

토픽모델링을 이용한 도시 분야 연구동향 분석

장선영, 정승현*

한국건설기술연구원 스마트시티연구센터

An Analysis of the Research Trends for Urban Study using Topic Modeling

Sun-Young Jang, Seunghyun Jung*

Smart Cities Research Center, Korea Institute of Civil Engineering and Building Technology

요약 연구동향은 시기별 연구주제에 대한 중요도 판단과 부족한 연구 분야를 파악하고 신규 분야를 발굴하는데 유용하게 활용될 수 있다. 본 연구에서는 인구집중과 도시화로 인해 다양한 문제가 발생하고 있는 도시공간을 대상으로 한 논문들을 대상으로 시기별 연구동향을 분석하였다. 이를 위해 2002년부터 2019년 사이에 게재된 한국학술지인용색인(KCI)에 등재된 논문의 초록을 대상으로 데이터마이닝 기법 중 하나인 토픽모델링 분석을 수행하였다. 토픽모델링은 전체 내용에서 일정한 패턴을 발견해낼 수 있는 알고리즘 기반의 텍스트마이닝 기법으로 방대한 문헌에서 주제를 찾아내고 군집하는데 용이하다. 본 연구에서는 키워드 빈도, 연도별 경향, 토픽 도출, 토픽별 군집, 토픽유형별 경향에 대한 분석을 실시하였다. 그 결과 먼저 도시재생 분야연구가 지속적으로 증가되고 있고 앞으로도 세부 주제가 확대될 수 있는 분야로 분석되었다. 그리고 도시재생 주제는 이제 정규 연구분야로 자리 잡고 있는 것으로 파악되었다. 반면, 개발/성장과 에너지/환경과 같은 주제는 정체기에 들어간 것으로 분석되었다. 본 연구는 국내 전체 도시분야 연구를 대상으로 데이터마이닝 기법인 토픽모델링을 이용하여 키워드 간 연관성과 경향을 함께 분석하였다는 데 의의가 있다.

Abstract Research trends can be usefully used to determine the importance of research topics by period, identify insufficient research fields, and discover new fields. In this study, research trends of urban spaces, where various problems are occurring due to population concentration and urbanization, were analyzed by topic modeling. The analysis target was the abstracts of papers listed in the Korea Citation Index (KCI) published between 2002 and 2019. Topic modeling is an algorithm-based text mining technique that can discover a certain pattern in the entire content, and it is easy to cluster. In this study, the frequency of keywords, trends by year, topic derivation, cluster by topic, and trend by topic type were analyzed. Research in urban regeneration is increasing continuously, and it was analyzed as a field where detailed topics could be expanded in the future. Furthermore, urban regeneration is now becoming a regular research field. On the other hand, topics related to development/growth and energy/environment have entered a stagnation period. This study is meaningful because the correlation and trends between keywords were analyzed using topic modeling targeting all domestic urban studies.

Keywords : Urban Study, Research Trend Analysis, Text Mining, Topic Modeling, Topic Clustering

본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었습니다. (과제번호 21TSRD-B151228-03).

*Corresponding Author : Seunghyun Jung(Korea Institute of Civil Engineering and Building Technology)
email: shjung@kict.re.kr

Received December 15, 2020

Revised January 12, 2021

Accepted March 5, 2021

Published March 31, 2021

1. 서론

1.1 연구의 배경 및 목적

지구상에서 발생하는 다양한 문제들은 대부분 도시에서 발생하고 있다. 그 원인 또한 인구증가와 인구의 도시 집중, 도시화에 의한 것이 대부분이다. 도시문제는 도시 확장에 따른 자연훼손, 대기질, 수질, 폐기물 처리와 같은 환경문제, 도시과밀에 따른 주택부족과 범죄발생, 차량 증가에 의한 교통혼잡과 사고, 공동체의 해체 등 다양한 부분에서 발생하고 있다. 이 같은 도시문제들은 사회경제적인 여건과 공동체의 인식체계에 기반하여 시기별로 중요도가 다르게 인식되어 왔다. 그리고 관련된 연구들도 이러한 현실을 반영하여 주제가 선정되어왔다. 과거 경제발전시기 건설 중심의 도시문제 해결 이슈에서부터 환경문제에 대한 관심으로 주제들이 확대되었고, 최근에는 기후변화에 대한 대응과 스마트시티, 도시재생 등으로 연구 분야가 확대되고 있다.

연구동향에 대한 연구는 시기별 연구주제에 대한 중요도 판단과 연구가 부족한 분야를 파악하고 신규 분야를 발굴하는데 유용하게 활용될 수 있다. 하지만 기존의 동향과 관련된 연구들[1-5]은 연구자의 주관적인 판단요소가 크게 작용하였다는 한계가 있다. 그러나 최근에는 데이터 마이닝 기법의 발달로 대용량의 텍스트 자료에 대한 전처리와 분석이 쉬워 이전보다 체계적으로 연구동향을 분석할 수 있게 되었다. 따라서 본 연구는 객관적이고 과학적인 방법을 이용하여 최근의 도시분야 연구동향을 분석하고자 한다. 구체적으로 데이터 마이닝 기법인 토픽 모델링을 활용하여 키워드간 연관성과 경향을 함께 분석하여 향후 연구주제 선정과 관련한 시사점을 제공하는데 목적이 있다.

1.2 연구의 방법

도시분야 연구주제 발굴을 위한 방안으로 본 연구에서는 문헌조사를 통해 기존 연구논문자료에 대한 분석을 연구과정에 도입하였다. 구체적으로 연도별 연구논문 키워드에 대한 빈도분석을 실시하였으며, 토픽모델링 기법을 이용하여 논문초록 내용에 대한 분석을 시행하였다. 토픽모델링은 특정 문서의 단어 분포를 분석하여 해당 문서가 다루는 주제들을 예측할 수 있다. 따라서 다수의 연구논문에서 다루는 주제들을 통해 연구 동향을 파악하는데 유용하다. 본 연구에서는 토픽모델링 기법 중 하나인 Blei 외 2인이 제시한 LDA (Latent Dirichlet Allocation)를 활용하였다[6].

