토지주택연구
Land and Housing Review
pISSN: 2093-8829, eISSN: 2234-1765
Land and Housing Review 2025, 16(4):55-67 | https://doi.org/10.5804/LHR.2025.16.4.55

스마트도시 조성을 위한 리빙랩 연구동향에 관한 소고

A Review of Research Trends in Living Labs for Smart City Development

이호현* · 박신원**
Ho-Hyun Lee* · Shin-won Park**

Abstract
This study examines how academic research on Smart City-Living Labs, an innovative approach to solving various urban problems arising from rapid urbanisation worldwide, has developed domestically and internationally by applying text mining methodologies. The analysis shows that smart city related research has been active overseas since 2002 and in Korea since 2016. This time lag appears to reflect the international diffusion of the Smart City-Living Lab concept from Europe (EU), followed by the gradual expansion of research in Korea as institutional foundations were established and relevant networks strengthened. In addition, an examination of key terms based on TF-IDF values reveals that, in the Korean literature, keywords such as 'citizen participation', 'user-centred', and 'service' have high weights, indicating a focus on service provision centred on citizens and users. By contrast, in overseas studies, terms such as 'data', 'innovation', and 'research' emerge with higher weights. This suggests that, whereas Korean research emphasises citizen participation and user-centred service delivery, international research seeks to more explicitly define and highlight the role of citizens through research grounded in technology and data. Overall, Smart City-Living Lab research, which has become a key paradigm for urban development in the twenty-first century, is establishing itself as an innovative and highly effective methodology for addressing the complex and multidimensional problems of cities. Accordingly, it is expected to continue to contribute to the development of integrated and holistic urban management strategies in which technology and social contexts are closely interconnected, serving as an important framework for future cities.

Keywords: Smart City-Living Lab, Text Mining, TF-IDF, Citizen Participation, Data

1. 서 론

1.1 연구의 배경 및 목적

전 세계적으로 산업화와 도시화에서 유발되는 도시문제 해결을 위한 혁신적인 접근 방식으로 스마트도시(Smart City)가 확산되었고, 산업구조의 변화에 따라 정보통신기술(ICT)을 활용한 도시의 관리는 사회·경제적으로 중요한 과제가 되고 있다. 이러한 스마트도시는 단순하게 공공부문에서 새로운 기술을 도입하여 도시문제를 해결하려는 차원을 넘어 시민참여와 협력을 통한 지속가능한 도시 발전을 목표로 한다는 점에서 리빙랩(Living Lab)과 밀접한 관련이 있으며, 도시문제 해결에 있어 기술 중심의 솔루션 및 서비스 도입에서 확장되어 민(民)·관(官)·산(産)·학(學)과 같은 다양한 주체 간 협력이 강조되면서 실증적인 접근 방식을 가능하게 한다.

*중앙대학교 대학원 도시계획 부동산학과 박사과정 수료(주저자: winteran93@nate.com)
**한국토지주택공사 토지주택연구원 스마트기후환경센터 수석연구원(교신저자: soka21@lh.or.kr)
Received: July 8, 2025 / Revised: September 10, 2025 / Accepted: November 18, 2025

최근 스마트도시는 방대한 양의 데이터와 정보가 온라인 기사, 도서, 기술문서, 누리소통망 서비스(Social Network Services, SNS), 웹페이지 등에서 텍스트 형식으로 저장 및 제공되고 있는데, 다양한 텍스트 정보에서 어느 특정한 주제와 관련된 부분을 추출하여 의미를 분석하고, 사회 현상이나 연구 경향 등 고품질의 정보를 효율적으로 추출하기 위한 방법으로 텍스트 마이닝 방법론이 활용되고 있다. 이러한 패러다임의 확산은 스마트도시 리빙랩(Smart City-Living Lab)과 관련된 다양한 연구를 통하여 도시문제 해결방안을 제시하기 위한 주요 연구 주제로 자리 잡고 있다.

기존 김은지(2019), 배기련 외(2022), 성지은·한규영(2017), 최민주(2022) 등의 선행연구에서는 시민 중심의 리빙랩 지표와 모델을 구축하는 것에 주안점을 두어 이론적 토대를 마련하고자 하였다. 또한, 김성묵·김영준(2020), 김기연(2024) 등의 선행연구에서는 리빙랩과 관련된 국내 키워드 분석이 이루어졌으며, 스마트도시와 리빙랩이 매우 중요한 핵심 키워드로 도출된 것을 확인하였다. 하지만 국내 연구에만 치우쳐져 있어 스마트도시 조성에 필요한 리빙랩과 관련된 해외 연구도 함께 수행되어 국내와의 키워드를 비교할 필요가 있다. 따라서 본 연구에서는 선행연구에서 공통적으로 도출되었던 핵심 키워드인 스마트도시 조성을 통해 도시문제를 해결하는 주요 혁신 도구인 시민참여를 기반으로 한 국내·외 스마트도시 리빙랩(Smart City Living Lab) 연구의 동향을 시계열적 관점에서 텍스트 마이닝 방법론을 활용하여 분석하고, 향후 스마트도시 조성을 위한 리빙랩의 연구 방향 설정을 위한 유의미한 학문적 기초자료를 제공하는 것에 그 의의가 있다.

1.2 연구의 범위와 방법

본 연구에서 활용한 분석자료는 학술자료 대표 검색엔진 사이트인 학술연구정보서비스(RISS; Research Information Sharing Services)와 전세계 학술 논문이 대량으로 수집되어 있는 Google 검색엔진 사이트의 자료를 활용하였다. 특히, 연구를 수행함에 있어 검색 대상에 초록을 포함시켰는데, 이는 연구의 핵심 주제와 주요 키워드를 가장 압축적으로 담고 있는 초록을 분석 대상으로 포함함으로써, 분석 결과의 정확성과 관련성을 높이기 위함이다. 즉, 초록에는 연구의 모든 과정이 모두 압축되어 제시되고 있다. 이에 초록을 검색 대상에 포함하는 것은 분석의 신뢰도와 타당성을 높이는 매우 중요한 연구 설계라고 판단되어 초록의 내용을 모두 포함하여 분석하고자 한다.

