

토픽 모델링을 활용한 스마트시티 연구동향 분석[☆]

A Study on the Research Trends for Smart City using Topic Modeling

박 건 철¹ 이 치 형^{2*}
Keon Chul Park Chi Hyung Lee

요 약

본 연구의 목표는 스마트시티에 대한 연구 동향을 파악하고 이를 바탕으로 스마트시티에 대한 정책적·산업적·학술적 방향성을 제시하는 데 있다. 산업화와 더불어 급격하게 진행된 도시화 및 도시인구 증가로 세계 주요 도시들은 대부분 교통, 환경, 주거 등 다양한 도시문제에 직면하고 있다. 세계 각 도시는 이러한 도시문제를 해결하여 시민 삶의 질을 향상시킨다는 공통된 목표를 추구하기 위해 스마트시티의 도입을 서두르고 있다. 하지만 스마트시티에 대한 다양한 개념적 접근은 정책 목표 설정 및 추진전략 수립을 위한 방향성 도출에 불확실성을 야기하고 있다. 본 연구는 이러한 문제 인식하에 스마트시티 정책에 대한 방향성을 제시하고자 Scopus DB 및 Springer DB에서 스마트시티와 관련된 학술논문 11,527건의 제목과 초록, 발행연도 등의 정보를 수집하여 연구현황, 연구주제, 연구분야 차이 등을 LDA기반 토픽모델링 기법을 활용하여 분석하였다. 분석결과, 스마트시티 관련 연구주제는 크게 서비스 및 애플리케이션 분야, 기술 분야, 시민·사용자 관점의 8가지 세부주제로 유형화되었으며, 이중 '시민중심 스마트시티 추진을 통한 지속가능성의 확보'와 관련된 주제가 가장 많이 언급된 주제로 분석되었다. 또한 주제 간 연관관계를 분석한 결과, 데이터와 프라이버시 관련 연구의 중심성이 가장 높게 나타났다. 이는 데이터가 전반적인 스마트시티의 공동기반으로서 역할을 하며 기술, 서비스, 정책과 관련된 다양한 의사결정에 영향을 미치는 한편, 위치정보 등 개인 정보가 수집되는 과정에서 프라이버시 침해 등이 잠재적인 위협요인이 될 수 있음을 암시한다. 스마트시티 연구현황에 대한 객관적 분석을 통해 스마트시티 개념의 발전 방향을 이해하고 향후 대응방안을 모색함으로써, 스마트시티 추진 및 연구에 정책적, 산업적, 학술적 근거자료를 제공할 수 있을 것으로 기대된다.

☞ 주제어 : 스마트시티, 연구동향분석, 텍스트마이닝, 토픽모델링, 잠재디리클레할당

ABSTRACT

This study aims to analyze the research trends on Smart City and to present implications to policy maker, industry professional, and researcher. Cities around globe have undergone the rapid progress in urbanization and the consequent dramatic increase in urban dwellings over the past few decades, and faced many urban problems in such areas as transportation, environment and housing. Cities around the globe are in a hurry to introduce Smart City to pursue a common goal of solving these urban problems and improving the quality of their lives. However, various conceptual approaches to smart city are causing uncertainty in setting policy goals and establishing direction for implementation. The study collected 11,527 papers titled "Smart City(cities)" from the Scopus DB and Springer DB, and then analyze research status, topic, trends based on abstracts and publication date(year) information using the LDA based Topic Modeling approaches. Research topics are classified into three categories(Services, Technologies, and User Perspective) and eight regarding topics. Out of eight topics, citizen-driven innovation is the most frequently referred. Additional topic network analysis reveals that data and privacy/security are the most prevailing topics affecting others. This study is expected to helps understand the trends of Smart City researches and predict the future researches.

☞ keyword : Smart City, Research Trend Analysis, Text Mining, Topic Modeling, Latent Dirichlet Allocation(LDA)

1. 서 론

UN Habitat 보고서에 따르면 현재 전 세계 인구의 약 54%(약 39 억 6천만 명)가 도시에 거주하고 있으며 이러한 수치는 2050년 약 70%를 상회할 것으로 전망되고 있다[1]. 도시화의 급속한 진전과 도시 거주 인구의 증가에 따라 교통, 환경, 주거 등과 관련한 각종 사회문제가 도시를 중심으로 확산 및 심화되고 있으며, 이에 따라 도시의 지속가능성을 확보하고 도시민의 삶을 질적으로 개선하

¹ Dept. of Policy Research, Seoul Digital Foundation, Seoul, Korea

² Department of Advertising and Public Relations, Pyeongtaek University, Pyeongtaek, Korea

* Corresponding author (chilee@gmail.com)

[Received 07 April 2019, Reviewed 25 April 2019(R2 24 May 2019), Accepted 03 June 2019]

☆ 본 논문은 2018년도 한국인터넷정보학회 추계학술발표대회 우수 논문 추천에 따라 확장 및 수정된 논문임.

Smart				City	
Structure	Functions	Focus	Semiotics	Stakeholders	Outcomes
Architecture	[to] Sense	[+] Cultural	[-] Data	Citizens	[to] Sustainability
Infrastructure	Monitor	Economic	Information	Professionals	QoL
Systems	Process	Demographic	Knowledge	Communities	Equity
Services	Translate	Environmental		Institutions	Livability
Policies	Communicate	Political		Businesses	Resilience
Processes		Social		Governments	
Personnel		Technological			
		Infrastructural			

Illustrative Components (total components = 7*5*8*3*6*5 = 25,200):
 Architecture to sense economic information by/from citizens for QoL.
 Systems to process environmental data by governments for livability.
 Policies to communicate technological knowledge by professionals for resilience.
 Processes to translate political information to citizens for sustainability.

(그림 1) Ramaprasad 외(5)의 스마트시티 개념 정의
 (Figure 1) Conceptual Definition of Smart City by Ramaprasad et al.(5)

는 방법으로서 스마트시티가 부상하고 있다[2].