1.3 연구의 범위

연구의 범위는 대상과 기간으로 구분할 수 있다. 먼저 연구주제의 분석대상은 도시문제 해결과 관련된 연구주제로 도시 전 분야로 설정하였다. 국내 대표적인 학술논문 데이터베이스에 등록된 학술연구 논문 전체가 대상이다. 시간적 범위는 데이터베이스를 통해 분석 가능하고 등재지 구분이 본격적으로 시행된 시기인 2002년부터 2019년으로 설정하였다. 2002년 이전은 등재지 데이터가 충분하지 않으므로 2002년부터의 자료를 수집하였다.

2. 선행 연구 및 관련 이론 고찰

2.1 선행 연구

국내 도시분야의 동향을 분석한 연구는 크게 세 가지로 유형화할 수 있다. 첫째, 도시분야 전반에 걸친 정책 및 학술논문의 추이 분석연구, 둘째, 도시분야 안에서도 스마트시티나 관광 등과 같은 세부 주제에 한정된 동향 분석연구, 셋째, 특정 기술적 방법론의 활용에 초점을 맞춰서 동향을 분석한 사례이다.

도시분야 전반에 걸친 추이 분석연구는 누적된 다량의 데이터를 바탕으로 잠재적인 의미 도출과 관계 분석을 통해 동향을 분석하였다. 기존 연구로는 정책 동향 분석 [7,8]과 건축·도시계획에 관한 학술논문 분석 연구[9]가 이루어졌다.

도시분야 중에서도 세부 주제에 집중된 동향 분석은 스마트시티[10-12], 도시재생[13-15], 도시이미지 및 경관[16,17], 관광[18], 리모델링[19]에 관한 영역을 대상으로 수행되었다.

Table 1. Classification according to analysis methodology

Analysis Methodology		Analysis Theme	
Qualitative Analysis Methodology		Urban Regeneration [13]	
		Eye Tracking Method [16]	
Quantitative Analysis Methodology	Frequency and Keyword Analysis	Urban Regeneration [14]	
		City Image [17]	
		Tourism [18]	
	Text Mining	Network Analysis	Government Policy & Research Trends [8]
			Smart City [10]
		Topic Modeling	Urban Regeneration [15]
			Remodeling [19]
		Government Policy & Research Trends [7,9]	
		Smart City [11,12]	

특정 기술적 방법론의 활용에 초점을 맞춘 동향 분석 연구는 아이트래킹 기법을 적용하여 도시건축경관을 연구한 문헌들에 대한 동향분석[16]과 건축분야에서의 인공지능 기계학습 연구동향 분석[20], 텍스트 마이닝을 적용한 기존 문헌들의 동향을 분석[21]한 연구들이 있다.

도시 분야를 대상으로 동향분석을 수행한 연구는 그 수가 많지 않으나 지난 5년간 급격히 증가하였다. 최근 연구들은 누적된 데이터를 바탕으로 키워드 네트워크 분석 및 텍스트 마이닝의 다양한 기법들을 분석 방법론으로서 적용하고 있다(Table 1).

주요 분석기법으로는 텍스트 마이닝 및 토픽모델링을 적용하고자 하는 연구들이 증가하고 있다. 키워드 자료만을 활용한 기존 동향분석 연구들의 한계에서 나아가 워드 클라우드와 N-그램 모델링, 토픽모델링 등을 활용하여 국토 및 도시계획 연구주제의 분류와 동향을 분석[9]하거나 도시재생에 대한 국제 연구 동향과 정책결정자의 담화 분석을 의미 연결망 분석과 토픽모델링을 중심으로 수행하여 정부간 시각차를 확인한 연구[7]도 존재한다. 그리고 최근에는 스마트시티에 관한 관심이 증대되면서 관련 연구동향 분석이 증가하고 있다. 박건철 외 1인은 스마트시티 관련 연구들에 대하여 토픽 모델링과 주제 간 연관관계를 분석하여 스마트시티 연구현황에 대한 객관적 분석 결과 얻었으며[11], 박한샘 외 2인[12]의 연구에서는 토픽 모델링을 통해 드러난 주제들을 핫토픽과 콜드토픽으로 분류하였다. 임시영 외 2인은 텍스트 마이닝을 통해 U-City와 스마트시티의 연구동향에 대해 분석하고 차이점을 확인하고자 하였다[10].

위의 연구들과 같이 토픽모델링은 문서에 존재하는 주제를 도출해내고 다수의 문헌으로부터 연구동향 파악에 유용하게 활용되고 있다. 따라서 기존의 정성적 방법론이 가진 분석자의 관점과 기준에 의해 분석 결과가 달라질 수 있으며, 문서에 내포된 잠재된 의미 또는 주제의 파악이 어려운 한계를 해결하기 위해서는 토픽모델링이 유용하게 활용될 수 있다.

2.3 토픽 모델링

연구동향을 조사하는데 있어서 핵심키워드 추출이 중요하다. 텍스트 마이닝 기법 중 하나인 토픽 모델링은 비구조화된(unstructured) 대량의 문서집합에서 잠재된 주제(topic)들을 추출하는 확률 모델이다[6]. 토픽 모델은 문서 모음으로부터 말뭉치(corpus)에 포함된 단어들을 사용해 주제를 식별한다. 이 모델은 동일한 문서 내에 있는 단어들이 연관되어 있다고 가정하고 그 가정을 사

용해 반복적으로 유사하게 나타나는 단어 그룹을 찾아 주제를 정의한다[22].

LDA (Latent Dirichlet Allocation)는 대표적인 토픽모델링 기법으로 문서 포함되는 주제들의 확률 모델이다. LDA는 문서나 단어와 같은 관측되는 변수(observed variable)를 통해 문맥, 문서의 구조와 같은 보이지 않는 변수(hidden variable)를 추론하는 방법으로 각 단어들이 특정 주제에 포함될 확률, 문서별 주제 비율, 전체 문서집합의 주제를 파악할 수 있다[23]. LDA는 단어(w)가 특정 주제(t)에 해당할 확률과 단어(w)가 포함된 모든 문서 집합에서 특정 주제(t)의 확률을 결합 확률로 추정하여 주제를 추출한다. Blei의 LDA 그래픽 모델[24]은 이를 도식화하여 나타내고 있다(Fig. 1). 토픽 추출과정을 나타내는 그래픽 모델에서 θ 는 문서의 각 주제별 비율(topic proportions)이고 Z 는 특정 주제에 대한 단어의 확률로, θ 값에 따라 Z 가 결정된다. β 는 주제(topic)이며, 단어 W 는 Z 와 β 의 값에 따라 결정된다.