국내 자료는 3개(스마트도시 리빙랩, 스마트시티 리빙랩, Smart City Living Lab)의 단어를 검색하여 제목(주제어) 또는 초록으로 사용된 149편의 논문을 활용하고자 한다. 또한, 해외 자료는 2개(Smart City Living Lab, Urban City Living Lab)의 단어를 검색하여 제목(주제어) 또는 Abstract에 사용된 총 583편의 논문을 최종적으로 분석에 활용하고자 한다.

종합적으로 본 연구를 수행하기 위한 기초자료로써, 국내·외 학술 동향을 비교·분석하기 위해 총 732편의 논문을 활용하고자 한다.

2. 이론적 배경

2.1 스마트도시 리빙랩의 개념

스마트도시(Smart City)는 다양하게 정의되고 있으나, 일반적으로 교통·안전·환경 분야 등 도시에서 발생하는 다양한 문제를 첨단정보통신기술(ICT; Information & Communications Technology)을 활용하여 시민들의 불편함을 해소하고, 쾌적한 삶을 누릴 수 있게 하는 일명 '똑똑한 도시'를 의미한다.

대한민국은 2008년 세계 최초의 스마트도시 관련 법률인 「유비쿼터스도시의 건설 등에 관한 법률」을 제정하고, 신도시 중심의 U-City 개발을 추진하였다. 이후 2017년에는 기존 법률을 인프라 구축 중심의 스

마트도시 추진 체계를 기성 도시를 포함한 개발을 추진하기 위해 「스마트도시의 조성 및 산업진흥 등에 관한 법률」로 전면 개편하였다(정승현·김민주, 2021).

리빙랩(Living Lab)은 인공지능(AI)·사물인터넷(IoT) 등의 신(新)기술을 활용하여 사용자가 도시에서 발생되는 문제를 해결하기 위한 시민사회 영역의 혁신적인 공간을 의미하는데, '살아있는 실험실', '우리 마을 실험실', '일상생활 실험실' 등으로 해석되어 상용화되었다. 특히, 리빙랩에서는 기술의 개발 및 고도화, 사업화 과정에서 해당 기술을 이용하는 수요자들이 수동적 역할 개념에서 해당 기술 개발 과정에 밀접하게 참여함으로써, 기술 혁신을 이끌어나가는 핵심 기능을 수행하는 주체로 역할이 변화된다(곽병창 외, 2023).

스마트도시에 적용되고 있는 스마트도시 리빙랩(Smart City- Living Lab)은 도시에서 발생하는 다양한 문제를 해결하기 위해 민(民)·관(官)·산(産)·학(學) 등이 함께 참여하여 실제 생활 환경에 혁신적인 기술이나 서비스를 개발 및 테스트하는 실험적 플랫폼을 의미한다. 이는 지역 주민들이 직·간접적으로 참여하는 방식으로 장기적인 관점에서 살기 좋은 도시를 형성하여 도시를 보다 발전시키고자 하는 것으로써, 의제 발굴부터 심층(주민) 조사, 문제의 해결방안 도출, 문제의 해결(실험), 보완점 도출, 피드백 반영까지의 전 과정을 통해 발생하는 각종 문제점들을 지역에 거주하고 있는 주민들과 함께 해결하고자 하는 것이 가장 큰 특징이다(이호현·박신원, 2023).

2.2 텍스트 마이닝(Text Mining)

텍스트 마이닝은 방대한 양의 텍스트 데이터에서 유용한 정보를 수집·전처리·표현·분석 및 해석하여 데이터 기반 의사결정을 지원할 수 있는 정보를 효율적으로 찾아내는 기법이다.

텍스트 마이닝 기반 기술은 빅데이터 기술과 자연어 처리(NLP; Natural Language Processing)로 논문과 같은 비정형 텍스트에서 유의미한 정보 또는 패턴을 찾는 연구의 한 영역으로 일련의 텍스트 분석과 처리 과정을 통해 이루어진다.

텍스트 마이닝의 첫 단계는 데이터의 품질과 관련성을 고려하여 웹사이트, 도서, 논문 등 다양한 자료를 찾아 텍스트 데이터를 수집하고, 수집된 텍스트 데이터를 전처리 과정을 통해 토큰화, 불용어 제거, 어간 추출 및 표제어 추출, 대·소문자 변환 등의 작업을 수행한다. 전처리된 텍스트는 단어 빈도-역문서 빈도(TF-IDF; Term Frequency-Inverse Document Frequency)와 단어 임베딩(Word Embedding)을 통해 컴퓨터가 이해할 수 있는 형태로 변환한다.

대표적으로 네트워크 분석(Network Analysis)은 텍스트의 주요 의미를 가지는 단어들을 추출하여 연결 관계를 시각적 네트워크로 재구조화하여 보다 다양한 특성을 분류하는 데 유용하며, 주제 모델링(Topic Modeling)은 구조화되지 않은 대량의 문서 집합에서 내재된 토픽 구조를 발견하고, 통계적 추론 알고리즘으로 동일 또는 유사한 의미를 가진 텍스트 데이터를 클러스터링하는 방식으로 토픽을 추론하는 데 사용되고 있다. 이외에도 문서 분류와 연관 규칙 학습 등의 방법도 널리 사용되고 있다.

전반적으로 텍스트 마이닝 분석 결과는 인사이트를 도출하여 연구자의 의사결정을 지원하는 데 중요한 역할을 하기 때문에 데이터를 시각화하여 결과를 효과적으로 전달하는 것도 필수적이라고 할 수 있으며, 분석 결과를 통해 연구자는 다양한 분야에서 명확한 연구의 방향이나 전략을 세울 수 있게 도움을 주는 강점이 있다.

3. 분석모형 설정 및 기초 분석

3.1 LDA 토픽 모델링

잠재적 디리클레 할당(LDA; Latent Dirichlet Allocation)은 텍스트 데이터 내에 잠재된 주제(Topic)

를 확률적으로 추론하는 대표적인 토픽 모델링 기법이다. 문서를 여러 주제들의 혼합으로 구성되어 있다고 가정하고, 각 주제는 특정한 단어들의 확률적 분포로 정의된다. 이러한 LDA를 텍스트 마이닝 기법에 적용하여 다음 4가지의 특성을 반영한 국내·외 스마트도시 리빙랩(Smart City-Living Lab)의 분포 및 변화를 효율적으로 분석하고자 한다.