해외 주요 선진도시들은 각 도시가 처한 사회·경제·문화적 환경에 따라 방법은 상이할지 모르지만 기본적으로 도시문제를 해결하여 시민 삶의 질을 향상시킨다는 공통된 목표를 실현하기 위해 스마트시티를 경쟁적으로 추진하고 있다. 이에 따라 스마트시티 관련 시장 규모는 2016년 368억 달러에서 2025년 887억 달러로 급증할 것으로 전망되고 있다[3].

스마트시티(Smart City)가 내포한 개념을 구조적으로 살펴보면 ‘Smart’의 의미는 스마트 기술을 포함한 다양한 ICT기술을 의미할 뿐 아니라 이러한 기술을 활용한 새로운 방식의 접근체계를 의미한다. 또한 ‘City’의 경우 스마트한 기술과 접근체계가 적용되는 공간적 범위를 의미할 뿐 아니라 이러한 기술과 새로운 접근을 도입하는 궁극적 목표, 즉 해당 도시가 처한 사회·경제·문화적 환경에 따른 도시문제의 정의와 경제적 성장의 주안점이 된다고 할 수 있다.

한편 ITU-T[4]는 스마트시티에 대한 개념정의를 위해 전 세계 국가 및 유관기관에서 정의한 스마트시티 개념을 조사하였는데, 당시 전 세계적으로 스마트시티에 대한 정의는 116개에 달하는 것으로 조사되었다[4]. 뿐만 아니라 그림 1에서 보는 바와 같이 Ramaprasad 외[5]는 스마트시티를 ‘스마트’를 구성하는 기반구조, 기능, 목표, 핵심요소와 ‘도시’를 구성하는 이해관계자 및 도시의 성과를 바탕으로 총 25,200(7*5*8*3*6*5)개의 하위 개념으로 정의했다.

이러한 개념적 다양성은 각 도시가 처한 지정학적 상황에 따라 특화된 도시전략을 구축하는 데에는 유용할지 모르지만, 관련 기술의 개발 및 표준화, 산업의 확산 및 융합 등 산업적·정책적 관점에서는 방향성 부재에 따른

불확실성을 야기하고 있다. 따라서 이러한 시점에 스마트시티에 대한 연구 동향을 파악하여 그간의 연구 흐름에 대한 이해와 함께 향후 발전 방향을 전망하는 것이 시급한 상황이다.

이에 본 연구에서는 Scopus DB와 Springer DB에서 “Smart City” 키워드를 바탕으로 검색된 문헌의 제목과 초록을 수집하여 텍스트마이닝 기법 중 하나인 토픽 모델링 분석방법을 적용하여 스마트시티와 관련된 연구 주제를 유형화하였다. 또한 이들 주제의 시기별 변화 트렌드를 분석함으로써 전반적인 스마트시티에 대한 연구 동향을 파악하고자 한다. 이를 통해 학술적 관점에서는 스마트시티에 대한 향후 연구 방향 수립에 기여하며, 정책적·산업적 관점에서는 성공적인 스마트시티 전략 수립을 위한 방향성을 제시할 것으로 기대된다.

2. 관련 연구

2.1 스마트시티

2.1.1 스마트시티의 개념 변화

스마트시티는 1990년대 중반 정보인프라를 도시전역에 확대하여 도시효율화를 추구하는 목적으로 시작되었다[6]. 유시티(U-city) 또는 디지털 시티(Digital city)라는 명칭으로 시작된 당시의 스마트시티에 대한 접근은 주로 통신인프라의 구축 등 대규모 기술의 적용과 확대가 핵심목표로 추진되었다. 하지만 최근 추진되고 있는 스마트시티의 경우 단순히 기술적 적용과 확대에만 초점을 두는 것이 아니라 이러한 기술이 도시문제의 해결과 시민의 생활을 개선하는데 활용되는 실용적 측면을 강조하며

변화하고 있다[7]. 따라서 현재 최근 세계 각 도시에서 추진하는 스마트시티의 공통된 모습은 과거 공급자 중심의 하향식 거버넌스로 추진되었던 한계를 극복하여 수요자 중심으로 진화하고 있으며 스마트시티는 기술을 기반으로 도시의 여러 이해당사자를 조정하는 도시플랫폼으로서의 의미도 내포하고 있다[8]. 과거와 비교하여 최근 발전하고 있는 스마트시티 개념은 다음 표 1과 같다[2].

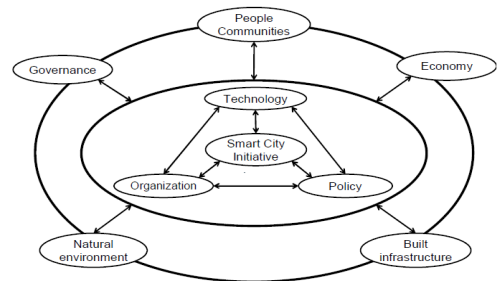
(표 1) 과거와 현재의 스마트시티의 개념 차이(2)
(Table 1) The difference between the past and the present in smart city(2)

구분	과거	현재
유사 개념	· Digital City, U-City 등	· Smart (Sustainable) City, Intelligent City
핵심 기반	· 통신망 등 기술중심	· 사회적 자산 등 기능중심
해결 방식	· 물리적 투입량 증가 · 도시문제 해결을 위한 신규 인프라 확대 (예: 교통체증→도로건설)	· 창의·지능·신기술 · 기존 인프라의 지능적 활용 (예: 교통체증→신호시스템 조정)
추진 목표	· 도시 정보화(효율화)	· 도시 지능화(가용성)
해결 과정	· (사전정의) 프로세스 기반	· (상황인식) 데이터 인지기반
추진 체계	· (도시중심) 공공의 Top-down 방식 · 개인인프라 silo에 한정(수직)	· (시민중심) 다양한 이해관계자, 특히 시민이 참여하고 주도하는 Bottom-up 방식 · 시스템 간 연계와 지능화(수평)
적용 / 운영	· 통신, 방법, 방재 등 기본 인프라를 신도시에 도입 · 도시 내에서 기능별로 분절적 운영, 도시데이터 공유 불가, 시민이 도시운영체계에 적용	· 행정, 교통, 에너지, 물관리, 복지, 환경 등 다양한 기능을 신도시 외 기존도시에도 적용 · 도시 전체가 플랫폼으로 연결, 도시데이터 공유로 단절없는 시민 맞춤형 서비스 제공