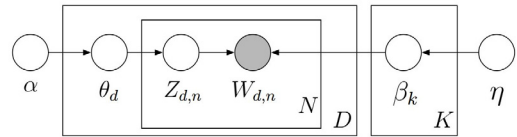


Fig. 1. The graphical model for latent Dirichlet allocation [24]

3. 연구 방법

3.1 데이터 수집

연구문헌은 한국학술지인용색인(KCI)[25] 데이터베이스를 통해 수집되었다. KCI 데이터베이스로부터 국내

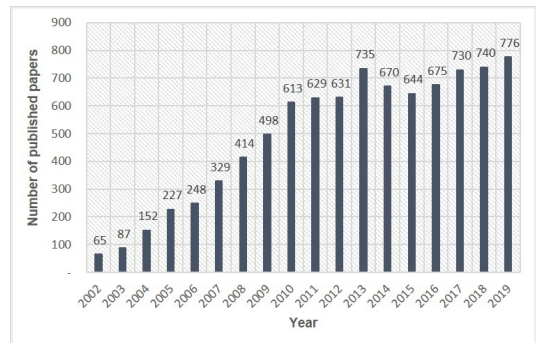


Fig. 2. Statistics of collected documents by year

KCI 등재지 중 기간인 2002년 1월부터 2019년 12월까지를 대상으로 하였다. 수집된 전체 문헌은 8,863건이며, 발행연도별 통계는 Fig. 2와 같다.

본 연구에서는 수집된 연구문헌에서 제목, 저자, 연도, 키워드, 초록의 정보를 추출하여 분석에 활용하였다.

3.2 분석 과정

수집된 문헌 데이터를 활용하여 키워드 분석과 LDA 토픽 모델링을 수행하였다. 키워드 분석은 수집된 키워드들 중 같은 의미를 다르게 표현한 데이터들을 전처리하였다. 예를 들어, 셉티드는 원어인 CPTED와 한글로 표현한 셉티드, 셉테드, 범죄예방환경설계 등으로 조금씩 다르게 표현하고 있다. 이와 같은 표현은 빈도수 분석에 영향을 주므로 분석 전에 동일한 건으로 정의할 필요가 있다.

키워드 분석에서는 2002년부터 2019년의 전체 데이터를 대상으로 빈출 키워드와 그 빈도수를 도출하였다. 20개의 상위 빈출 키워드를 통해 지난 18년간 도시 분야의 주요 키워드를 살펴보고 상위 6개 키워드를 선별하여

연도별로 변화 추이를 도출하였다.

LDA 토픽모델링은 연구문헌의 초록데이터를 대상으로 외국어 초록의 번역, 불용어 제거, 형태소 추출, 사용자 사전 등록의 전처리 과정을 거친 후에 수행되었다. 분석과정은 먼저, 전체 연도를 대상으로 10개의 토픽을 추출하고 각 연도별로 다시 토픽모델링을 수행하여 연도별로 가장 빈출한 키워드를 추출하였다. 전체 연도를 대상으로 추출한 토픽 10개와 연도별로 추출한 빈출 키워드를 맵핑하여 하나의 토픽 분야(예: 재생/커뮤니티)에서 빈출하는 키워드가 연도별로 어떻게 변화되는지를 살펴 보았다.

종합적으로 본 연구는 키워드 분석과 LDA 토픽 모델링 분석을 통해 도시 분야의 정책 및 연구가 어떤 주요한 이슈를 다루고 있으며, 시계열적으로 어떤 정책, 의제, 사건에 영향을 받았는지 분석하였다. 또한 분석된 연구 결과를 통해 도시 관련 연구의 유망 분야와 향후 동향을 예측하였다.

4. 분석 결과

4.1 도시 관련 핵심 키워드 분석

도시 연구논문의 키워드는 중복을 제외하고 총 34,925개가 도출되었다. 빈출 키워드 중 상위 20개의 순위와 빈도수는 Fig. 4와 같다.

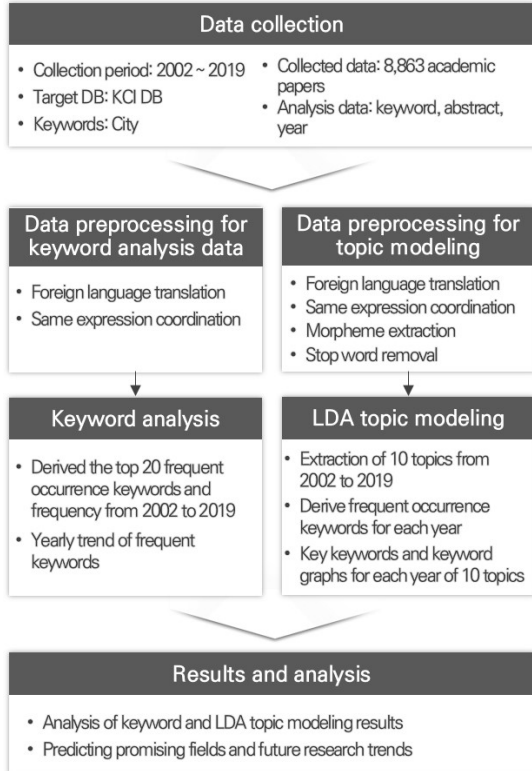


Fig. 3. Data analysis method and procedure

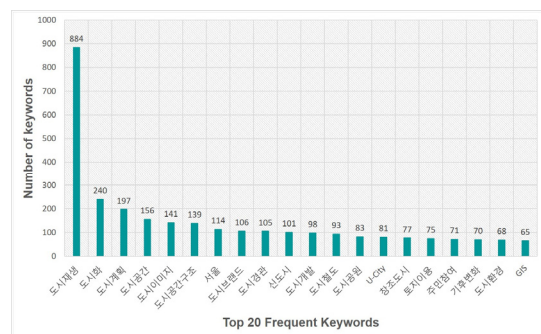


Fig. 4. Top 20 Frequent Keywords and Frequency

가장 빈출한 키워드는 “도시재생”이었으며, 총 884개로 집계되었다. 이어서 “도시화”(240개), “도시계획”(197개) 등이 빈출 키워드로 도출되었다.

