 papers in Web of Science with the keyword "artificial intelligence"	2000 ~ 2017	7개
W.S.Lee (2014)	Analyzing trends in space big data	“1,621 papers in Scopus with the keyword “space big data"	2014	19개
S.S.Han (2017)	Analysis of research trends related to start-up	1,010 papers in Riss with the keyword "start-up"	1974 ~ 2017	10개
K.C.Park (2018)	Smart City Research Trend Analysis	11,527 papers in Scopus DB and Springer DB with the keyword "Smart City"	2008 ~ 2019	8개
K.W.Cho (2019)	Fourth Industry Research Trend Analysis	685 papers in KCI with the keyword "Fourth Industry"	2016 ~ 2019	9개

1) Perplexity : 확률 분포 또는 확률 모델이 표본을 정확하게 예측하는지 평가하는 지표

2.3 이론적 고찰(LDA:Latent Dirichlet Allocation)

토픽 모델링과 관련하여 다양한 분야의 연구들이 진행되고 있으며, 연구 동향을 파악함으로써 현재까지 진행된 연구에 대한 이해와 함께 향후 연구 방향 제언 및 수립 등에 기여할 수 있다.[13] 무엇보다 연구 동향 및 유망 연구주제에 대한 파악은 전문가에 의해 이루어지는 것이 이상적일 것이다. 그러나 시간과 비용에 제약이 있어 어려운 부분이 있다. 그러므로, 연구의 결과물인 다수의 관련 문헌들을 대상으로 진행한 연구 동향 파악이 효과적일 수 있다. 이에 연구 동향 파악 및 분석에 주로 사용되는 토픽 모델링에 대해 살펴보고자 한다.

토픽 모델링은 전체 문서 집합에서 의미 구조를 파악하기 위해 사용되는 텍스트 마이닝 기법이며, 그 중 LDA 알고리즘은 토픽 모델링의 기법 중 대표적인 방법론으로 이산자료들에 대한 확률적 생성 모델로, 단어들의 출현 확률을 이용하여 문서 집합 내의 잠재된 토픽들을 찾아내는 기법이다[13]. 초기에는 잠재의미분석(LSA:Latent Sematic Analysis)로 시작하여 확률 기반 잠재의미분석(PLSA:Probabilistic LSA)으로 발전했다가, 2003년 Blei가 고안한 LDA 알고리즘을 발표한 이후 LDA가 토픽 모델링의 주요 기법으로 사용되고 있다.[14] LDA는 전체 문서 집합의 주제, 문서별 주제 비율, 단어들이 각 주제에 포함될 확률 등을 파악할 수 있다.[15]

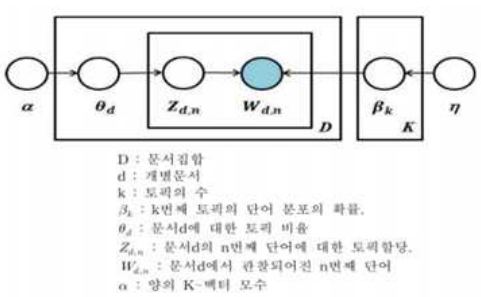


Fig. 1. Latent Direct Allocation Model

Fig. 1.은 LDA 그래프 모델이며 구체적인 내용은 다음과 같다.[16] Fig. 1을 살펴보면 K는 토픽의 개수, α는 θ값을 결정하는 파라미터이며, η는 β값을 결정하는 파라미터이다.[17] θ는 문서별 토픽의 비율. β는 토픽별 단어 w의 생성 비율이며 Zd,n은 문서 d의 n번째 단어의 토픽, Wd,n은 문서 d의 n번째 단어로 문서에서 관측되는 변수를 뜻한다.[18] θ는 각 문서 집합에 대한 주제 비율 값으로 Dirichlet 분포를 따르며 θ 값에 따라 문서 집합에서 존재하는 단어들의 주제인, z가 결정된다.[15] 또한 각 단어의 주제를 나타내는 값 Z와 토픽별 단어 생성 비율인 β값에 따라 단어 W가 결정된다.[15] 이처럼 LDA는 파라미터 값에 따라 결과가 달라진다.[17] LDA 토픽 모델링은 사전에 이용자가 K, α, β 값을 설정해줘야 하며 K값에 따라 토픽 모델링의 결과가 달라진다. 따라서 본 연구에선 Perplexity를 이용하여 적절한 K값을 찾아내 토픽 모델링을 수행하였으며, 해당 분야 정책적 전문가 1명과 연구원 1명의 의견을 반영하여 수행하였다.

3. 데이터 수집 및 분석 방법

3.1 데이터 수집 및 전처리

본 연구에서는 공공데이터를 활용한 연구 동향을 분석하기 위해 국내 학술정보는 한국연구재단 등재지 KCI, 국외 학술정보는 Springer DB를 통해 '공공데이터' 또는 'Public Data'를 키워드로 포함하고 있는 학술논문들의 자료를 수집하여 논문 제목과 초록, 발행 연도 등의 정보를 활용하여 공공데이터 연구 동향을 분석하였다.

국내 논문(KCI)의 경우 2007년 3월부터 현재(2020년 10월) 시점까지의 1,471편의 공공데이터 관련 논문이 수집되었으며, 국외 논문(Springer)에서는 2007년 11월부터 현재(2020년 10월)까지 10,000건의 공공데이터 논문이 수집되었다. 수집된 논문 중 중복 논문을 제거한 후 국내 1,437편, 국외 논문 9,607편을 본 연구의 분석에 활용하였다.