2.1.2 스마트시티의 구성요소

일반적으로 스마트시티의 구성요소는 기술적, 인적, 제도적 부분으로 구분된다[9]. 먼저 기술적 측면에서는 도시를 구성하는 다양한 사물(Node)과 사용자간의 원활한 상호작용을 지원하는 유무선 통신 환경과 센서기술

(사물인터넷; IoT), 이를 통해 도시데이터를 확보하며 이를 일련의 과정을 통해 도시에서 필요로 하는 가치로 변환할 수 있는 데이터 수집·처리 기술 등이 중요한 요소로 부각되고 있다. 인적 측면의 경우 사회적 공공성과 문화적 다양성을 담보하기 위한 방법론으로서 시민의 참여과정을 양질로 보장하는 참여정당성에 대한 부분이 중요하게 작용하며, 시민의 아이디어와 정책수요를 정책화하려는 노력이 추진되고 있다[10]. 마지막으로 제도적 요소는 추진체계, 정책, 규제 명령 등이 포함되는데, 특히 혁신을 방해하는 규제를 철폐하고 부서 간의 벽을 허무는 조직적 관점뿐 아니라 앞선 시민참여를 제도적으로 활성화하기 위한 시민 주도의 커뮤니티, 민·관 협력 구조를 강조하고 있다. Chourabi[11]는 스마트시티의 구성요소를 더욱 세분화하여 조직, 기술, 거버넌스, 정책적 맥락, 사람과 커뮤니티, 경제, 경제, 도시 인프라, 천연자원 등 8개로 구분하고 이들 요소 간의 상관관계를 바탕으로 그림 2와 같이 스마트시티 프레임워크를 제안하였다. 여기서 내부에 위치한 기술, 조직, 정책은 스마트시티를 추진하는 가장 핵심 직접요인으로 작용하며, 외부의 요인들은 내부 핵심 요인의 활성화를 통해 스마트시티에 요소로 작용하는 간접적 역할을 수행한다.



(그림 2) 스마트시티 추진 프레임워크
(Figure 2) Smart City Initiatives Framework

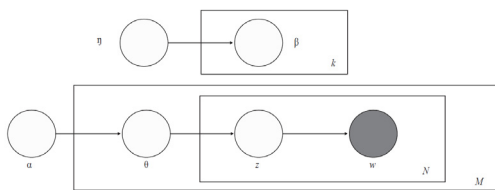
3.2 토픽모델링

토픽모델링은 문서를 이루고 있는 키워드들을 바탕으로 문서에서 주제(Topic)을 도출하기 위해 사용되는 통계적 분석기법으로 주로 비정형 텍스트 분석에 많이 사용되고 있다. 토픽모델링은 확률적 기반으로 구성된 주제별 키워드 분포를 바탕으로 다시 주어진 문서에서 발견된 키워드 분포를 분석함으로써 해당 문서에 사용된 주제들을 추정한다. 이를 통해 문서들을 주제별로 분류할 수 있으며, 또한 문서들이 어떤 주제들을 함께 다루고 있는지

예측이 가능하고 시간에 따라 주제들의 동향 파악과 신규 문서가 어느 주제에 대한 것인지 등을 추정할 수 있는 특징이 있다[12]. 이러한 특징으로 토픽모델링은 특히 특정주제와 관련된 연구 동향을 파악하는데 유용하게 활용되고 있다.

Griffiths와 Steyvers[13]는 마코브 체인(Markov Chain)과 몬테카를로(Monte Carlo)알고리즘을 기반으로 문서에서 주제와 주제를 이루는 단어의 확률분포를 유추하는 추정모형을 제시하였으며, 이를 PNAS 저널에 수록된 논문들에서 수집된 초록정보에 적용하여 PNAS저널의 연구 주제를 그룹화하고 이들 주제의 동향 변화를 분석하였다. Jeong과 Song[14]은 토픽모델링을 통해 Computer Science와 Medical Science 분야에서 수집한 논문, 특히, 웹 뉴스 등의 데이터를 분석하여 연구동향을 파악하였다.

한편 토픽모델링을 적용하기 위한 기법 중 가장 대표적인 알고리즘은 디리클레 분포를 이용하여 주어진 문서에 숨겨져 있는 주제들을 추론하는 확률모델 알고리즘인 LDA(Latent Dirichlet Allocation)이다[12, 15, 20]. LDA는 각 주제에서 도출할 수 있는 단어들의 확률 분포를 알고 있을 때, 무작위 과정(random process)에 의해 문서가 생성될 수 있다고 가정하는 모형으로서 하나의 문서는 여러 주제로 구성되고, 문서의 주제 분포에 따라 단어의 분포가 결정된다는 가정 하에 단위 문서가 생성되는 과정을 그림 3과 같이 가정한다. 이를 통해 문서 전체의 주제, 각 문서별 주제 비율, 그리고 각 주제에 포함될 단어들의 분포를 알아낼 수 있다.



(그림 3) LDA의 그래프 모형[12]

(Figure 3) Graphical Model Representation of LDA[12]

조성배 외[16]는 LDA기반의 토픽모델링을 이용하여 Web of Science에서 수집된 1,788개의 논문을 분석하여 ‘개방형 혁신’과 관련된 연구 동향을 분석하였다. 아울러 이원상과 손소영[17] 역시 LDA 기반의 토픽모델링을 통해 Scopus DB에서 “Spatial Big Data”라는 키워드로 검색된 전체문헌의 초록을 분석하여 공간빅데이터의 글로벌 연구 동향을 파악하였다.

3. 데이터 및 분석 방법

3.1 데이터 수집

본 연구에서는 스마트시티 연구 동향을 분석하기 위해 Scopus DB와 Springer DB를 통해 “Smart City(ies)”를 관련 키워드에 포함하고 있는 논문들을 모집단으로 구축한 후 제목과 초록, 발행연도 등의 정보를 수집하여 스마트시티에 대한 그동안의 연구동향을 분석하고자 한다.