빈출 상위 20개의 키워드는 몇 개의 주요 이슈로도 분류해 볼 수 있다. 첫째, 신도시 개발과 U-city, 토지이용

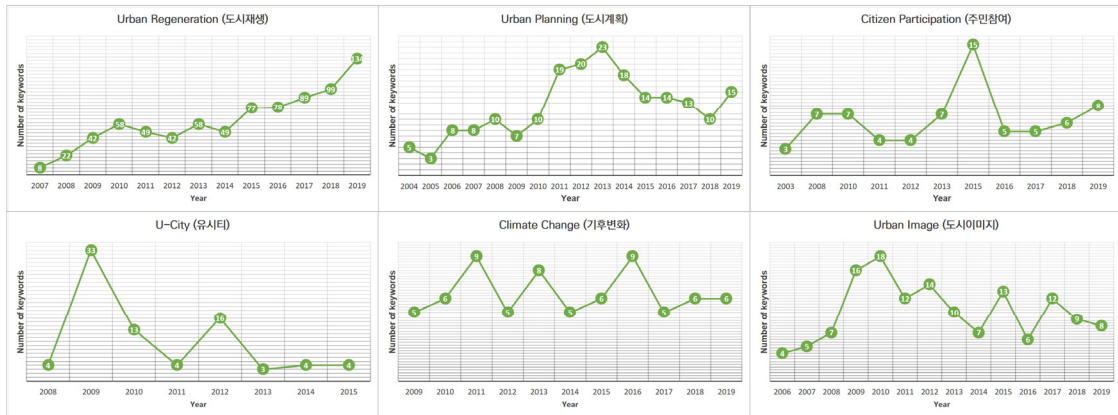


Fig. 5. Yearly trend of frequent keywords

사업 등의 도시개발 중심 이슈, 둘째, 급격한 도시화로 인해 낙후된 도시의 재생 및 회복과 관련된 이슈, 셋째, 도시계획, 도시공간, 도시공간구조와 같은 계획 및 설계와 관련된 이슈, 넷째, 주민참여 관련 이슈, 다섯째, 기후변화로 인한 도시공원, 도시환경, 도시열섬(60번째 빈출) 등의 환경적 이슈, 여섯째, 도시이미지, 도시브랜드, 도시경관과 같은 도시의 시각적 이미지화 및 브랜딩과 관련된 이슈 등으로 유형화할 수 있다.

이를 기준으로 연도별 각 유형의 추세를 Fig. 5와 같이 표현해보면, 먼저 가장 빈출한 키워드인 “도시재생”은 2007년부터 꾸준히 상승하는 형태의 그래프를 나타낸다. 2010년도와 2015년도, 2019년도가 전 해 대비 크게 상승하는 모습을 보였다. 도시재생 연구의 필요성에 대한 공감과 관련 특별법의 제정(2013년)을 중심으로 그 전후의 해당 주제에 대한 관심이 두드러진 것으로 분석된다.

두 번째 “도시계획” 키워드는 2011년부터 2013년 사이에 큰 폭으로 상승하였다. 도시계획 연구는 여러 키워드 중에서도 다양한 세부 항목을 가지고 있다. 도시계획 연구의 범주에는 도시계획 및 관리기법, 도시역사, 재해 예방 도시계획, 여성친화도시와 같은 하위 주제를 포함하고 있다.

세 번째 “주민참여” 키워드는 2015년도에 상대적으로 높은 수치를 나타내고 있다. 특히 주민참여 키워드는 도시재생 키워드와 동시에 등장하는 사례가 많은데 이는 4.2장에서 다루어질 도시 분야 토픽 추출에서도 확인할 수 있다. 주민참여 키워드의 연구들에서는 재생사업이 법제화되고 본격적으로 사업을 진행하는 과정에서 이해관계자로서 주민의 참여와 상호작용을 중요하게 여기고 있다.

네 번째 “U-city” 키워드는 현재 스마트도시의 전신으로 지난 2008년 유비쿼터스 도시 건설 법률을 제정한다 있다. U-city 키워드는 법률 제정과 유사한 시기인 2008년 등장해서 2009년에 많은 관심을 받은 것으로 보인다. 그러나 이후 차츰 감소세를 보이다가 2016년도 이후로는 등장하지 않는 키워드가 되었다. 이는 ‘스마트시티’라는 새로운 용어가 사용되었으며, 2017년 ‘스마트도시의 조성 및 산업진흥에 관한 법률’에 의해 전면 개정된 영향으로 판단된다.

다섯 번째 “기후변화” 키워드는 매년 꾸준히 제시되는 주제어로 도시화, 폭염, 홍수, 탄소저감 등 자연환경 변화와 관련된 주제어와 시뮬레이션, 모델과 같은 실험 및 정량화하는 방법론의 키워드와 함께 나타나는 양상을 보인다.

여섯 번째 “도시이미지” 키워드는 2009~2010년에 가장 상승폭이 두드러졌으며, 이후에는 서서히 감소하는 형태를 보인다. 도시이미지는 도시설계와 디자인 분야의 키워드 유형으로 전통적인 도시계획분야와 흐름을 같이하는 것으로 판단된다.

4.2 도시 분야 토픽 추출

앞서 분석은 단순히 상위 빈출 키워드를 산출한 것으로 토픽모델링은 키워드 간에도 활용된 문헌의 내용에 따라 상호 간의 관련성에 기반하여 군집화가 가능하며 주제(topic)화 할 수 있다. 과거 연구논문들의 주제를 도출하기 위해 논문 초록을 대상으로 토픽 모델링을 수행하였다.