첫째, LDA는 비지도 학습 방식으로 동작하기 때문에 레이블링이 되지 않은 대규모 텍스트 데이터를 효과적으로 분석할 수 있다.

둘째, 문서 내 주제 분포와 주제 내 단어 분포를 확률적으로 표현하여 직관적인 해석과 주제의 시각화가 가능하므로, 주제 간 비교나 텍스트 내 주요 이슈 파악에 유리하다.

셋째, LDA는 단어와 문서 간의 관계를 명확하게 모델링하여 특정 텍스트가 어떤 주제들로 구성되어 있는지 정량적으로 나타낼 수 있어 다양한 분야에서 실무적으로 활용이 가능하다.

넷째, 대규모 텍스트 데이터 셋에서도 계산 효율성이 높아 빠른 분석이 가능하며, 분석의 확장성과 유연성이 뛰어나다.

이러한 LDA 모델을 적용하기 위해서는 전체 텍스트를 몇 개의 주제(k)로 분류할 것인지를 사전에 결정해야 한다. 즉, k값은 "해당 분야의 연구 동향이 총 몇 개의 핵심 흐름으로 구성되어 있는가?"라는 질문과 직결되며, 이 값의 선정 과정은 연구의 타당성을 입증하는 핵심 절차이다.

본 연구는 k값을 임의로 선정하지 않고, 통계적·정량적 지표를 통해 가장 객관적인 최적의 k값을 도출하고자 한다. 이를 위해 토픽 모델링의 품질을 평가하는 가장 보편적인 지표 중 하나인 토픽 일관성 점수(Coherence Score)를 사용하였다. Coherence Score는 특정 토픽을 구성하는 상위 단어들이 원본 텍스트에서 실제로 얼마나 자주 함께 등장하는지를 측정한다. 점수가 높을수록 해당 토픽이 우연의 일치가 아닌, 의미론적으로 명확하고 일관성 있는 주제(동향)로 구성되었음을 의미한다. 특히, LDA 모델의 주제 일관성을 평가하기 위해 R의 CalcProbCoherence 함수를 활용하여 토픽 응집도(Topic Coherence)를 측정하였다. 또한, KoNLP 패키지를 활용해 텍스트 데이터에서 명사를 추출하였으며, dplyr 패키지로 단어 빈도를 분석한 후 wordcloud 패키지를 이용해 핵심 키워드를 시각화하여 주요 키워드 분석을 도출하고자 한다. 이와 관련하여 토픽 응집도(Topic Coherence) 지표의 수식은 다음 식 (1)과 같다.

$$C_P(t) = \frac{2}{N(N-1)} \sum_{i=2}^{N} \sum_{j=1}^{i-1} \log \frac{DF(w_i, w_j) \cdot D}{DF(w_i) DF(w_j)} \quad (1)$$

* D: 전체 문서 수
* $DF(w_m, w_i)$: 두 단어가 함께 등장한 문서 수
* N: 토픽 내 단어 수

분석을 위해 k값을 2(최소 분류)부터 10(세분화)까지 순차적으로 변화시켜 각 모델의 Coherence Score를 측정하였다(Fig. 1).

그래프의 x축은 토픽 수(k)를, y축은 Coherence Score를 나타내며, 점수의 범위는 약 0.01에서 0.07 사이로 산출되었다. Coherence Score는 무차원 지표로써, 값이 높을수록 토픽 내 단어들의 의미적 결속력이 강하고 모델의 주제 해석 품질이 우수함을 의미한다. 이와 관련하여 토픽 수(k)를 2에서 10까지 순차적으로 변화시키며 Coherence Score를 산출한 결과, k=2에서는 약 0.015로 비교적 낮은 값을 보였으나 이후 k가 증가함에 따라 점수가 점진적으로 상승하였다. 특히, k=5부터 k=8에 이르기까지 Coherence Score는 전반적으로 상승하는 추세를 보이고 있었으며, k=8에서 약 0.066으로 최고점을 기록하였다. 이는 토픽의 수를 8개까지 세분화하는 과정이 '스마트도시 리빙랩' 연구들의 이질적인 특성들을 더 명확하게 분리하여 토픽의 품질을 향상시켰

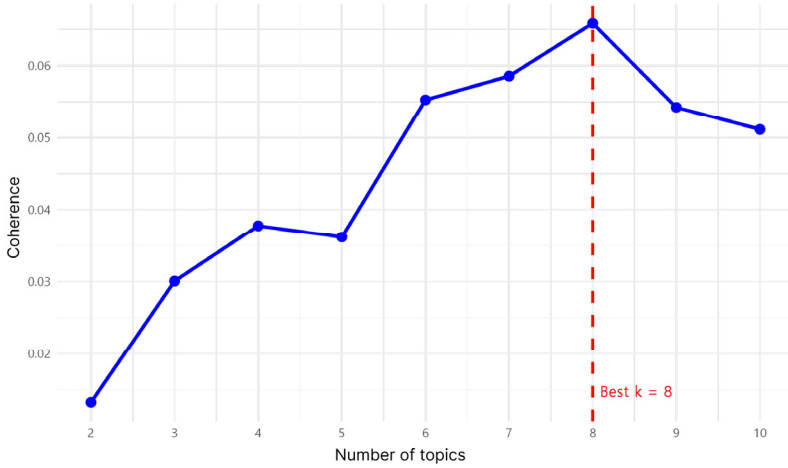

Fig. 1. Changes in the Coherence Score

음을 의미한다. 이 때 주목할 점은 k=8에서 점수가 가장 뚜렷한 정점(Peak)을 기록한 후, k=9부터는 오히려 점수가 하락한다는 것이다. 이것은 연구 동향을 8개보다 과도하게 세분화할 경우, 실제로는 하나의 일관된 동향이었던 주제가 불필요하게 쪼개지거나 의미 없는 단어들의 조합이 생성되어 토픽의 전반적인 품질과 일관성이 저하됨을 시사한다.