Springer DB의 경우 2008년부터 현재(2019년 3월)시점까지 총 1,219편의 스마트시티 관련 논문이 수집되었으며, Scopus DB의 경우 2008년부터 현재(2019년 3월)까지 총 10,318건의 스마트시티 관련 논문이 수집되었다. 이중 중복된 논문 10건을 제외하고 총 11,527개의 논문을 분석에 활용하였다.

(표 2) 연도별 발행편수

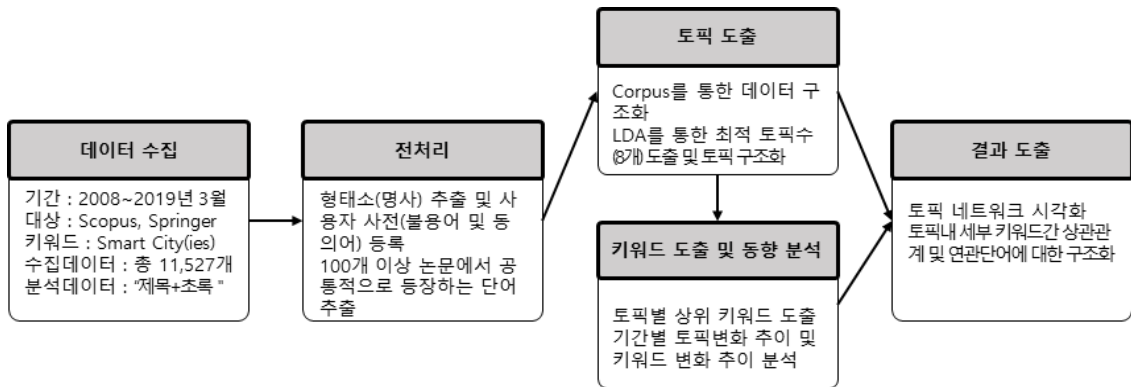
(Table 2) Publications by Year

연도	편수	연도	편수
2000	17	2011	82
2001	1	2012	146
2003	1	2013	291
2004	1	2014	521
2005	1	2015	770
2006	5	2016	1231
2007	5	2017	3630
2008	8	2018	3858
2009	20	2019	903
2010	25	2020	11

3.2 분석과정

먼저 수집된 11,527개의 논문의 제목과 초록 컬럼을 병합한 후 Python 3.4를 통해 전처리 과정을 진행하였다. 전처리 과정은 우선 대문자를 소문자로 변경(Lowercase)하고 띄어쓰기 공백과 모든 구두점을 제거하여 연속하는 단어의 열로 분할하는 토큰화(Tokenize)작업을 수행하였다. 또한 문서에서 빈번하게 사용되는 전치사, 관사, 접속사, 특수 문자 등 분석에 불필요한 불용어(Stopwords)를 제거하고 동일 의미 단어를 하나의 단어로 변환하는 어간추출(Stemming) 작업을 수행하였다.

아울러 본 연구는 스마트시티와 관련된 모든 주제를



(그림 4) 데이터 분석 프레임워크
(Figure 4) Data Analysis Framework

파악하는 것보다 공통적으로 나타나는 핵심적인 연구의 흐름을 파악하는 것이 주된 목적이기에 최소 100개 이상의 논문에서 공통적으로 등장하는 키워드들만으로 분석 단위를 한정하였다. 이러한 과정을 통해 애초 76,169개로 구성된 단어노드를 883개로 재구성하였다.

일련의 정제과정이 이루어진 텍스트 데이터는 LDA모형을 활용하여 토픽모델링 분석을 진행하였다. LDA를 통한 토픽의 유형화 및 토픽 간 네트워크 관계의 시각화는 NetMiner 4.2를 활용하여 진행하였으며, 비지도학습(unsupervised learning)과정을 수차례 반복하여 최종적으로 8개의 스마트시티 관련 주제군과 주제별 핵심키워드를 도출하였다.

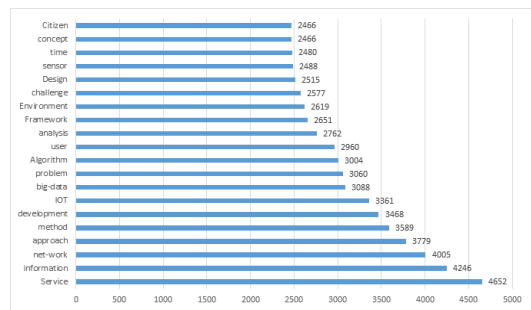
4. 분석 결과

4.1 스마트시티 관련 핵심 키워드

먼저, 분석에 활용된 전체 단어노드에 대한 빈도를 분석하여 스마트시티와 관련된 핵심 키워드를 도출하였다. 전처리 과정을 통해 정의된 총 874개의 단어노드 중 빈도수를 분석한 결과 스마트시티와 관련하여 연구자들이 가장 많이 언급한 키워드는 “service”로 나타났다. 아울러 “network”, “IoT”, “big-data”, “algorithm”과 같은 기술중심의 키워드들이 상위에 위치한 가운데, “citizen”, “process”, “design”과 같은 참여주체 및 추진과정에 대한 내용도 상위에 언급되고 있음을 알 수 있다. 스마트시티 관련 핵심 키워드를 바탕으로 도식화한 워드클라우드와 상위 20개 단어의 빈도표는 다음 그림 5 및 그림 6과 같다.



(그림 5) 스마트시티 워드클라우드
(Figure 5) Word-cloud for Smart City



(그림 6) 가장 빈번히 언급된 상위 20개 단어
(Figure 6) Top 20 Words Most Frequently Cited

4.2 스마트시티 연구주제 도출

다음으로, NetMiner 4.2의 LDA 패키지를 통해 스마트 시티와 관련된 총 8개의 연구주제를 도출하였다. 주제의 도출을 위해 활용한 기준은 TF-IDF 가중치를 활용하였으며 전체 단어노드의 평균 TF-IDF는 284.40으로 나타났다.