토픽모델링의 첫 번째 단계는 토픽의 개수를 선정하는 것이다. 토픽의 개수는 사용자가 임의로 설정할 수 있는데, 토픽의 개수를 적게 설정하면 하나의 주제군 안에 여

Table 2. Major keywords of LDA topics from 2002 to 2019

Topic (ratio)	Theme	Keywords
Topic1 (9%)	서비스/정보기술 Service/ Information Technology	시스템, 정보, 프로젝트, 기술, 통합, 서비스, 관리, 데이터, 네트워크, 인프라, 유시티, 스마트 System, Information, Project, Technology, Integration, Service, Management, Data, Network, Infrastructure, U-city, Smart
Topic2 (9%)	제도/법규 Institution/Regulation	계획, 제도, 사업, 관리, 토지, 지방, 정비, 지구, 자치, 행정구역, 국토, 법률 Plan, Institution, Business, Management, Land, Province, Maintenance, District, Autonomy, Administrative district, Land, Law
Topic3 (10%)	문화/관광 Culture/Tourism	문화, 이미지, 역사, 가치, 전략, 자원, 브랜드, 관광, 경제, 장소, 마케팅, 콘텐츠 Culture, Image, History, Value, Strategy, Resource, Brand, Tourism, Identity, Place, Marketing, Content
Topic4 (9%)	개발/성장 Development/Growth	산업, 인구, 서울시, 성장, 도심, 수도, 쇠퇴, 이동, 범죄, 혁신, 지방, 균형 Industry, Population, Seoul, Growth, Downtown, Capital city, Decline, Movement, Crime, Innovation, Province, Balance
Topic5 (10%)	외교/역사 Diplomacy/History	일본, 중국, 시장, 정치, 전통, 조선, 자본, 건설, 식민지, 근대, 세계, 역사 Japan, China, Market, Politics, Tradition, Shipbuilding, Capital, Construction, Colonial, Modern, World, History
Topic6 (10%)	주거/생활 Housing/Life	주택, 주거, 생활, 농촌, 안전, 건강, 거주, 노인, 아파트, 공급, 소득, 복지 Housing, Residential, Life, Rural, Safety, Health, Residence, Elderly, Apartment, Supply, Income, Welfare
Topic7 (9%)	재생/커뮤니티 Regeneration/Community	재생, 사업, 주민, 참여, 교육, 공동체, 마을, 프로그램, 협력, 전문가, 시민, 커뮤니티 Regeneration, Business, Residents, Participation, Education, Community, Village, Program, Cooperation, Professional, Citizen, Community
Topic8 (12%)	담론/문학 Discourse/Literature	현대, 사람, 근대, 인간, 세계, 문학, 의식, 장소, 시대, 담론, 비판, 대중 Modern, Person, Modernity, Human, World, Literature, Ritual, Place, Times, Discourse, Criticism, Public
Topic9 (9%)	계획/설계 Planning/Design	시설, 디자인, 공원, 공공, 경관, 건물, 건축, 계획, 설계, 거리, 생태, 녹지 Facility, Design, Park, Public, Landscape, Building, Architecture, Plan, Design, Street, Ecology, Green area
Topic10 (13%)	에너지/환경 Energy/Environment	평가, 지표, 교통, 시간, 철도, 에너지, 기후, 추정, 예측, 온도, 도로, 녹색 Evaluation, Indicator, Transportation, Time, Railroad, Energy, Climate, Estimate, Forecast, Temperature, Road, Green

러 유형의 주제가 혼합되고, 개수가 많게 설정되면 토픽이 중복되는 문제점이 있다. 따라서 본 연구에서는 토픽의 개수를 점진적으로 변화시켜 일관성 있는 적정 토픽

개수를 산출하였다. 토픽의 개수가 10개일 경우 주제의 혼재가 최소로 판단되어 이를 기반으로 주제군과 주제군별 빈출 키워드를 도출하였다.

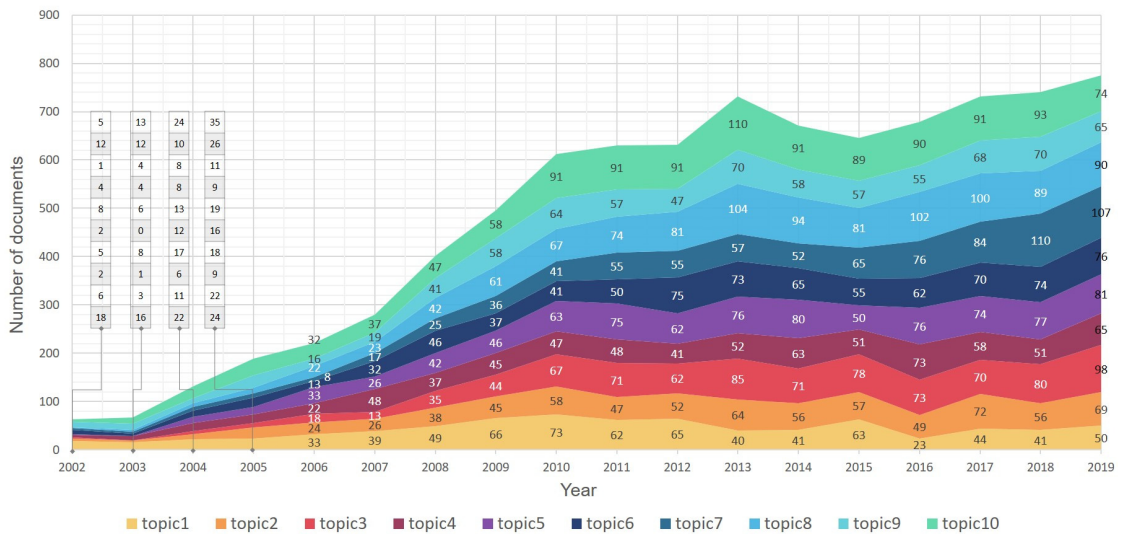


Fig. 6. Topic trend by year (2002~2019)

Table 2는 전체 문서 중에서 각 토픽이 차지하는 비율과 토픽의 주요 키워드 고려하여 의미적으로 포괄할 수 있는 주제(theme)를 선정할 결과다. 전체 문서에 대한 토픽 분포는 LDA 토픽 모델링 결과를 다차원 척도법을 기반으로 시각화하는 도구인 LDAvis로 표현하여 토픽 군집별 관련 정도를 파악할 수 있다(Fig. 7).