Table 3. Gathering data

Item	KCI	Springer
Search keyword	공공데이터	Public Data
Thesis number	1,437	9,607
Publication period	'07.03~'20.10	'07.11~'20.10
Data component	Title, Abstract, Year of issue	Title, Abstract, Year of issue

데이터의 전처리는 파이썬(Python)을 활용하여 수집된 논문의 제목과 초록을 결합하여 별도의 열을 생성하였으며, 영문의 경우 대문자를 소문자로 변환, 동사의 어근 추출, 특수문자 및 공백 제거 등을 수행하였다. 이를 분석하기 위해 자연어를 별도의 형태소 단위로 분리하는 토큰화(Tokenization) 작업이 필요하므로 국내 논문의 경우 Konlpy를 활용하였고, 국외 논문의 경우 nltk를 활용하여 명사의 품사를 가지는 단어들을 추출하여 작업을

수행하여 데이터 셋을 구축하였다. 또한, 국내외 논문별 불용어를 Table 4와 같이 설정하여 언어의 특성상 논문에서 자주 사용되는 키워드들을 추가로 제거하여 데이터 전처리를 진행하였다.

Table 4. Stopword list

Item	KCI (Domestic)	Springer (Foreign)	
Stopword list	데이터	Datum	Technology
	연구	Analysis	Type
	정보	Study	Chapter
	분석	Result	Case
	활용	Model	Role
	결과	Method	Problem
	이용	System	Source
	빅	Information	Performance
	영향	Expression	Activity
	기반	Approach	Function
	관련	Research	Technique
	제공	Network	Framework
	효율	Paper	Knowledge
	평가	Level	Mechanism
	효과	Process	Conclusion
	사용	Factor	Response
	자료	Use	
	가능	Application	
	조사	User	
	관계	Time	
	적용	Set	
	구축	Tool	
	방법	Number	
	문제	Sample	
	대상	Time	
	목적	Set	
	경우	Tool	
	성과	Number	
	특성	Sample	
	수집	Feature	
	모형	Dataset	

3.2 분석 방법

앞서 전처리한 데이터 셋을 활용하여 본 연구에서는 다음과 같은 절차로 국내, 국외 공공데이터 연구 동향을 비교 분석하였다.

첫 번째, 국내외 논문별로 추출된 키워드의 빈도수를 계산하여 공공데이터 분야의 핵심 키워드를 추출하였으며, 이를 국내외 비교 분석하였다.

두 번째, 논문 발행기관별 분석을 통해 분야별 공공데이터 연구영역을 간접적으로 추정하고, 국내외 비교 분석하였다.

세 번째, 전처리된 데이터 셋을 기준으로 LDA 기반의 토픽 모델링을 진행하였으며, 최적의 토픽 개수를 Perplexity로 비교하여 국내는 7개, 국외는 13개로 도출하였다. 이를 통해 토픽별 주제를 선정하였고, 해당 분야 전문가 2명에게 토픽 개수 및 토픽별 주제 선정에 대한 자문을 수행하였다.

마지막으로 분류된 토픽별로 시계열 분석을 진행하여 국내외 비교 분석 후 정책적 시사점을 도출하였다.

4. 분석 결과

4.1 키워드 빈도 분석

앞서 전처리한 데이터 셋을 활용하여 공공데이터 연구 동향에 관한 명사들의 키워드를 분석한 결과 Table 5와 같이 나타나며, 국내의 경우 '공공'이 가장 높게 나타나며, '기술', '서비스', '기관' 등의 순으로 나타나며, 그 외 '관리', '정책', '정부' 등 정부와 관련성이 높은 키워드들이 높게 나타남을 알 수 있다. 반면, 국외의 경우는 'Gene'이 가장 높게 나타나며, 'Cell', 'Cancer', 'Patient', 'Disease' 등의 의료 부문과 관련성이 높은 키워드들이 높게 나타났다.

이를 통해 국내와 국외의 공공데이터 연구 동향 차이를 개략적으로 파악할 수 있으며, 국내는 '공공', '기술', '서비스', '기관'과 관련된 연구가 활성화되었지만, 국외의 경우는 'Gene', 'Cell', 'Cancer', 'Patient'와 같이 의료 관련된 분야의 연구가 공공데이터 분야에 있어 활성화되어 있음을 추정할 수 있다.

Table 5. Keyword frequency analysis

Rank	Domestic		Foreign	
	Key word	frequency	Key word	frequency
1	공공	3,109	Gene	5,942
2	기술	1,189	Cell	4,020
3	서비스	1,123	Cancer	3,237
4	기관	1,074	Patient	2,625
5	관리	1,045	Development	2,322
6	필요	967	Disease	2,218
7	정책	937	Protein	1,912
8	정부	931	Risk	1,851

9	개인	912	Algorithm	1,799
10	사회	903	Health	1,759
11	지역	786	Effect	1,725
12	시스템	717	Service	1,721
13	제시	699	Change	1,646
14	개발	682	Database	1,510
15	도서관	681	Population	1,472
16	기업	631	Group	1,466
17	환경	617	Region	1,391
18	공간	598	Community	1,378
19	분야	595	Challenge	1,362
20	국가	548	Value	1,322
21	산업	548	Sequence	1,310
22	요인	543	Area	1,283
23	사업	538	Quality	1,265
24	방안	532	Tumor	1,257
25	조직	489	Image	1,247
26	운영	484	Impact	1,243
27	보호	460	Rate	1,225
28	민간	449	Resource	1,219
29	법	443	Genome	1,204
30	개선	436	Data	1,200
31	의료	420	Pathway	1,187
32	지원	416	Target	1,177
33	제안	413	Management	1,168
34	중심	410	Work	1,159
35	제도	409	Interaction	1,155
36	중요	400	Year	1,148
37	체계	400	Structure	1,137
38	교육	395	Policy	1,127
39	논문	388	Control	1,110
40	변화	387	DNA	1,109