(표 3) 토픽 모델링을 통해 도출된 스마트시티 연구주제 및 주제별 키워드

(Table 3) Smart City Research Topics and Keywords by Topics derived from Topic Modeling

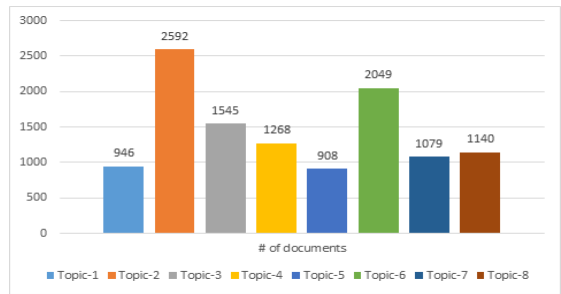
주제	키워드
Topic-1 : 건축물/모니터링	building, construction, water, area, method, analysis, monitoring, design, authors, development 건축물, 건설, 물, 지역, 방법론, 분석, 모니터링, 설계, 작성자, 개발
Topic-2 : 시민중심혁신/지속가능성	development, citizen, concept, project, challenge, innovation, policy, service, approach, sustainability 개발, 시민, 개념, 프로젝트, 문제, 혁신, 정책, 서비스, 접근 방식, 지속 가능성
Topic-3 : 빅데이터/알고리즘	algorithm, method, performance, big-data, problem, approach, image, time, accuracy, feature 알고리즘, 방법, 성능, 빅데이터, 문제, 접근 방식, 이미지, 시간, 정확도, 특징
Topic-4 : 프라이버시/사용자 위치정보	user, information, service, person, citizen, sensor, platform, privacy, location, event 사용자, 정보, 서비스, 사용자, 시민, 센서, 플랫폼, 개인 정보, 위치, 이벤트
Topic-5 : 에너지/스마트그리드	energy, power, smart-grid, energy-efficiency, energy consumption, control, load, cost, optimization, demand 에너지, 전력, 스마트 그리드, 에너지 효율성, 에너지 소비, 제어, 부하, 비용, 최적화, 수요
Topic-6 : 사물인터넷/클라우드	iot, service, net-work, device, architecture, cloud, sensor, security 사물인터넷, 서비스, 통신망, 장치, 아키텍처, 클라우드, 센서, 보안
Topic-7 : 데이터/의사결정지원	big-data, design, information, development, process, framework, decision-making, platform, management, approach 빅데이터, 설계, 정보, 개발, 프로세스, 프레임워크, 의사결정, 플랫폼, 관리, 접근 방식
Topic-8 : 교통문제/효율성	vehicle, traffic, road, net-work, transportation, parking, mobility, bus, traffic congestion, time 차량, 교통, 도로, 통신망, 교통, 주차, 이동성, 버스, 교통 혼잡, 시간

본 연구에서는 TF-IDF 값 0.1을 임계치로 설정하여 이보다 큰 가중치 값을 갖는 단어들을 LDA 분석에 활용하였으며[18, 19], 반복된 비지도학습 과정(unsupervised learning)을 통해 8개의 최적 주제군을 도출하였다.

LDA 분석을 통해 도출되는 결과는 주제별로 군집화되는 키워드들의 조합으로 나타나는데, 이러한 키워드들의 조합을 바탕으로 워드네트워크(Word-Network)분석을 수행하여 단어 조합 간 가중치를 분석하였으며, 상위 가중치 조합을 바탕으로 스마트시티 관련 연구를 5건 이상 수행한 2명의 전문가를 대상으로 심층인터뷰를 통해 각 주제별 키워드를 대표하는 주제에 대한 명명(Naming)을 수행하였다.

그림 7과 같이, 도출된 8개의 주제 중 가장 많은 비중을 차지하는 주제는 “Topic-2 : 시민중심혁신/지속가능성 (2,592건, 22%)”였으며, 다음으로 “Topic-6 : 사물인터넷/클라우드(2,049, 18%)”, “Topic-3 : 빅데이터/알고리즘 (1,545, 13%)”, “Topic-4 : 프라이버시/사용자 위치정보 (1,268, 11%)”, “Topic-8 : 교통문제/효율성(1,140, 10%)”, “Topic-7 : 데이터/의사결정지원(1,079, 9%)”, “Topic-1 : 건축물/모니터링(946, 8%)”, “Topic-5 : 에너지/스마트그리드 (908, 8%)”의 순으로 스마트시티 관련 주제가 도출되었

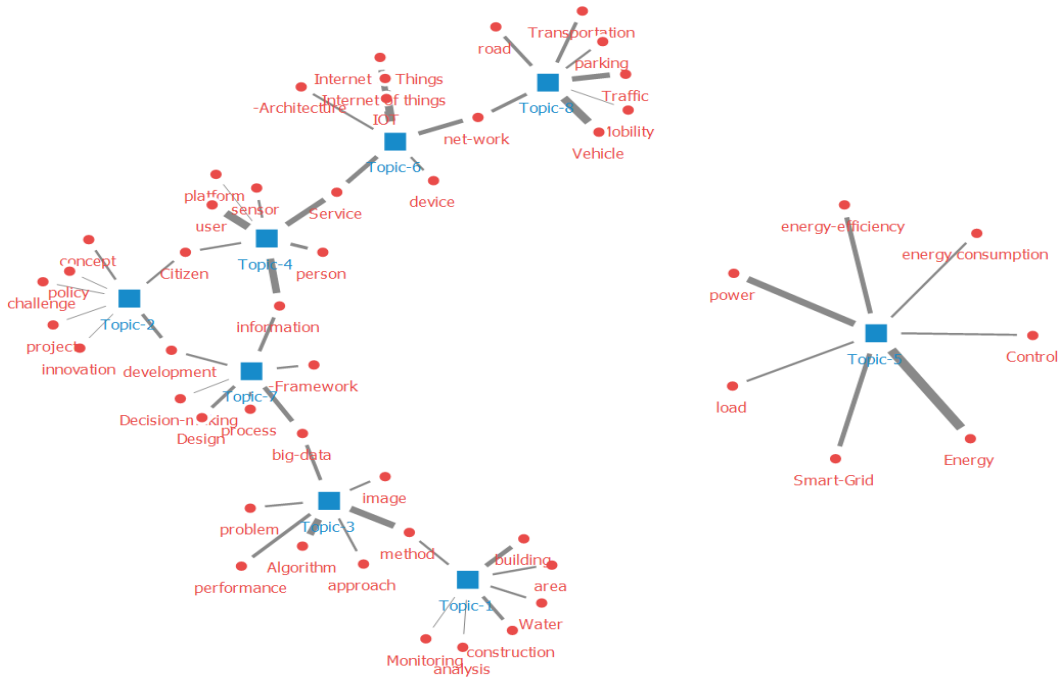
다.