3.2 연도별 데이터 수집 및 전처리

본 연구에서는 국내·외 스마트도시(Smart City) 리빙랩 관련 연구를 수행한 논문의 제목(주제어), 국문초록의 비정형 데이터를 활용하여 스마트도시 리빙랩 연구의 주제 흐름, 핵심어 출현 빈도, 연구 방향성 등을 체계적으로 분석하고자 한다. 단, 복수형 단어(-s)를 단수형 단어로 모두 통일하였다. 이는 표제어 추출을 함으로써, 동일한 의미를 가진 단어의 다양한 형태를 하나로 통일하여 분석의 정확도를 향상시켜 데이터의 복잡성을 줄이기 위함이다. 이렇게 정리된 데이터는 R의 KoNLP 패키지를 활용해 텍스트 데이터에서 명사를 추출하였으며, dplyr 패키지로 단어 빈도를 분석한 후 wordcloud 패키지를 이용해 핵심 키워드를 시각화하여 주요 키워드 분석을 도출하고자 한다.

먼저 국내 스마트도시 리빙랩 관련 연도별 연구논문 수는 다음 Fig. 2와 같다. 2016년에 처음으로 연구논문이 발표되어 지속적으로 성장세를 보이다가 2022년 이후에는 일정한 연구 패턴을 보이고 있다. 특히, 2017년 한국리빙랩네트워크(KNoLL)의 출범과 U-City법(유비쿼터스도시의 건설 등에 관한 법률)에서 스마트도시법(스마트도시 조성 및 산업진흥 등에 관한 법률)으로 개정된 이후 2019년에서 2021년에는 뚜렷한 양적 성장세를 보였다. 이는 스마트도시의 제도적 기반 정비와 리빙랩 네트워크의 활성화가 결합되어 연구의 양적 확대에 기여한 결과라고 해석된다.

해외 Smart City Living Lab 관련 연도별 연구논문 수는 다음 Fig. 3과 같다. 2002년에 처음으로 연구가 시작되어 2000년대 초기에는 다소 미미한 수준의 논의가 이루어졌으나 2010년을 기점으로 급격한 양적 성장이 나타났다. 이후 2012년부터는 연간 논문 발간 수가 일정한 수준이 유지되면서 비교적 안정적인 연구 패턴을 보이고 있는데, 2010년에 들어서면서 스마트도시 리빙랩의 개념이 국제적으로 널리 확산되어 연구의 초기 탐색 단계를 지나 성

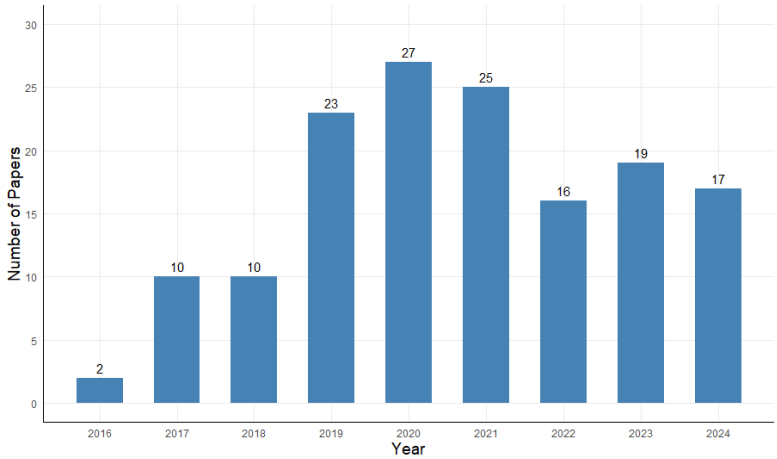

Fig. 2. Number of Papers Published by Year (Domestic)

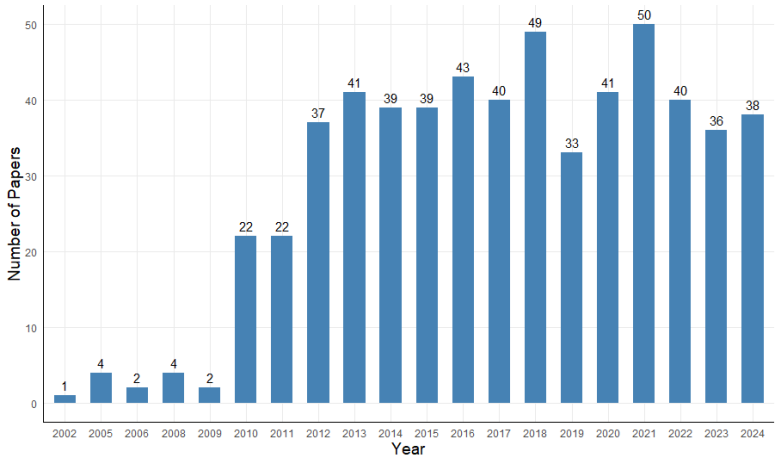

Fig. 3. Number of Papers Published by Year (Overseas)

숙기에 접어들었다고 이해할 수 있다. 특히, 이 시기부터 스마트도시 리빙랩이 단순히 이론적 논의에 머무르지 않고, 실증적·응용적 접근을 강조하는 실용 중심의 학문 분야로 자리 잡게 되면서 연구의 주제도 보다 다변화 및 구체화되는 경향을 보이고 있다. 또한, 스마트도시 리빙랩 개념이 전 세계적 확산과 더불어 연구의 초점이 점차 기존의 이론적 기반 확립이나 개념 정립 중심에서 벗어나 다양한 도시 환경과 사회적 맥락 속에서 실질적으로 발생하는 도시문제를 해결하기 위한 구체적·실무적인 연구와 적용 사례 중심으로 활발히 이루어진 결과로 이해할 수 있다.

4. 텍스트 마이닝 분석

4.1 워드 클라우드 분석

워드 클라우드(Word Cloud)는 특정 텍스트 데이터에서 단어들의 빈도수 또는 중요도를 시각적으로

표현한 것으로서, 자주 등장한 단어일수록 더 크게 나타난다. 사용자는 주로 문헌, 인터뷰, 설문조사 응답, 정책 문서 등을 분석할 때 워드 클라우드를 활용하고 있으며, 이러한 분석을 통해 주요 키워드를 직관적으로 파악하고 있다. 즉, 스마트도시 리빙랩 연구의 과거 경향과 방향성을 요약해 종합적인 인사이트를 도출할 수 있는 강점이 있다. 이에 본 연구에서는 국내·외 스마트도시 리빙랩(Smart City-Living Lab) 관련 자료에서 추출한 데이터를 기반으로 워드 클라우드를 생성하고자 한다.