이를 워드클라우드 형태로 시각화하여 살펴보면 Fig. 2와 같이 나타나며, Table 5와 유사하게 국내와 국외의 연구 동향의 차이를 살펴볼 수 있다. 빈도수에 비례하여 텍스트의 크기가 결정되며, 국내는 '공공', '기술', '서비스', '기관' 등의 키워드가 높게 나타나며, 국외는 'Gene', 'Cell', 'Cancer', 'Patient' 등의 의료 관련 키워드가 높음을 시각적으로 알 수 있다.

4.2 분야별 연구 동향 분석

공공데이터를 활용한 발행기관별 빈도수를 통해 공공데이터 연구 동향 분야를 간접적으로 추정하였으며, 저널별 빈도수를 살펴보면 Table 6과 같이 나타난다.

국내의 경우 과학기술학, 정보 관리학, 문헌정보학, 컴퓨터공학 등의 분야에서 공공데이터 활용 연구가 높은

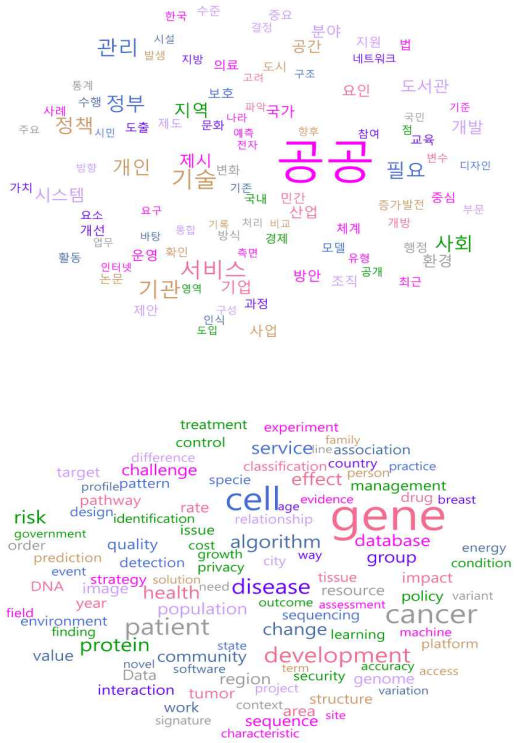


Fig. 2. Public Data Research Wordcloud (Top:Domesitc, Below:Outside)

것으로 판단되며, 국외의 경우는 자연과학, 유전학, 생물학 등의 분야에서 연구가 활성화되어 있는 것으로 나타났다. 이는 앞서 분석한 키워드 분석과 유사하게 나타나는 것으로 보인다.

Table 6. Number of papers published by domestic and foreign public data journals

Rank	Domestic		Foreign	
	Journal Name	Num	Journal Name	Num
1	Journal of Digital Convergence	34	Scientific Reports	330
2	Journal of the Korean Society for Information Management	29	Nature Communications	211
3	Journal of Korean Library and Information Science Society	28	BMC Genomics	204
4	Journal of the Korean Biblia Society For Library And Information Science	27	BMC Bioinformatics	191
5	International Journal of Contents	27	Nature	91
6	Journal of Korean Association for Regional Information Society	22	Genome Biology	87
7	Journal of the Korean Society for Library and Information Science	19	Scientific Data	76

8	The Korean Journal of Archival Studies	18	Multimedia Tools and Applications	73
9	The Journal of Information Technology and Architecture	17	Oncogene	59
10	Korean Journal of Construction Engineering and Management	17	Nature Genetics	55
11	Journal of Korea Academia-Industrial cooperation Society	16	BMC Cancer	52
12	Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology	15	Nature Methods	49
13	Journal of Intelligence and Information Systems	15	BMC Medical Genomics	46
14	Journal of Korean Society of Archives and Records Management	14	BMC Health Services Research	41
15	Journal of Cadastre & Land InformatiX	14	Nature Biotechnology	41
16	Law Review	14	Cell Death & Disease	40
17	Journal of Korean Institute of Information Technology	14	BMC Public Health	39
18	Informatization Policy	12	Genome Medicine	38
19	Korean Journal of Public Administration	11	Journal of Cheminformatics	37
20	Korean policy studies review	11	Journal of Business Ethics	35

공공데이터를 활용한 논문이 많이 게재된 상위 20개 저널을 추출하여 트리맵으로 시각화하여 살펴보면 Fig. 3.과 같이 나타난다. 국내의 경우 '디지털융복합연구', '정보관리학회', '한국도서관·정보학회'가 높게 나타나며, 국외는 'Scientific Reports', 'Nature Communications' 등의 자연과학 종합저널이 높음을 알 수 있다.