분석결과를 4분면으로 구분하면 1사분면은 5(외교/역사)와 8번(담론/문학) 토픽의 유사도가 매우 높은 것을 확인할 수 있다. 세부 키워드를 보면 해당 토픽은 주로 인문학 관련 연구들로 판단할 수 있다.

2사분면은 1, 4, 6, 10번 토픽으로 구성되어 있으며, 이중 1(서비스/정보기술)과 4번(개발/성장)과 유사도가 매우 높다. 세부 키워드를 보면 1번은 주로 스마트시티, 유시티와 관련된 내용으로 판단되고, 4번은 스마트시티와 유시티에서 제공되는 서비스와 유발기대효과 관련 키워드로 구성되어 있음을 알 수 있다. 그리고 6번과 10번의 경우에도 스마트시티의 서비스 분야에 해당하는 주제들로 구성되어 있다.

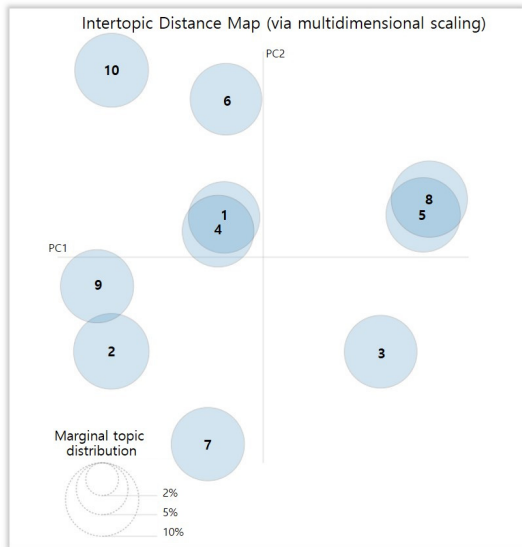


Fig. 7. Visualization of topic clustering

3사분면은 2, 7, 9번으로 주로 도시재생과 관련된 주제로 구성되어 있다. 세부 키워드들은 도시재생의 물리적 계획, 정책, 법률, 공동체와 관련되어 있으며 이를 통해 도시재생 키워드들이 주로 연구된 분야를 파악할 수 있다. 특히 2번(제도/법규)은 9번(계획/설계)과 유사성이 높은 것으로 분류되었다.

4사분면은 3번(문화/관광)으로 타 분야와 다르게 장소

마케팅, 관광차원에서의 역사, 문화, 이미지에 대한 내용을 다루는 연구키워드로 구성되어 있다. 이 분야는 1사분면의 5번(외교/역사)와 가까운 키워드들로 구성되어 있으며, 인문학 연구주제다.

4.3 토픽유형별 추세 분석

Fig. 6은 10개의 토픽들에 해당하는 논문수를 연도별로 나타낸 것이다. 10개의 토픽 중에서 Topic 10이 13%로 가장 많았다. Topic 10은 “에너지/환경”에 관한 주제로 도시의 기후환경 변화, 온도 상승, 녹지 구조 등에 대해 다루고 있으며 도시의 교통, 도로, 철도 관련 연구들을 포함하고 있다. 두 번째로 많은 비중을 차지하는 토픽 유형은 Topic 8 “담론/문학”으로 12%를 차지한다. Topic 8은 도시 공간에 대한 역사적 의미와 문화사적 의미를 찾는 연구, 유형학과 같은 특성 연구, 도시를 배경으로 한 작품들에 대한 연구를 포함한다. 그 외 나머지 토픽들의 경향들은 모두 유사하게 나타났다.

이 같이 기존의 토픽별 분류에서는 추세나 경향을 파악하기 어려우나, 앞서 사분면별로 구분한 토픽들로 묶어서 추세를 Fig. 8과 같이 표현하면 그 차이를 확인할 수 있다. 먼저 4사분면의 경우 3번(문화/관광)만 해당되며, 가장 낮은 비율을 보이고 있다. 2017년 이후 논문수가 증가하는 것으로 보이나 1사분면에 해당되는 5, 8번 토픽과 마찬가지로 전반적으로 급격한 증가는 보이지 않는다.

3사분면은 도시재생과 관련된 토픽들로 전반적으로 관련 주제의 논문들이 증가하는 추세를 보인다. 도시재생은 앞서 분석된 상위 빈출 키워드 분석에도 가장 높은 빈도를 보였고 추세도 증가하는 것으로 나타났다.

2사분면은 가장 많은 수의 논문을 포함하는 토픽유형으로 2010년까지 증가세를 보이다가 이후 정체되는 경향을 보이고 있다. 이는 2사분면의 주제들이 도시분야에

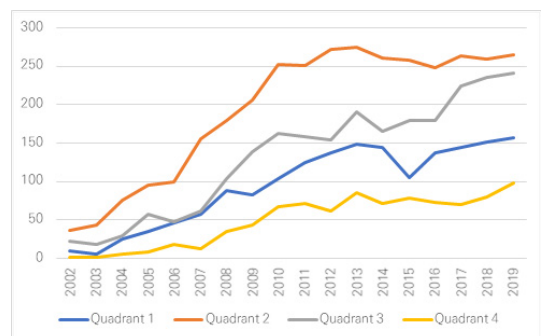


Fig. 8. Topic Group Trend by year (2002~2019)

서 일반적으로 다루지는 주제들로 도시재생과 같은 새로운 키워드 보다는 일반적인 도시분야 연구주제들로 구성되어 연구분야 확대의 동력이 크지 않은 이유로 판단된다.

5. 요약 및 결론

본 연구는 도시분야 연구동향을 파악하기 위해서 2002년부터 2019년까지 발간된 국내 학술등재지 중 '도시'라는 주제로 검색되는 논문의 초록을 분석하였다. 초록별 키워드의 빈도와 시기별 경향을 파악하였고, 데이터마이닝 기법 중 하나인 토픽모델링을 통해 주요 토픽을 도출하였다. 그리고 도출된 토픽간의 관련성을 바탕으로 유형화하고 각 유형별 연도별 경향을 파악하였다.