먼저 국내에서 연구되었던 주요 단어와 의미를 Fig. 4에서 살펴보면 리빙랩, 스마트, 도시, 연구, 시민, 서비스 등이 가장 핵심적인 수단으로 자리잡고 있다. 이는 스마트도시 리빙랩이 단순한 실험 공간을 넘어서 시민참여 기반의 도시 혁신 거버넌스 모델로 부상하고 있음을 알 수 있다. 즉, 스마트도시 조성에 있어 기술 기반(TF, ICT) 못지않게 시민 중심,

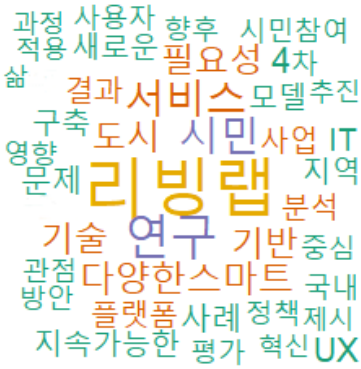

Fig. 4. Results of Domestic Word Cloud Analysis



Fig. 5. Results of Overseas Word Cloud Analysis

사용자 경험, 서비스 수요가 매우 중요하게 인식되고 있다는 것을 시사한다.

분석 대상은 실행 전략(운영·서비스), 제도적 모델(구축·평가), 정책 방향(지속가능성·도시재생) 등의 주제를 균형 있게 다루고 있다. 궁극적으로 이 모든 흐름은 스마트도시 정책이 기술 중심에서 사람 중심으로 전환되고 있음을 뚜렷하게 보여주고 있는 것이다. 종합적으로 스마트도시와 리빙랩을 단순한 구상 수준을 넘어 실행(Operationalization) 단계로 끌어올리고자 하는 시도가 강하게 드러나고 있는 것으로 확인된다.

해외에서는 Fig. 5에서 나타나듯이 Smart City-Living Lab 단어를 중심으로 Urban, Innovation, Data 등이 스마트도시 리빙랩 연구 분야에서 가장 핵심 수단이자 주요 키워드로 자리 잡고 있다. 특히, 도시 혁신과 지속가능성을 위한 리빙랩 개념의 확산과 더불어 도시문제 해결의 중심에 시민의 역할을 더욱 강조하고, 도시공간에서의 혁신을 촉진하는 방법론적 접근이 연구의 주요 흐름으로 정착되고 있다. 이와 관련하여 Data, Research, Development, Governance, Design, Citizens, Sustainability, Technology, Approach, Systems 등은 다음 단계인 중간 규모 키워드로 나타나고 있다. 이는 스마트도시 리빙랩 연구가 단순히 기술적인 측면을 강조하는 것을 넘어서 데이터 기반 연구를 활용해 도시의 지속가능한 발전을 핵심 가치로 삼고 있으며, 다양한 이해관계자들과의 협력을 바탕으로 도시의 복잡한 문제들을 실질적으로 해결하려는 통합적·실증적인 접근 방식을 중심으로 꾸준히 발전하고 있음을 의미한다.

또한, 데이터 기반의 연구 및 의사결정, 거버넌스 체계를 통한 정책적 지원, 사용자 중심의 디자인 접근법과 같은 실무 중심적인 요소들이 스마트도시 리빙랩 개념의 확산과 정착에 크게 기여하고 있음을 보여준다.

전반적으로 국내·외 스마트도시 리빙랩 연구는

모두 기술과 사람의 균형 잡힌 결합을 추구하고 있으며, 시민 중심적 접근과 도시 혁신 거버넌스 모델의 중요성을 뚜렷하게 강조하고 있는 공통점이 있다. 하지만 이와 관련하여 접근 방식의 경향성에서 일부 차이가 관찰된다. 국내 연구의 경우에는 세종특별자치시와 부산광역시 국가 시범도시 리빙랩 사례에서 나타나듯 개발된 기술·서비스를 실제 환경에 적용하고, 그 과정에서 얻은 결과를 바탕으로 정책적 효용성을 검증하여 제도적 안착을 지원하려는 실행적 측면에 집중하는 경향이 강하다. 반면, 해외에서는 MIT Senseable City Lab과 유럽 리빙랩 네트워크(ENoLL) 사례에서 나타나듯 전 세계 다양한 도시에서 생성되는 방대한 데이터를 심층적으로 연구하여 문제의 본질을 탐구하거나 시민참여 자체를 혁신의 핵심 동력으로 보아 방법론을 체계화하는 등 보다 통합적인 연구의 기반을 구축하는 데 무게를 싣고 있는 것으로 본다.

4.2 TF-IDF 분석

TF-IDF 분석은 비정형적인 텍스트 데이터 분석에서 특정한 단어가 문서 안에서 얼마나 중요한지를 평가하는 분석기법이다. 이는 단어 빈도(TF; Term Frequency)와 역문서 빈도(IDF; Inverse Document Frequency)의 곱으로 정의된다. 이 때 중요한 것은 앞서 분석한 Coherence Score는 LDA 모델을 통해 도출된 토픽의 해석 가능성과 품질을 평가하는 데 활용되는 반면, TF-IDF는 문서 내 단어의 중요도를 산정하는 데 활용된다는 점에서 그 차이가 있다.

TF는 특정 단어가 해당 문서 내에서 얼마나 자주 등장하는지를 나타내며, 일반적으로 단어가 문서 내에서 많이 등장할수록 높은 가치를 가진다. 그러나 모든 문서에서 흔히 등장하는 단어들은 TF 값만으로는 그 중요성을 충분히 평가할 수 없기 때문에 IDF로 이러한 문제를 보완하고자 해당 단어가 전체 문서의 집합에서 얼마나 드문지 희소성을 측정한다(김기연, 2024). 즉, 이 기법은 특정 단어의 중요도를 평가하는 데 사용되고 있으며, 단어 임베딩은 단어를 벡터 공간에 매핑하여 의미적 유사성을 반영하고 있다. 이와 관련하여 TF-IDF의 수식은 다음 식 (2)와 같다.

$$tf(t,d) = 0.5 + \frac{0.5 \times f(t,d)}{\max\{f(w,d) : d \ni w\}}$$

$$idf(t,D) = \log \frac{|D|}{\{D \ni d : d \ni t\}}$$

$$tfidf(t,d,D) = tf(t,d) \times idf(t,D) \quad (2)$$

* t: 특정 단어
* d: 특정 문서
* $|D|$: 문서집합 D의 크기
* $\{D \ni d : d \ni t\}$: t가 포함된 문서의 수

수식에서는 t(단어)가 d(문서)에서 얼마나 독특하고 중요한 단어인지를 수치로 표현한 값을 나타낸다. 이것은 한 문서에서 자주 등장하면서도 다른 문서에서는 보기 힘든 단어일수록 높은 가중치(TF-IDF)를 갖는다. 이러한 단어의 중요도를 효과적으로 측정하기 위해 먼저 명사 형태의 단어로 정제한 주요 키워드 빈도를 검토한 결과는 다음 Table 1과 같다.