4.3 토픽 모델링 분석

본 연구에서는 공공데이터를 활용한 토픽 모델링을 수행하기 위해 LDA 알고리즘을 적용하여 분석하였다. LDA는 문서들에 들어 있는 토픽을 확률·통계적으로 분석할 수 있는 좋은 도구이지만 토픽(K)의 개수를 지정해야 한다. 따라서 토픽의 최적 개수를 산정하기 위해 Perplexity의 변화가 최소가 되는 지점을 추출하였으며, 토픽의 추출 범위는 선행 연구에서 실시한 연구 동향 분석의 범위를 참고하여 5개에서 15까지로 정하여 Perplexity 변화를 확인하였다. 토픽의 수를 지속하여 증가시켜 살펴볼 수도 있으나, 결과해석의 용이성과 간결성을 고려하여 최대 토픽 수를 15개로 한정하여 분석한 결과 Fig 4.와 같이 나타났다.

Fig. 3. Public Data Research Trends by Sector (Top:Domesitc, Below:Foreign)

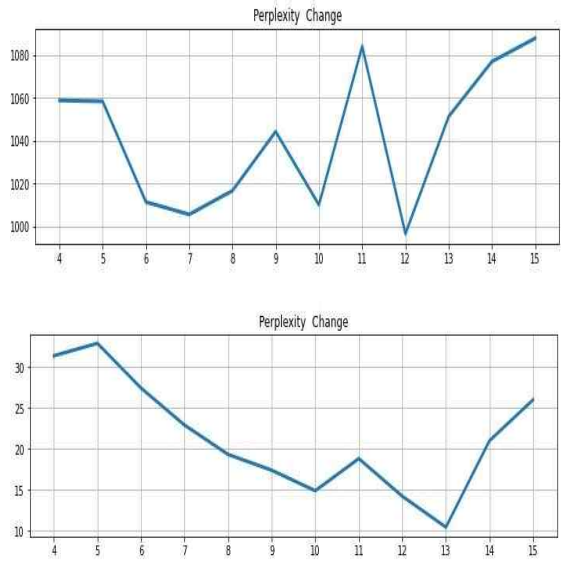


Fig. 4. Perplexity as the number of topics varies (Top:Domesitc, Below:Foreign)

국내의 경우 토픽의 수가 7개, 10개, 12개인 지점에서 Perplexity가 비교적 낮게 나타났으며, 이를 해당 분야의 전문가 2명에게 토픽 개수 및 토픽별 주제 선정에 대한 자문을 구한 결과, 국내의 경우 7개일 경우 가장 적합한 것으로 판단되어 토픽의 수를 7개로 설정하였다.

국외의 경우 토픽의 수가 13개일 경우 Perplexity가 가장 낮게 나타났기에 13개로 토픽을 설정하여 분석하였다. 이를 통해 토픽별 핵심 키워드를 추출하였으며, 키워드를 참고하여 토픽 번호별 주제 선정을 수행하였다.

공공데이터를 활용한 국내 연구 토픽 모델링은 크게 7개의 토픽으로 구분되며, Table 7과 같이 나타난다.

Table 7. Keywords by topic using Topic Modeling (domestic)

Topic (Topic Name)	Keyword	Number	Rank
Topic-1 (Information Protection)	개인, 기술, 보호, 법, 산업	225	2
Topic-2 (Analysis model)	예측, 모델, 공공 발생, 시스템	160	7
Topic-3 (Public library)	도서관, 공공 개발, 사업, 운영	189	4
Topic-4 (Public data management)	공공, 관리, 서비스, 시스템, 방안	322	1
Topic-5 (Public policy)	공공, 정책, 정부, 기관, 기업	201	3
Topic-6 (Public data recognition)	조직, 기술, 인식, 디자인, 시민	170	5
Topic-7 (Urban environment)	지역, 공간, 도시, 환경, 사회	170	5

토픽 1은 '개인', '기술', '보호', '법', '산업' 등이 연관 단어로 구성되어 있으며, 개인정보와 관련된 단어들이 중심이며, 빅데이터 활성화 및 공공데이터 개방에 따른 개인정보보호 관련 연구들이 다수 이루어졌음을 알 수 있다.

토픽 2는 '예측', '모델', '공공', '발생', '시스템' 등이 연관 단어로 구성되어 있으며, 이는 공공 주도 하의 빅데이터 모델, 시스템 구축에 관한 연구들이 이루어졌음을 알 수 있다.

토픽 3은 '도서관', '공공', '개발', '사업', '운영' 등이 연관 단어로 구성되어 있다. 이는 공공도서관 입지선정과 개발에 관한 연구들이 진행되고 있음을 알 수 있다.

토픽 4는 '공공', '관리', '서비스', '시스템', '방안' 등이 연관 단어로 구성되어 있다. 이는 공공데이터 관리와 시스템 구축 관련 연구들이 활발하게 이루어졌음을 알 수 있다.

토픽 5는 '공공', '정책', '정부', '기관', '기업' 등이 연관 단어로 구성되어 있다. 이는 정부 주도 하의 공공과 민간 분야에서의 데이터 활성화 정책 연구로 추정된다.

토픽 6은 '조직', '기술', '인식', '디자인', '시민' 등이 연관 단어로 구성되어 있다. 이는 공공데이터 관련 기술과 정책, 그리고 시민들의 인식 등의 연구들이 이루어짐을 알 수 있다.

마지막 토픽 7은 '지역', '공간', '도시', '환경', '사회' 등이 연관 단어로 구성되어 있다. 이는 도시 환경과 관련된 데이터로 도시 분야에 있어 공공데이터 활용이 높게 이루어짐을 알 수 있다.