전체 키워드 빈도 분석에서는 도시재생이 가장 높게 나타났고, 연도별 빈도에서도 지속적으로 증가하는 추세를 보였다. 토픽모델링에 의한 총 10개의 토픽(서비스/정보기술, 제도/법규, 문화/관광, 개발/성장, 외교/역사, 주거/생활, 재생/커뮤니티, 담론/문학, 계획/설계, 에너지/환경)이 도출되었으며 초록 내 키워드들 간의 연관성을 바탕으로 다시 4개로 분류하였다. 분류된 토픽들의 연도별 경향을 분석한 결과에서도 도시재생과 관련된 토픽들이 지속적으로 증가하는 것으로 파악되었으며, 그 외 일반적인 도시연구주제는 증가하지 않거나 정체되고 있는 것으로 분석되었다. 이 같은 연구결과를 바탕으로 다음과 같은 시사점을 도출할 수 있다.

첫째, 향후 연구분야에 있어서 도시재생분야 연구는 지속적으로 증가되고 있다. 타 분야에 비해서 도시재생연구는 연구가 필요한 세부 주제가 다양하고, 연구분야의 확대가 지속될 수 있는 분야로 판단된다. 키워드로 도출되는 규모로 판단했을 때 도시재생 관련 연구들은 제도화된 하나의 정규 연구분야로 자리잡고 있다고 볼 수 있다. 도시재생 토픽과 관련성이 높은 토픽들로 계획/설계, 제도/법규 토픽들이 선정되었는데, 이는 이전의 전통적인 도시계획과 설계 영역의 연구가 대부분 도시재생이라는 키워드를 중심으로 전환되고 있다고 볼 수 있는 것이다.

둘째, 서비스/정보기술 부문의 중요성이 증가하고 있다. 서비스/정보기술은 개발/성장 관련 키워드들과 연관성이 가장 높게 나타났고 다른 키워드과의 관련도도 유사성 분석에서 높게 나타났다. 이는 최근 전세계적인 스마트시티 추진 경향과 국내 동향을 고려했을 때 4차산업혁명으로 인한 지능정보기술이 도시문제해결에 적극 도입되고 있는 상황이 반영된 결과로 판단되며, 경향분석에서

도 나타난바와 같이 향후 앞으로도 높은 비중을 차지할 것으로 예상된다.

셋째, 개발/성장과 에너지/환경과 같은 주제는 정체기에 들어간 것으로 분석되었다. 기후변화가 심화되고 있는 상황에서 에너지와 환경관련 연구들은 개별 토픽의 연도별 변화에서는 증가양상을 보이고 있으나, 반대로 개발과 성장에 대한 연구들이 줄어들고 있어 결과적으로 전체 경향이 둔화 또는 정체되고 있는 것으로 판단된다. 실제 에너지/환경 분야의 연구들은 개발/성장과 관련된 논의가 함께 진행되는 특징이 반영된 결과로도 해석할 수 있다.

넷째, 사회과학분야로의 도시연구가 확대되고 있어 공간과 장소에 대한 관심이 증대되고 있음을 알 수 있다. 특히 외교/역사, 담론/문학에 해당되는 주제들은 문화/관광보다도 높은 비율을 차지하고 있고 증가세를 보이고 있어 도시관련 주제가 계획분야 외에 사회과학전체 분야로 확대되고 있음을 알 수 있다.

본 연구는 국내 전체 도시분야 연구를 대상으로 데이터마이닝 기법을 적용하여 연구주제의 연관성과 경향을 함께 분석하였다는데 의의가 있다. 그러나 분석자료 구득의 한계로 국내연구에만 한정하여 국제적인 도시연구동향까지 분석은 실행하지 못했다. 향후 국외 연구주제를 연구대상에 포함하여 국내와 국외의 비교, 시기별 연구주제별 논문 발표의 시차 등에 대한 분석이 수행될 수 있다면 연구주제의 시의성과 적절성을 판단하는데 도움 되는 결과를 도출할 수 있을 것이다. 그리고 기계적인 과정을 거쳐 키워드를 도출하다보니 논문 내용과 직접적인 연관된 분야보다는 일반적인 키워드들이 우선적으로 선정되는 한계가 있었다. 예를 들어 분석에 활용된 키워드들 중 도시계획, 도시설계는 일반적인 연구분야를 통칭하는 경우가 많기 때문에 실제 논문상의 구체적인 내용들은 연관도가 낮을 수 있다. 후속연구에서는 지칭하는 대상과 범위가 넓은 일반적인 주제어를 선처리할 수 있는 방법론 개발하여 동향분석을 수행하는 연구를 제안한다.

References

- [1] J. W. Hur, J. S. Choo, "A critical survey on regional studies in Korea", *Journal of the Korea Planning Association*, 28(2), pp.107-119, 1993.
- [2] K. Kim, Y. W. Nam, "Gentrification: Research trends and arguments", *Journal of the Korea Planning Association*, 33(5), pp.83-97, 1998.
- [3] B. H. Ko, "A study on tile paradigm shift and planning