국내에서는 스마트도시 기술과 디지털 서비스를 활용한 지역사회 발전 전략을 연구하는 과정에서 시민의 참여와 사용자 중심적 서비스 개발이 중요한 축으로 자리 잡고 있음을 발견할 수 있다. '시민참여', '사용자 중심', '서비스' 등과 같이 시민과 사용자를 중심으로 한 서비스 제공을 기반으로 한 키워드들이 높은 TF-IDF 값으로 나타나는 점은 스마트도시 리빙랩 연구가 기존의 기술 위주 접근에서 더 나아가 발전된 기술을 활용해 주민 참여와 사용자 경험을 핵심 가치로 설정하고 있음을 명확히 보여준다. 이는 기술과 도시 혁신의 방향성을 설정할 때 지

Table 1. Domestic and Overseas Occurrence Frequency of Key Word

Division	Frequency	Keywords
Domestic	Top 30	Smart City (531) Living Lab (359) Research (191) Citizen (144) Service (131) City (110) Various (97) Smart (95) Foundation (86) Necessity (80) Technology (77) Business (74) Platform (71) Results (67) Analysis (67) Problem (60) Case (54) Deployment (48) Model, New (47) Sustainable (44) Region (44) Impact (39) Policy (37) Center (34) Future (33) Perspectives (32) Methods (32) Evaluation (31) Citizen Participation (29)
Overseas		City (2573) Living (1651) Smart (1488) Lab (1438) Urban (1101) Innovation (624) Research (492) Technology (429) User (423) Data (366) Development (365) Paper (330) Approach (284) Design (276) Citizens (271) Based (251) Services (238) Project (230) Study (230) Social (228) Public (224) Process (215) Energy (206) Sustainable (204) Solutions (190) Challenges (186) Local (185) Project (185) Concept (184) Governance (161)

역 주민과의 협력을 통해 주민이 실제로 필요로 하는 실효성 있는 서비스 개발이 필수적이라는 점을 강력하게 시사하고 있다. 더하여 '사례(Case)', '개발(Development)', '참여(Participation)' 등의 융합 관련 단어들이 높게 등장하는 것은 지역사회에서 발생하는 다양한 도시문제를 해결하는 데 있어 단순히 이론적 차원에 머물지 않고, 실제 도시 현장에 적용이 가능한 구체적인 사례 연구와 이를 통해 실효성 높은 해결 방안을 모색하는 노력이 활발히 진행되고 있음을 의미한다. 즉, 스마트도시 기술과 서비스를 기반으로 지역사회가 겪고 있는 사회적·경제적·환경적 문제들을 효율적으로 해결하기 위해 지역 내 다양한 이해관계자들이 적극적으로 참여하고 협력하는 새로운 형태의 도시 혁신 모델이 대두되고 있음을 의미한다.

해외에서는 'Smart City(Urban)-Living Lab', 'Data', 'innovation' 'research' 등의 단어는 상대적으로 높은 TF-IDF 값을 보여 스마트도시 조성을 위해 관련 데이터를 활용한 연구를 통해 혁신을 하고자 하는 것으로 해석된다. 특히, 혁신을 의미하는 'Innovation' 키워드는 향후 스마트도시 리빙랩 연구의 근본적인 개념과 매우 밀접하게 연결된 의미 있는 핵심 단어로 본다. 이러한 혁신 개념은 단순한 기술 도입이나 인프라 구축 수준을 넘어 도시 내에서 지역사회가 당면한 복잡한 문제들을 해결하기 위한 포괄적·실현적인 접근 방식에서 출현하고 있다. 즉, 리빙랩의 개념이 도시 혁신과 지역사회 문제 해결이라는 구체적인 목표를 달성하기 위한 효과적인 방법론으로 자리 잡고 있으며, 연구에서는 이러한 혁신적 접근 방식을 다양한 지역사회 맥락에서 적용하고 있음을 알 수 있다.

더 나아가 주요 핵심 키워드가 시계열적 관점에서 얼마만큼 활용되었는지 살펴보고자 한다. Fig. 6에서 보면, 국내·외 스마트도시 리빙랩 관련 연구는 공통적으로 2010년대를 전·후로 급격한 관심과 확산