(그림 7) 주제별 발행편수
(Figure 7) Publications by Topics

4.3 스마트시티 연구 동향

토픽모델링을 통해 도출된 8개 주제에 대한 연구동향을 파악하기 위해 먼저 8개의 주제로 군집하는 주제별 세부 키워드간의 네트워크 관계를 구성하였다. 네트워크관계는 각 세부 키워드들 간의 문서 내 동시출현 빈도를 바탕으로 단어와 단어 간의 거리를 계산한 후, 이를 행렬화하여 단어와 단어 간의 관계에 대한 가중치를 부여하여 도출 하였다. 이를 통해 전반적으로 스마트시티와 관련된



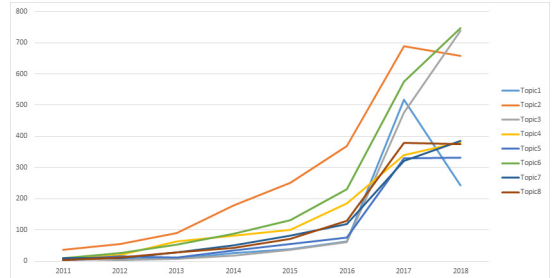
(그림 8) 스마트시티 연구 주제 네트워크

(Figure 8) Topic Network of Smart City Research

8개 주제들의 상호 연관관계를 도식화 하였으며 그 결과는 그림 8과 같다. 각 주제의 중심에 연결된 단어 노드들은 각 주제를 구성하는 핵심키워드를 의미하며 여기서 노드의 굵기와 진하기는 주제간 연관성 정도를 나타낸다. 도식화된 주제간 연관관계는 각 주제별로 상위 7개 단어를 중심으로 구성하였으며, 실제로는 각 주제에 속하는 수천개의 단어가 서로 연결관계를 이루고 있다. Topic-5 : 에너지/스마트그리드의 경우 타 주제들과의 연관관계가 높지 않고 독립된 군집을 이루고 있는 반면 나머지 7개 주제들은 서로 상호 연결 관계를 갖는 것으로 나타났다. 특히, 이 중 주제의 비중과 관련 없이 Topic-4 : 프라이버시/사용자 위치정보와 Topic-7 : 데이터/의사결정지원 분야는 가장 높은 연결중심성을 보여 다른 여러 주제에 포괄적으로 나타나는 문제로 인식할 수 있다.

다음으로 연구동향의 동적 변화를 파악하기 위해 그림 9와 같이 8개 주제와 관련된 논문의 발행편수를 연도별로 분석하였다. 본격적으로 스마트시티와 관련된 연구논문이 급증하기 시작한 2011년부터의 동향을 살펴보면 Topic-2 : 시민중심혁신/지속가능성과 관련된 연구가 계속해서 가장 많은 부분을 차지하여왔으나, 최근에는

Topic-5 : 에너지/스마트그리드, Topic-3 : 빅데이터/알고

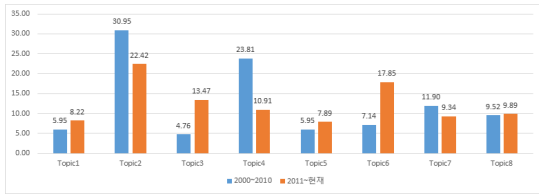


(그림 9) 연도별 주제 변화 추이

(Figure 9) Trends in Topic Change by Year

리즘과 관련된 논문의 비중이 급증하는 것을 알 수 있다. 아울러 연구 초기인 2000년부터 2010년까지의 기간(1기) 동안 주제별 연구 비중과 연구가 본격적으로 확산되기 시작한 2011년 이후 기간(2기) 동안의 주제별 연구 비중을 비교해보면 그림 10과 같다. Topic-1 : 건축물/모니터링, Topic-3 : 빅데이터/알고리즘, Topic-5 : 에너지/스마트그리드, Topic-6 : 사물인터넷/클라우드, Topic-8 : 교통문제/효율성의 경우 초기 1기 연구에 비해 최근 2기에 비중

이 늘었으며, 특히 Topic-3과 6은 약 10%p가까이 큰 폭으로 그 비중이 늘었다. 반대로 Topic-2 : 시민중심혁신/지속가능성, Topic-4 : 프라이버시/사용자 위치정보, Topic-7 : 데이터/의사결정지원의 경우 초기 1기 연구에 비해 최근 2기 연구의 비중이 줄어들고 있으며, 특히 Topic-4의 경우 약 13% 가까이 그 비중이 줄었다.



(그림 10) 시기에 따른 연구 비중 변화

(Figure 10) Changes in % of research over period

5. 결론 및 시사점

본 연구에서는 스마트시티 분야의 연구동향을 파악하기 위해 해외 저널 DB(Scopus 및 Springer)에서 수집한 11,527개 문헌들의 제목과 초록을 LDA 기반의 토픽모델링방법을 적용하여 분석하였다.

이를 통해 총 8개의 스마트시티 연구 주제를 도출하였으며 이들 주제를 구성하는 세부 키워드를 분석하였다. 아울러 논문에서 수집한 발행연도 등의 정보를 통해 연구 동향의 동적 변화를 시기별로 분석하였다.

토픽 모델링을 활용한 스마트시티 분야의 연구동향을 분석한 결과는 다음과 같다. 첫째, 스마트시티 관련 연구 주제는 스마트빌딩, 스마트그리드, 지능형 교통 체계 등 스마트기술의 서비스 및 애플리케이션 적용 분야와 (Topic-1, Topic-5, Topic-8), 빅데이터, 인공지능 알고리즘, 사물인터넷, 클라우드 컴퓨팅 등 기술 분야(Topic-3, Topic-6), 그리고 시민중심 도시혁신체계, 프라이버시 및 사용자 보호, 의사결정 지원 등 시민·사용자 관점 (Topic-2, Topic-4, Topic-7)의 연구분야로 구분할 수 있다.