- issues in urban and regional planning”, *Journal of the Korean Urban Management Association*, 13(2), pp.101-120, 2000.
- [4] H. M. Bae, “Research trend on the landscape area examined through the Journal Analysis in Korea”, *Journal of the Korea Planning Association*, 42(3), pp.179-190, 2007.
- [5] Y. S. Kim, C. Y. Park, “Types and trends of indicators in the study of urban landscape planning”, *Journal of the Korea Planning Association*, 41(5), pp.117-129, 2006.
- [6] D. M. Blei, A. Y. Ng, M. I. Jordan, “Latent dirichlet allocation”, *Journal of machine Learning research*, 3(jan), pp. 993-1022, 2003.
- [7] J. M. Park, J. H. Kim, S. J. Yun, “International Research Trends and Domestic Policymakers Discourse with regards to Urban Regeneration - Focusing on Semantic Network Analysis and Topic Modeling”, *Journal of the Korean Urban Management Association*, 31(2), pp. 31-59, 2018.
DOI : <https://doi.org/10.36700/KRUMA.2018.06.31.2.31>
- [8] Y. M. Song, E. S. Lee, “A Relation between Policy and Research in the Korean Architecture & Urban Field based on a Keyword Network Analysis”, *Journal of the Urban Design Institute of Korea Urban Design*, 19(6), pp. 83-92, 2018.
DOI : <https://doi.org/10.38195/judik.2018.12.19.6.83>
- [9] H. G. Sung, I. K. Park, K. H. Ji, J. Y. Kim, K. S. Jang, J. H. Kim, J. S. Ko, C. H. Jin, “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*, pp. 3-26, 2019.
DOI : <https://doi.org/10.15793/kspr.2019.103..001>
- [10] S. Y. Lim, Y. M. Lim, J. Y. Lee, “Study on the Trends of U-City and Smart City Researches using Text Mining Technology”, *Journal of the Korean Society for Geospatial Information Science*, 22(3), pp. 87-97, 2014.
DOI : <https://doi.org/10.7319/kogsis.2014.22.3.087>
- [11] K. C. Park, C. H. Lee, “A Study on the Research Trends for Smart City using Topic Modeling”, *Journal of Internet Computing and Services*, 20(3), pp. 119-128, 2019.
DOI : <http://dx.doi.org/10.7472/ikcii.2019.20.3.119>
- [12] H. S. Park, D. H. Kim, S. J. Chang, “Research Trend Analysis on Smart City based on Structural Topic Modeling(STM)”, *Journal of Digital Contents Society*, 20(9), pp. 1839-1846, 2019.
DOI : <https://doi.org/10.9728/dcs.2019.20.9.1839>
- [13] K. H. Jo, S. Y. Lee, “A Study on Regeneration Methods of Old Downtown Through Analyzing Recent Trends of Establishing Cultural Place Characteristics”, *Journal Of the Architectural Institute of Korea Planning & Design*, 32(8), pp. 123-130, 2016.
DOI: https://doi.org/10.5659/JAIK_PD.2016.32.8.123
- [14] Y. J. Kim, M. J. Kim, “Identifying Urban Regeneration Research Trend”, *Journal of the Korean Institute of Interior Design*, 28(5), pp. 51-59, 2019.
DOI : <https://doi.org/10.14774/KIID.2019.28.5.051>
- [15] B. H. Jeong, J. W. Kim, “Topic Modeling of the Transition of the Sense of Place in Urban Regeneration Area: A Case Study of Daegu Bangcheon Market”, *Journal of Daegu Gyeongbuk Studies*, 19(1), pp. 27-44, 2020.
- [16] H. K. Cho, Y. J. Jun, S. H. Ha, J. H. Choi, “Literature Review on the Research of Urban and Architectural Landscape using Eye Tracking Method”, *Proceedings of the KAIS Spring Conference 2015*, The Korea Academia-Industrial cooperation Society, pp. 205-207, 2015.
- [17] B. R. Yang, J. J. Ryu, “Analysis on The Academic Research Trend of The City Image in Korea”, *Journal of the Korea Landscape Council*, 7(1), pp. 17-28, 2015.
- [18] K. H. Byun, C. G. Yoo, J. R. Lee, E. J. Kang, “Characteristics on Literatures of Architecture and Urban Design related to Tourism”, *Journal of the Regional Association of Architectural Institute of Korea*, 16(1), pp. 149-158, 2014.
- [19] S. Y. Lee, S. E. Yoo, “Analysis of Domestic Research Trend on Remodeling - Focused on Design and Planing Research”, *Journal of the Korea Institute of Spatial Design*, 14(2), pp. 167-178, 2019.
DOI : <https://doi.org/10.35216/kisd.2019.14.2.167>
- [20] I. S. Kang, J. W. Moon, J. C. Park, “Recent Research Trends of Artificial Intelligent Machine Learning in Architectural Field - Review of Domestic and International Journal Papers”, *Journal Of the Architectural Institute of Korea Structure & Construction*, 33(4), pp. 63-68, 2017.
DOI : https://doi.org/10.5659/JAIK_SC.2017.33.4.63
- [21] J. E. Park, K. H. Lee, “An Analysis of Domestic Text Mining Research Trend in Architectural and Urban Planning”, *Spring Annual Conference of AIK 2019.4*, Architectural Institute of Korea, 39(1), pp. 110-113, 2019.
- [22] B. Johnston, C. Kruger, A. Jones (trans. K. B. Jo.), *Applied Unsupervised Learning with Python: Discover hidden patterns and relationships in unstructured data with Python*, p.528, Packt Publishing Ltd (acorn publishing Co. in Korea), 2019, p. 256.
- [23] J. H. Park, M. Song, “A Study on the Research Trends in Library & Information Science in Korea using Topic Modeling”, *Journal of the Korean Society for information management*, 30(1), pp. 7-32, 2013.
DOI : <https://doi.org/10.3743/KOSIM.2013.30.1.007>
- [24] D. M. Blei, “Probabilistic topic models”, *Communications of the ACM*, 55(4), pp. 77-84, 2012.
DOI: <https://doi.org/10.1145/2133806.2133826>
- [25] Korea Citation Index, [Internet]. National Research Foundation of Korea, [cited 2020 December 14], Available From: <https://www.kci.go.kr> (accessed Nov. 12, 2020)

장 선 영(Sun-Young Jang)

[정회원]



- 2010년 2월 : 건국대학교 건축대학 건축설계전공 (공학사)
- 2014년 2월 : 성균관대학교 대학원 건축학과 (건축학석사)
- 2019년 8월 : 성균관대학교 대학원 건축학과 (건축학박사)
- 2019년 9월 ~ 현재 : 한국건설기술연구원 박사후연구원

〈관심분야〉

공간경험, 증강공간, 건축문화유산, Digital Heritage, Ontology, Content Management System (CMS)

정 승 현(Seunghyun Jung)

[정회원]



- 2009년 8월 : 한양대학교 대학원 도시공학과 (공학박사)
- 2009년 9월 ~ 2011년 10월 : 한양대학교 BK21사업단 박사후연구원
- 2011년 11월 ~ 2012년 12월 : (재)인천발전연구원 초빙연구위원
- 2012년 12월 ~ 현재 : 한국건설기술연구원 수석연구원

〈관심분야〉

스마트시티, 도시계획 및 설계