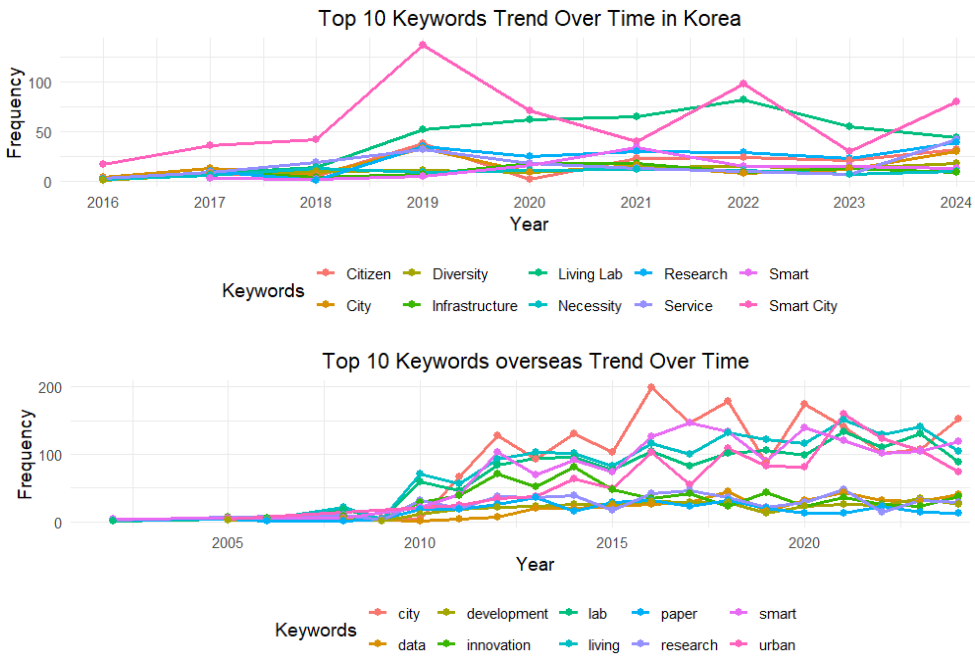

Fig. 6. Primary Keywords Trends Over Time

이 이루어졌으며, 2020년을 기점으로 '스마트도시', '리빙랩'과 같은 키워드가 주된 중심축으로 자리 잡는 양상이 뚜렷하다. 이러한 시계열적 변화는 국내·외 연구자들이 세계적인 스마트도시 담론을 적극적으로 수용하면서 이를 자국의 사회적·정책적 환경에 맞게 재해석하려는 경향과 밀접하게 연관되어 있음을 시사한다.

이와 관련하여 국내 스마트 기반 도시의 필요성이 높아짐에 따라 리빙랩과 같은 세계적인 방법론을 적극 수용했다는 큰 흐름을 읽을 수 있다. 이와 동시에 해외의 기술과 데이터 중심을 바탕으로 시민의 역할을 더욱 명시적으로 강조하는 것을 고려하여 '한국형 스마트도시' 모델을 구축하고자 지속적으로 노력하고 있음을 엿볼 수 있다. 즉, 글로벌 트렌드의 수용과 국내에서만의 독자적인 발전을 통한 로컬라이제이션(Localization) 과정이 키워드 빈도 변화에 그대로 나타나 있다고 본다. 예를 들어, 글로벌 사례를 수용하되 국내 지역별 실정에 맞게 변형한 스마트도시 계획이 대표적이라고 할 수 있다.

해외에서는 2000년대 초반에는 주요 키워드의 빈도가 전반적으로 낮은 수준을 유지하였으나, 2010년을 기점으로 뚜렷한 증가세가 나타났으며, 2012년부터는 다수의 키워드에서 뚜렷한 절정이 확인된다. 이 가운데 스마트도시 리빙랩 관련 계열이 가장 높은 빈도를 보여 주도적인 위치를 꾸준히 차지하는 추세를 보였다. 또한, 연구를 활용해 도시문제 해결을 위한 실증적 근거를 마련하고자 하는 빈도도 상당히 높게 나타나는 것을 확인할 수 있다. 특히, 도출된 핵심 키워드 중 "paper"와 "research" 등의 단어는 분석 대상 문헌의 연구 성격(예: 이론 중심 연구에서 실증적 연구로의 전환), 연구의 축적 정도 등을 반영하는 중심 개념으로 해석된다.

전반적으로 21세기 도시 발전의 핵심 패러다임으로 자리 잡은 스마트도시는 더 이상 낯선 개념이 아니다. 첨단 기술을 활용하여 도시문제를 해결하여 시민의 삶의 질을 향상하려는 이 거대한 흐름은 전

세계적으로 확산되었으며, 국내·외에서 저마다 다른 방식으로 새로운 도시 모델을 구현하고자 노력하고 있다.

따라서 스마트도시 리빙랩 연구는 단순히 스마트 기술이나 ICT 기반 기술 적용에만 머무르지 않고, 지역 공동체를 중심으로 시민들이 능동적으로 참여하여 도시가 직면한 다양한 사회적·경제적·환경적 현안을 보다 체계적인 경험(사례)을 기반으로 해결하고자 하는 융합적·통합적인 접근 방식을 지향하고 있다고 본다. 향후 이러한 접근 방식은 도시 정책과 리빙랩의 실행 전략을 수립하는 데 있어 더욱 중요한 프레임워크로 기능할 것으로 기대되며, 주민 친화적인 스마트도시 구현에 기여할 것으로 전망된다. 이러한 과정을 통해 스마트도시 리빙랩 연구는 도시 문제 해결을 위한 새로운 학제적 연구 프레임워크를 제시하고, 실증적 자료의 축적을 통해 연구의 신뢰성과 타당성을 향상할 수 있으며, 더 나아가 스마트도시 연구의 이론적 지평을 확장하는 데 기여할 수 있을 것이라고 사료된다. 즉, 이것은 향후 스마트도시 리빙랩이 4대 요소(시민참여, 기술, 정책, 사회적 측면)를 모두 아우르는 융합적인 연구 방식을 더욱 강조하고자 하는 것이다.

5. 결 론

본 연구는 전 세계적으로 도시화가 급속하게 진행됨에 따라 도시에서 발생하는 각종 문제를 해결하기 위한 혁신적인 접근 방식인 스마트도시 리빙랩(Smart City-Living Lab) 연구에 대해서 텍스트 마이닝 방법론을 활용하여 국내·외에서 학문적으로 어떻게 발전해왔는지 고찰하였다. 분석을 위해 국내 기초자료는 학술자료 대표 검색엔진 사이트인 RISS(학술연구정보서비스)의 자료를 활용하였으며, 해외 기초자료는 전 세계 학술 논문이 대량으로 수집되어 있는 Google 검색엔진 사이트의 자료를 활용하였다. 이를 통해 스마트도시 리빙랩 연구의 주제 흐름, 핵심어 출현 빈도, 연구 방향성 등을 체계적으로 도출하고자 하였다. 이와 관련하여 본 논문에서 분석 결과는 크게 다음 4가지로 구분할 수 있다.

첫째, 먼저 LDA 모델의 주제 일관성을 평가하기 위해 토픽 응집도(Topic Coherence) 지표를 활용해 k값을 2(최소 분류)부터 10(세분화)까지 순차적으로 변화시켜 각 모델의 Coherence Score를 측정하였다. 점수의 범위는 약 0.01에서 0.07 사이로 산출되었으며, 토픽의 수를 8개까지 세분화하는 과정이 '스마트도시 리빙랩' 연구의 이질적인 특성들을 더 명확하게 분리하여 토픽의 품질을 향상시켰음을 확인하였다.