둘째, 이러한 연구 주제들 중 지난 20여 년 간 진행된 스마트시티 관련 연구에서 연구자들의 가장 많은 관심을 받은 연구분야는 ‘시민중심 스마트시티 추진을 통한 지속가능성의 확보’와 관련된 연구들이었으며, 이를 통해 스마트시티의 성공을 담보하기 위한 핵심요소로서 시민 참여, 시민 권한증진, 수요기반 정책수립 등이 중요한 과제임을 유추할 수 있다.

셋째, 스마트시티와 관련된 연구는 2000년부터 이루어

지긴 했지만 사실상 2011년 이후 본격적으로 시작 된 것으로 판단할 수 있다. 아울러 2017년 이후 발행된 논문이 전체 연구의 약 82%를 상회하는 것으로 나타나 2000년대 초기부터 이루어졌던 스마트시티에 대한 학술적 담론들이 2017년을 기점으로 본격적으로 실무적 경험과 성공 사례를 만들어내며 관련 데이터를 양산, 스마트시티 연구의 본격적인 활성화를 촉진하였다고 판단할 수 있다.

넷째, 스마트시티 연구를 2011년을 기준으로 전기, 후기로 구분할 경우 기술적 관점에서 특히 사물인터넷 및 클라우드 컴퓨팅과 관련된 연구가 최근 들어 급증한 것으로 나타났다. 이는 2011년 이후 4G LTE의 상용화 등 무선통신시장을 중심으로 한 통신 기술의 비약적 발전과 함께 다양한 센서들이 도시전역에 확대되어 데이터를 수집하는 사물인터넷 환경이 성공적으로 정착한 결과를 반영한다고 할 수 있다. 아울러 이를 통해 최근 5G 기술의 상용화와 더불어 더욱 많은 새로운 연구 사례들이 새로이 양산 될 것으로 기대된다.

다섯째, 토픽 모델링을 통해 도출된 8개 스마트시티 연구주제들의 상호 연관관계를 분석한 결과 Topic-4와 Topic-7의 중심성이 가장 높은 것으로 나타났다. 이는 효용적 측면에서는 데이터가 전반적인 스마트시티의 공통 기반으로서 역할을 하며 다른 기술, 서비스, 정책과 관련된 모든 의사결정에서 핵심적인 역할을 하는 것으로 판단할 수 있으며, 반대로 비용적 측면에서는 위치정보와 같은 다양한 개인 정보가 수집되는 과정에서 프라이버시 침해 등의 문제가 공통적인 잠재적인 위협요인이 될 수 있음을 암시한다. 따라서 스마트시티와 관련된 제품, 서비스, 정책 추진 시 데이터의 긍정적 측면과 부정적 측면에서의 균형점을 수립하기 위한 정책적 가이드라인 또는 기준제시가 필요한 상황이다.

본 연구의 결과가 제시하는 정책적·학술적 시사점은 다음과 같다. 우선 스마트시티의 성공을 담보하기 위한 핵심요소로서 기술, 시민, 서비스 애플리케이션의 3요소가 중점이 된 전략 프레임워크를 도출할 수 있다. 아울러 특히 이중에서도 가장 많은 가중치가 부여되는 시민과 관련된 수요와 공급을 활성화 할 수 있는 정책모델로서 시민참여플랫폼, 시민역량강화, 시민수요분석, 시민참여형 도시문제 해결 리빙랩 등을 활성화 하는 전략이 필요할 것으로 판단된다. 아울러 이러한 정책우선순위의 도출은 지금까지 수년간 축적된 실제 연구정보 데이터를 기반으로 하였다라는 점에서 데이터 기반의 과학적 행정 구현의 좋은 사례로도 인용될 것으로 기대된다.

다음으로 연구 주제의 확산 시기와 당시 기술정책적

이슈를 비교할 경우 무엇보다 무선 네트워크 기술의 비약적 발전이 이를 활용하는 기술서비스에 관한 연구 확산으로 순환된 결과를 유추할 수 있다. 또한 이를 바탕으로 기술적 관점의 요소들 역시 스마트시티를 활성화하는 중요한 요소임을 알 수 있다. 최근 들어 이러한 기술적 관점의 연구들이 급속하게 증가하는 추세를 고려할 때 앞으로 5G 상용화 및 확산은 스마트시티가 고도화되는 중요한 전환점으로 작용할 것으로 기대된다.

세 번째로, 스마트시티의 개념을 유형화하거나 구성요소를 정의하는 기존 연구의 대부분은 정성적 접근을 통한 연구가 이루어졌으나, 본 연구는 광범위한 비정형 연구데이터를 수집하여 텍스트마이닝 기법의 일종인 토픽 모델링 방법과 네트워크 분석을 통해 스마트시티와 관련된 연구동향을 파악하고 주제를 유형화 하였다는 점에서 스마트시티 관련 연구의 학문적 다양성을 증진하는데 기여한다고 할 수 있다. 또한 연구주제의 시기별 변화 분석 결과는 후속 연구자들이 스마트시티와 관련된 후속 연구 주제 발굴 시 좋은 참고자료로 활용 될 수 있을 것으로 기대된다.

하지만 본 연구는 이러한 시사점과 더불어 몇 가지 한계점과 보완해야 할 부분을 내포하고 있다.

먼저, 연구과정에서 전처리를 위해 불용어 사전 등을 수기로 작성하여 어느 정도 높은 정제율을 확보하였으나, 반면 동의어 처리 결과는 상대적으로 완성도가 높지 않았다. 가령 IoT나 Internet of Things를 동의어로 처리하지 못하고 개별 분석 단위로 포함한 점은 향후 반드시 보완되어야 할 것으로 판단된다. 두 번째로 본 연구에서는 연구 주제 및 키워드간 네트워크 관계를 기초적인 수준으로만 진행하여 대략적인 주제간 연결관계정도만을 파악하였지만, 향후 word2vec과 Semantic Network Analysis 등 텍스트마이닝 및 네트워크 분석 분야의 최근 기법을 적용하여 주제와 키워드 간 더욱 심도 있는 분석이 가능할 것으로 기대한다. 마지막으로 제목, 초록, 발행연도 외에 논문에서 추가적으로 수집 가능한 다른 정보들(예: 저자 정보, 국가정보, 저널특성 등)을 활용 시 스마트시티 연구의 국가별 관점 차이, 연구자간 상호 협력 관계 및 지식 전파 경로 등 연구 동향 분석에 대한 더욱 심도 있는 분석이 가능할 것으로 예측된다.