국외 공공데이터 분야 토픽 모델링은 크게 13개로 구분되며, Table 8과 같이 나타난다.

토픽 1은 'Event', 'State', 'Evidence', 'Element', 'Firm' 등이 주요 연관 단어로 구성되어 있다. 이는 지역 현안과 관련된 단어들이 중심이며, 공공데이터를 활용한 지자체 현안 연구들이 이루어지고 있음을 알 수 있다.

토픽2는 'Recommendation', 'Student', 'Question', 'Survey', 'Program' 등이 주요 연관 단어로 구성되어 있다. 이는 공공데이터 활용 조사방법론에 관한 것으로 추정된다.

토픽 3은 'Risk', 'Year', 'Age', 'Association', 'Health' 등이 주요 연관 단어로 구성되어 있다. 이는 국민 건강과 관련된 연구들이 공공데이터를 활용하여 다수 수행된 것으로 보인다.

토픽 4는 'Energy', 'Value', 'Market', 'Power', 'Cost' 등이 주요 연관 단어로 구성되어 있다. 이는 에너지, 시장 분야에 있어 공공데이터 활용 연구가 이루어지고 있음을 알 수 있다.

토픽 5는 'Change', 'Area', 'Region', 'Population', 'Impact' 등이 주요 연관 단어로 구성되어 있다. 이는 지역별 인구 동향, 변화에 관해 공공데이터를 활용한 연구가 이루어지고 있음을 알 수 있다.

토픽 6은 'Community', 'Specie', 'Diversity', 'Strain', 'Environment' 등이 주요 연관 단어로 구성되어 있다. 이는 공동체 내의 다양성, 환경에 관한 연구들인 것으로 추정된다.

토픽 7은 'Development', 'Challenge', 'City', 'Policy', 'Data' 등이 주요 연관 단어로 구성되어 있다. 이는 국내 토픽 7과 유사한 도시 환경에 관한 토픽으로 국내와 같이 공공데이터를 활용한 도시 연구가 활성화되고 있는 것으로 보인다.

토픽 8은 'Database', 'Interaction', 'Pattern',

'Relationship', 'Cluster' 등이 주요 연관 단어로 구성되어 있다. 이는 공공데이터 기반의 데이터 인프라에 관한 연구로 보인다.

토픽 9는 'Cell', 'Cancer', 'Gene', 'Tumor', 'Protein' 등이 주요 연관 단어로 구성되어 있다. 이는 암세포, 종양 등과 같은 세포 생물학에 관한 연구가 공공데이터를 활용하여 이루어지고 있음을 알 수 있다.

토픽 10은 'Algorithm', 'Image', 'Learning', 'Detection', 'Classification' 등이 주요 연관 단어로 구성되어 있다. 이는 딥러닝 관련 키워드로 공공데이터를 활용한 딥러닝 관련 연구들이 주를 이루는 것으로 보인다.

토픽 11은 'Gene', 'Genome', 'Sequence', 'Sequencing', 'Protein' 등이 주요 연관 단어로 구성되어 있다. 이는 토픽 9와도 유사하나 유전학과 관련된 연구가 공공데이터를 활용하여 이루어지고 있음을 알 수 있다.

토픽 12는 'Service', 'Security', 'Privacy', 'Scheme', 'Cloud' 등이 주요 연관 단어로 구성되어 있다. 이는 클라우드, 보안 등의 토픽으로 보인다.

마지막으로 토픽 13은 'Patient', 'Disease', 'Health', 'Care', 'Treatment' 등이 주요 연관 단어로 구성되어 있다. 이는 질병 관리와 관련된 단어들로 토픽 3과 유사하게 건강과 관련된 연구 토픽인 것으로 나타났다.

Table 8. Keywords by topic using Topic Modeling (Foreign)

Topic (Topic Name)	Keyword	Number	Rank
Topic-1 (Regional issues)	Event, State, Evidence, Element, Firm	395	12
Topic-2 (Survey methodology)	Recommendation, Student, Question, Survey, Program	344	13
Topic-3 (Health)	Risk, Year, Age, Association, Health	563	8
Topic-4 (Energy)	Energy, Value, Market, Power, Cost	613	6
Topic-5 (Changes in area)	Change, Area, Region, Population, Impact	441	11
Topic-6 (Community diversity)	Community, Specie, Diversity, Strain, Environment	563	7
Topic-7 (Policy of city)	Development, Challenge, City, Policy, Data	1,751	1
Topic-8 (Data infrastructure)	Database, Interaction, Pattern, Relationship, Cluster	461	10
Topic-9 (Cell biology)	Cell, Cancer, Gene, Tumor, Protein	1,266	2
Topic-10 (Deep learning)	Algorithm, Image, Learning, Detection, Classification	957	4
Topic-11 (Genetics)	Gene, Genome, Sequence, Sequencing, Protein	857	5
Topic-12 (Cloud security)	Service, Security, Privacy, Scheme, Cloud	1,088	3
Topic-13 (Disease control)	Patient, Disease, Health, Care, Treatment	491	9

4.4 토픽 시계열 분석

추출된 토픽별 시계열 분석한 결과 Fig 5. 와 같이 나타나며, 국내의 모든 분야에서 지속하여 증가하고 있음을 알 수 있다.