참고문헌(Reference)

- [1] UN Habitat, "World Cities Report 2016", 2016.
<http://wcr.unhabitat.org/main-report/>
- [2] K. C. Park, M. Kang, and S. Lee, "Research on Smart City Model for Seoul", Seoul Digital Foundation, 2018.
<http://sdf.seoul.kr/web/pages/gc4491b.do?bbsFlag=View&bbsId=0034&ntId=2168&bbsTyCode=&bbsAttrbCode=&authFlag=&pageIndex=1>
- [3] Nacigant Research, Smart Cities, 2016
<https://www.navigantresearch.com/reports/smart-cities>
- [4] ITU-T, "Smart sustainable cities: An analysis of definitions", 2014.
https://www.itu.int/en/ITU-T/focusgroups/ssc/Documents/website/web-fg-ssc-0100-r9-definitions_technical_report.docx
- [5] A. Ramaprasad, A. Sánchez-Ortiz, and T. Syn, "A Unified Definition of a Smart City" International Conference on Electronic Government. Springer, Cham, 2017.
https://doi.org/10.1007/978-3-319-64677-0_2
- [6] V. Albino et al., "Smart Cities: Definitions, Dimensions, Performance, and Initiatives", Journal of Urban Technology, Vol.22, No.1, pp.3-21, 2015
- [7] D. Gooch et al., "Reimagining the role of citizens in Smart City projects", Proceedings of the 2015 ACM International Symposium on Wearable Computers, ACM, New York, pp.1587-1594, 2015.
<https://doi.org/10.1145/2800835.2801622>
- [8] S. Letaifa, "How to Strategize Smart Cities: Revealing the SMART model", Journal of Business Research, Vol.68, No.7, pp.1414-1419, 2015.
<https://doi.org/10.1016/j.jbusres.2015.01.024>
- [9] T. Nam, P. Theresa, "Conceptualizing Smart City with Dimensions of Technology, People, and Institutions", Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Time, pp. 282-291, 2011
<https://doi.org/10.1145/2037556.2037602>
- [10] V. A. Schmidt, "Democracy and legitimacy in the European Union revisited: Input, output and 'throughput'", Political Studies, Vol.61, No.1, pp.2-22, 2013.
<https://doi.org/10.1111/j.1467-9248.2012.00962.x>
- [11] H. Chourabi, "Understanding Smart Cities: An Integrative Framework", 45th Hawaii International

- Conference on System Sciences, 2012.
<https://doi.org/10.1109/HICSS.2012.615>
- [12] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent dirichlet allocation", Journal of machine Learning research, Vol.3, pp. 993-1022, 2003.
- [13] T. Griffiths, and M. Steyvers, "Finding scientific topics," Proceedings of the National Academy of Sciences, Vol. 101, pp. 5228-5235, 2004.
<https://doi.org/10.1073/pnas.0307752101>
- [14] D. H. Jeong, and M. Song, "Time gap analysis by the topic model-based temporal technique", Journal of Informetrics, Vol.8, No.3, pp.776-790, 2014.
<https://doi.org/10.1016/j.joi.2014.07.005>
- [15] S. Lee, S. Lee, J. Lee, and H. Youm, "Extraction of Network Threat Signatures Using Latent Dirichlet Allocation", Journal of Internet Computing and Services(JICS), Vol.19, No.1 pp.1-10, 2018.
<http://dx.doi.org/10.7472/jksii.2018.19.1.1>
- [16] S. Cho, S. Shin, and D. Kang, "A Study on the Research Trends on Open Innovation using Topic Modeling ", Informatization Policy, Vol.25, No.3, pp. 52-74, 2018.
<https://doi.org/10.22693/NIAIP.2018.25.3.052>
- [17] W. S. Lee, and S. Y. Sohn, "Topic Model Analysis of Research Trend on Spatial Big Data", Journal of the Korean Institute of Industrial Engineers, Vol. 41, No. 1, pp. 64-73, February 2015.
<http://dx.doi.org/10.7232/JKIE.2015.41.1.064>
- [18] G. Kim, D. Yoon, J. Hwan, and D. Sun, "Discovering the emerging technologies through patent topic modeling and growth curve model", Journal of Korean Institute of Intelligent Systems, Vol. 27, No. 4, pp. 357-363, August 2017.
<http://dx.doi.org/10.5391/JKIS.2017.27.4.357>
- [19] J. Chung, Y. Park, and W. Kim, "Social Media Analysis Based on Keyword Related to Educational Policy Using Topic Modeling ", Journal of Internet Computing and Services, Vol.19, No.4, pp.53-63, 2018.
<http://dx.doi.org/10.7472/jksii.2018.19.4.53>
- [20] Q. Lv, L. Pang, and X. Li, "Learning Probabilistic Kernel from Latent Dirichlet Allocation", KSII Transactions on Internet & Information Systems, Vol. 10, No.6, pp.2527-2545, 2016.
<http://dx.doi.org/10.3837/tiis.2016.06.005>

◎ 저 자 소 개 ◎



박 건 철(Keon Chul Park)

2011년 연세대학교 정보대학원 정보시스템학 석사
 2015년 연세대학교 정보대학원 정보시스템학 박사
 2016년~현재 서울디지털재단 책임연구원
 관심분야 : 스마트시티, 공유경제, 디지털경제, 지능형정부
 E-mail : parkkc07@sdf.seoul.kr



이 치 형(Chi Hyung Lee)

2011년 연세대학교 정보대학원 정보시스템 박사
 2003년~2004년 다음 부사장
 2004년~2009년 KT 상무
 2016년~2019년 서울디지털재단 이사장
 2013년~현재 평택대학교 광고홍보학과 교수
 관심분야 : 온라인광고, 디지털미디어, 스마트시티
 E-mail : chilee@gmail.com, chilee@ptu.ac.kr