특히 '개인정보보호', '공공데이터 관리', '도시 환경'을 나타내는 토픽 1, 토픽 4, 토픽 7의 경우 증가하는 추세를 보이며, '공공데이터 관리'의 경우 공공데이터에 관한 관심이 증가된 2012년부터 급격히 증가하는 것으로 나타났고, '개인정보보호'와 '도시 환경'의 경우 2014년부터 급격히 증가하는 것으로 나타났다.

반면, '공공데이터 정책'을 나타내는 토픽 5의 경우 2019년부터 하락하는 것으로 나타났다.

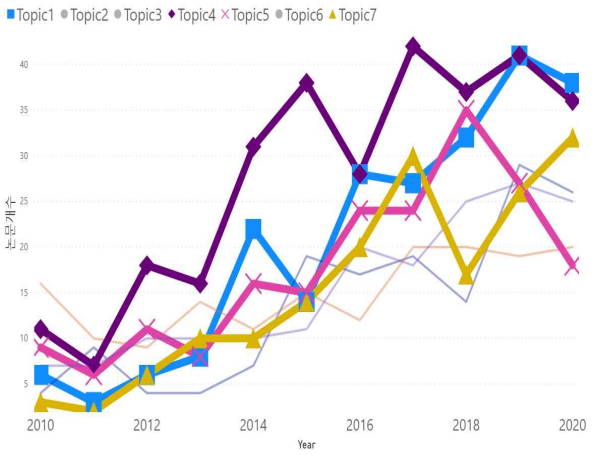


Fig. 5. Time Series Analysis by Topic(Domestic)

국외의 경우 2013년 이후 공공데이터 관련 모든 연구 분야에서 지속하여 증가하고 있으며, 특히 '도시정책'과 '세포 생물학', '딥러닝', '클라우드·보안'을 나타내는 토픽 7, 토픽 9, 토픽 10, 토픽 12의 경우 2012년부터 급격히 증가하고 있는 것으로 나타났다.

반면, ‘지역 현안’과 ‘조사방법론’, ‘건강’, ‘에너지’, ‘지역 변화’, ‘공동체 다양성’, ‘데이터 인프라’, ‘유전학’, ‘질병 관리’를 나타내는 토픽 1, 토픽 2, 토픽 3, 토픽 4, 토픽 5, 토픽 6, 토픽 8, 토픽 11, 토픽 13의 경우 증가추세가 상대적으로 완만한 것으로 보인다.

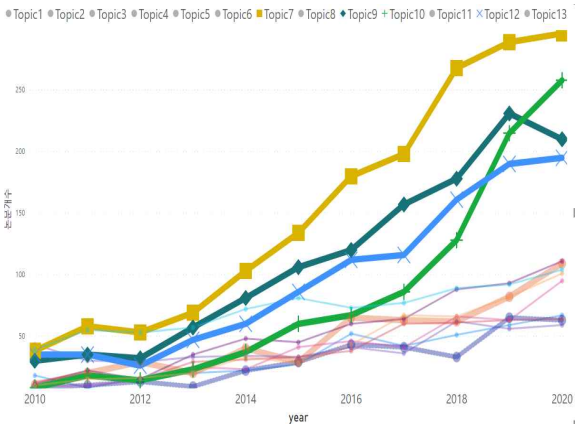


Fig. 6. Time Series Analysis by Topic(Outside)

5. 결론

5.1 분석 결과 요약

최근 4차 산업혁명으로 빅데이터의 성장과 가치는 지속하여 증가할 것으로 보인다. 특히 정부에서 디지털 정부로의 전환에 중점을 두고 있어, 공공데이터 개방과 활성화가 중요하며, 그 파급의 효과는 다른 산업과는 비교할 수가 없을 정도로 높으므로 해당 분야에 대한 면밀하고 장기적인 발전 계획 수립이 필수라 할 수 있다.

이에 본 연구에서는 공공데이터의 장기적인 발전 방향과 정책적 시사점을 제시하기 위해 국내외 학술논문 중 ‘공공데이터’, ‘Public Data'의 키워드가 포함된 논문을 수집하여 토픽 모델링으로 주로 활용되는 LDA 알고리즘을 통해 분석한 결과를 요약하면 다음과 같다.

첫째, 키워드 분석을 통해 국내의 경우 공공데이터 관련 활용, 정부 정책 등의 연구가 활성화되어 있으며, 국외는 의료 분야의 연구가 주를 이루는 것으로 나타났다.

둘째, 세부적인 연구 분야를 발행기관별로 살펴본 결과 국내는 과학기술학, 정보 관리학, 문헌정보학, 컴퓨터공학 등의 분야에서 공공데이터 연구가 높게 나타나며, 국외는 자연과학, 유전학, 생물학 등의 분야에서 공공데이터 연구가 높은 것으로 나타났다.

셋째, LDA 기반의 토픽 모델링을 수행한 결과 국내는 ‘개인정보보호’, ‘분석 모형’, ‘공공도서관’, ‘공공데이터 관리’, ‘공공 정책’, 공공데이터 인식’, ‘도시 환경’의 총 7개로 분류되었으며, 국외는 ‘지역 현안’, ‘조사방법론’, ‘건강’, ‘에너지’, ‘지역 변화’, ‘공동체 다양성’, ‘도시정책’, ‘데이터 클러스터’, ‘세포 생물학’, ‘딥러닝’, ‘유전학’, ‘클라우드·보안’, ‘질병 관리’로 총 13개의 토픽으로 분류되었다.

이를 세부적으로 살펴보면 국내는 공공데이터 관리, 정책, 개인정보보호 등의 토픽으로 크게 분류되며, 국외는 의료 분야인 건강, 세포 생물학, 유전학 등의 토픽 연구가 주를 이루는 점이 차이라 볼 수 있다.

또한 딥러닝, 클라우드·보안, 에너지 등의 토픽도 별도의 연구가 활성화되고 있음을 알 수 있다.

마지막으로 분류된 토픽별로 시계열 분석 결과 국내외 모두 2012년부터 공공데이터 연구가 지속하여 성장하고 있으며, 국내에서는 ‘개인정보보호’, ‘공공데이터 관리’, ‘도시 환경’ 연구가 상대적으로 급격히 증가하고 있으며, 국외에서는 ‘도시정책’, ‘세포 생물학’, ‘딥러닝’, ‘클라우드·보안’ 토픽의 연구가 활발히 진행되고 있다.

5.2 정책적 시사점 및 연구의 한계

본 연구의 분석 결과를 통해 정책적 시사점을 도출하면 다음과 같다.

첫째, 국내의 경우 정부 주도 하의 공공데이터 관리, 개인정보보호, 정책 등의 연구를 주인 반면, 국외는 의료, 에너지, 딥러닝 분야에 있어 공공데이터 연구가 활성화되어 향후 해당 분야에 있어 국내에서도 다양한 정책적 지원이 이뤄진다면 국외와 유사하게 관련 전문 분야 연구가 활성화 되어질 것으로 판단된다.

둘째, ‘14년부터 ‘도시 환경’ 토픽 분야의 논문 수가 지속하여 증가하고 있는데, 이는 국내 스마트 시티 조성과도 연관성이 높은 것으로 추정된다. 국내 스마트 시티 테스트 베드인 대구와 시흥시에서 수집되는 데이터들이 정제되어 시민들에게 개방되고 활용된다면 해당 분야 연구 및 산업이 지속하여 증가할 것으로 기대된다.

셋째, 국외의 경우 세포 생물학, 딥러닝, 클라우드·보안 분야 연구가 지속하여 증가하는데 이러한 국외 분야별 연구 방향성을 벤치마킹하여 국내에 적용하여 본다면 정책적인 측면에서도 다양한 시도가 될 수 있을 것으로 기대한다.

마지막으로 공공데이터 분야의 연구동향을 구조화하여 비교한 점이 학술적으로 기여한 바라고 판단하며, 다른 분야와의 융합을 통해 공공데이터 활성화 정책 수립에도 기초자료로 적용할 수 있을 것으로 기대된다.

한편, 본 연구의 한계는 다음과 같다.

첫째, 언어의 특성상 논문에서 자주 사용되는 단어들을 불용어 처리하여 전처리를 진행하였으나, 연구자의 주관적 견해가 포함되어 있어 향후 보완해야 할 사항으로 판단된다.

둘째, 공공데이터를 활용한 학술 연구를 논문 초록에서 '공공데이터', 'Public Data'로 정의하고 이를 추출하여 분석을 진행하였는데, 이에 대한 객관적인 논리가 다소 부족한 상황이라 본 연구의 포함되지 않은 논문이 다수 존재할 것으로 판단된다.

셋째, 국외의 경우 해당 분야의 전문가 확보 및 해석의 어려움으로 Perplexity의 정량적 수치로만 토픽의 수를 산정한 점이 연구 동향 분석의 한계로 보인다.

마지막으로 추후 학술논문 이외 언론, 소셜 네트워크, 공공데이터 관련 보고서 등의 다양한 자료와 결합하여 공공데이터 연구 동향을 분석한다면 심화된 공공데이터 연구 동향 분석이 가능할 것으로 판단된다.

REFERENCES

[1] D.M.Bae & H.S.Park, & G.H.Oh. (2013).Big data trends and policy lmplications.KISDI,25(10), 37-74.

[2] S.O.Yun, & J.W.Hyun. (2019). An Analysis of Open Data Policy in Korea: Focused on National Core Data in Open Data Portal . Korea Public Mangemnet Review, 33(1), 219-247.
DOI :10.24210/kapm.2018.33.1.010

[3] D.G.Kim & Y.H.Lee & W.K.Joo &, E.J.Kim, & Y.H.Lee(2014). A Case Study on Classification System Design for Public Sector Information Typology. *Journal of Digital Convergence, 12(4), 51-68.*
DOI :10.14400/JDC.2014.12.4.5

[4] Y.I.Cha & S.K.Choi & K.S.Han. (2017). An Empirical Study on the Influence on Public Data Usage in Private Business Sectors. *Journal of Digital Convergence, 15(6), 9-17.*
DOI :10.14400/JDC.2017.15.6.9

[5] S.B.Cho, & S.H.Ha. (2020). Analysis of Open Government Data Demand Using Structural Topic Modeling. *Journal of Information Technology and Architecturte 17(2),* 103-118.
DOI : 10.22865/jita.2020.17.2.103

[6] Wang, V., & Shepherd, D. (2020). Exploring the extent of openness of open government data-A critique of open government datasets in the UK. *Government Information Quarterly, 37(1),* 101405.
DOI : 10.1016/j.giq.2019.101405

[7] Wirtz, B. W., Weyerer, J. C., & Rösch, M. (2019).Open government and citizen participation: an empirical analysis of citizen expectancy towards open government data. *International Review of Administrative Sciences, 85(3),* 566-586.
DOI :10.1177

[8] W.S.Lee & S.Y.Sohn. (2015). Topic Model Analysis of Research Trend on Spatial Big Data .*Journal of the Korean Institute Of Industrial Engineers, 41(1),* 64-73.
DOI : 10.7232/JKIIE.2015.41.1.064

[9] S.S.Han & D.W.Yang. (2017). Analysis of Research Trends Related to Start-Up Using Text Mining. Asia-Pacific *Journal of Business Venturing and Entrepreneurship, 12(5*), 1-12

[10] M.S.Chung, & J.Y.Lee. (2018). Systemic Analysis of Research Activities and Trends Related to Artificial Intelligence(A.I.) Technology Based on Latent Dirichlet Allocation (LDA) Model. *Journal of the Korea Industrial Information Systems Research, 23(3),* 87-95..
DOI : 10.9723/JKSIIS.2018.23.3.087

[11] K.C.Park & C. H. Lee.(2019). A Study on the Research Trends for Smart City using Topic Modeling. *Journal of Internet Computing and Services, 20(3)*, 119-128.
DOI : 10.7472/JKSII.2019.20.3.119

[12] K.W.Cho & Y. W. Woo. (2019). Topic Modeling on Research Trends of Industry 4.0 Using Text Mining. J*ournal of the Korea Institute of Information and Communication Engineering,* 23(7), 764-770.
DOI : 10.6109/jkiice.2019.23.7.764

[13] T.K.Kim, H.R.Choi, H.C.Lee. (2016). A Study on the Research Trends in Fintech using Topic Modeling.,*Journal of Korea Academia-Industrial cooperation Society 17(11),* 670-681.
DOI : 10.5762/KAIS.2016.17.11.670

[14] C.H.Nahm (2016). An Illustrative Application of Topic Modeling Method to a Farmer's Diary. *Institute of Cultural Studies 22(1).*89-135
DOI : 10.3743/KOSIM.2013.30.1.007

[15] J.H.Park, & M.Song. (2013). A Study on the Research Trends in Library & Information Science in Korea using Topic Modeling. *Journal of the Korea Society for Information Management., 30(1)*, 7-32.
DOI : 10.3743/KOSIM.2013.30.1.007

[16] Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55(4),* 77-84.
DOI: 10.1145/2133806.2133826

[17] J.H.Park & H.J.Oh. (2017). 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, 48(4),* 235-258.
DOI : 10.16981/kliss.48.4.201712.235

[18] S.K.Kim, & S. Y. Jang. (2016). Analysis of Research Trends in Domestic Industrial Engineering Using Topic Modeling. Korean Institute Of Industrial

Engineers *Journal of the Spring Joint Academic Conference,* 3996-4018..

[19] J.H.Kim & S.Y.Yoon. (2019). Analysis of Policy Changes and User Satisfaction of Road Transportation Services using Opinion Mining Technique.International *journal of highway engineering, 21(5),* 65-74.
DOI : 10.7885/JHE.2019.21.5.065

[20] S.Y.Lee & H.J.Moon.(2017).Analysis of traffic research trend through big data analysis of world traffic related research . *Korean Academic Society Of Business Administration,* 299-313.

[21] J.Y.Sohn (2020). Big Markets, Big Names and Big Networks in Big Data Research: Urban Big Data Research Trends in International Academic Journals. *Journal of the Korea Geographical Society, 55(2),* 161-179
DOI : 10.22776/kgs.2020.55.2.161

[22] S.Y.Chung. (2019). Research Trends and Issues Analysis on the Use of Artificial Intelligence in Public Administration. *Journal of Korean Associastion for Regional Information Society, 22(4),* 175-207.

[23] H.J.Han & S.W.Hwang & J.M.Lee &H.J.Oh (2020).Analysis of Current Status and Improvement Plans of the User Service in Open Data Portal - Focusing on Citizen Participation Data Portal. *Journal of Korean Library and Information Science Society, 51(1),* 255-279.
DOI :10.16981/kliss.51.202003.225

[24] Y.W.Hong (2014). A study on the invigorating strategies for open government data. *Journal of the Korean data & information science society , 25(4),* 769-777.
DOI :10.7465/jkdi.2014.25.4.769

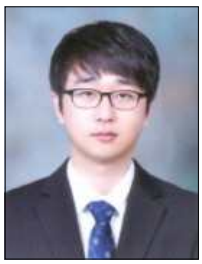
박 대 영(Daeyeong Park) [학생회원]

· 2015년 3월 ~ 현재 : 영남대학교 경영학과
· 2020년 10월 ~ 현재 : 대구디지털산업진흥원 공공 빅데이터 인턴
· 관심분야 : 서비스 경영, 서비스 이노베이션, 소비자행동, 빅데이터
· E-Mail : eodud6539@naver.com

김 덕 현(Deokhyeon Kim) [학생회원]

· 2015년 3월 ~ 현재 : 경북대학교 통계학과
· 2020년 10월 ~ 현재 : 대구디지털산업진흥원 공공 빅데이터 인턴
· 관심분야 : 빅데이터, 인공지능
· E-Mail : enumerator@naver.com

김 건 욱(Keunwook Kim) [정회원]



· 2009년 2월 : 영남대학교 도시공학(공학사)
· 2011년 8월 : 아주대학교 교통공학(공학석사-교통모델링)
· 2019년 7월 ~ 현재 : 대구디지털산업진흥원 빅데이터활용센터 센터장
· 관심분야 : 도시데이터분석, 빅데이터, 인공지능, 텍스트마이닝
· E-Mail : aut7767@dip.or.kr