둘째, 스마트도시 리빙랩이 학문적으로 발전하게 된 시점은 해외에서는 2002년부터 연구가 시작되었으나 국내에서는 2016년부터 연구가 시작되었다. 이는 해외 그 중에서도 유럽(EU)에서 스마트도시 리빙랩의 개념이 국제적으로 널리 확산되었고, 연구의 초기 탐색 단계를 지나 성숙기에 접어들면서 국내에서도 이러한 흐름에 발맞춰 2017년 한국리빙랩네트워크(KNoLL)의 출범과 U-City법(유비쿼터스도시의 건설 등에 관한 법률)에서 스마트도시법(스마트도시 조성 및 산업진흥 등에 관한 법률)으로 개정하는 등 스마트도시의 제도적 기반 정비와 리빙랩 네트워크의 활성화가 결합되어 양적 확대에 기여한 결과라고 해석된다. 이러한 흐름에 따라 현재 스마트도시 리빙랩은 단순히 이론적 논의에 머무르지 않고, 실증적·응용적 접근을 강조하는 실용 중심의 학문 분야로 확고히 자리 잡게 되면서 연구 주제도 보다 다변화 및 구체화되는 경향을 나타내게 되었다.

셋째, 주요 단어 키워드를 살펴봤을 때, 국내에서는 스마트도시 기술과 디지털 서비스를 활용한 지역사회 발전 전략을 연구하는 과정에서 시민의 참여와 사용자 중심적 서비스 개발이 중요한 축으로 자리잡고 있음을 발견할 수 있었다. 반면, 해외에서는 리

빙랩의 개념이 도시 혁신과 지역사회 문제 해결이라는 구체적인 목표를 달성하기 위한 효과적인 방법론으로 자리 잡고 있으며, 연구에서는 이러한 혁신적 접근 방식을 다양한 지역사회 맥락에서 적용하고 있음을 알 수 있었다.

넷째, 더 나아가 핵심 키워드에 대한 시계열적 관점에서 살펴본 결과, 국내 스마트 기반 도시의 필요성이 높아짐에 따라 리빙랩과 같은 세계적인 방법론을 적극 수용했다는 큰 흐름을 읽을 수 있었다. 이와 동시에 해외에서는 기술과 데이터를 바탕으로 시민의 역할을 더욱 명시적으로 연구를 통해 강조하고자 하였다.

전반적으로 21세기 도시 발전의 핵심 패러다임으로 자리 잡은 스마트도시는 더 이상 낯선 개념이 아니다. 첨단 기술을 활용하여 도시에서 발생되는 다양한 문제를 해결하여 시민의 삶의 질을 향상하려는 이 거대한 흐름은 전 세계적으로 확산되었으며, 국내·외에서 저마다 다른 방식으로 새로운 도시 모델을 구현하고자 노력하고 있다. 향후 이러한 접근 방식은 도시 정책과 리빙랩의 실행 전략을 수립하는 데 있어 더욱 중요한 프레임워크로 기능할 것으로 기대되며, 주민 친화적인 스마트도시 구현에 기여할 것으로 전망된다.

그럼에도 불구하고 본 연구에서는 분석에 활용한 자료가 학술 데이터베이스에 한정되어 있어 비학술적 현장 사례나 정책적 동향 등 실무적 측면을 충분히 반영하지 못한 한계가 있다. 이에 향후 연구에서는 정성적 연구 방법과 결합하여 데이터를 추가적으로 보완한다면 보다 발전된 연구가 진행될 수 있을 것이라고 사료된다.

참고문헌

1. 곽병창·지원길·이성재·김길태(2023), “스마트하우징 주거서비스 기술에 대한 이용자 인식 개선을 위한 리빙랩 활용성 분석 연구”, 「LHI Journal」, 14(3): 125~135.
2. 김기연(2024), “리빙랩에 관한 연구 주제 및 동향 분석: LDA 토픽모델링과 의미 연결망 분석을 통한 탐색적 접근”, 「한국인터넷정보학회논문지」, 25(5): 171~182.
3. 김성묵·김영준(2020), “텍스트 마이닝을 이용한 리빙랩 연구동향 분석”, 「우리말글」, 18(8): 37~48.
4. 김은지(2019), “리빙랩 평가지표 개발에 관한 연구: 도시재생 리빙랩을 중심으로”, 박사학위논문, 이화여자대학교 대학원.
5. 배기련·양정연·송경신·김재민·이준협·정혜주(2022), “보건복지 분야를 중심으로 본 국내 리빙랩 연구의 현황: 주제범위 문헌고찰”, 「보건사회연구」, 42(4): 248~265.
6. 성지은·한규영(2017), “중간지원조직의 리빙랩 현황과 플랫폼으로서의 발전 가능성 탐색”, 「기술혁신학회지」, 20(4): 915~938.
7. 이호현·박신원(2023), “효율적인 스마트시티 구현을 위한 리빙랩 활성화 방안에 관한 연구”, 「부동산학보」, 92: 115~132.
8. 정승현·김민주(2021), “EU 라이트하우스 프로젝트 분석을 통한 스마트시티 실증과 확산을 위한 시사점 도출”, 「LHI Journal」, 12(2): 1~12.
9. 최민주(2022), “시민 중심의 스마트도시 리빙랩 모델 구축”, 박사학위논문, 한밭대학교.

요 약

본 연구는 전 세계적으로 도시화가 급속하게 진행됨에 따라 도시에서 발생되는 각종 문제를 해결하기 위한 혁신적인 접근 방식인 스마트도시 리빙랩(Smart City-Living Lab) 연구에 대해서 텍스트 마이닝 방법론을 활용하여 국내·외에서 학문적으로 어떻게 발전해왔는지 고찰하였다. 분석 결과, 해외에서는 2002년부터, 국내는 2016년부터 스마트도시 관련 연구가 활성화되었으며, 이는 유럽(EU)에서 스마트도시 리빙랩의 개념이 국제적으로 널리 확산된 후 국내에서도 제도적 기반 정비와 네트워크 활성화로 연구가 점차 확대된 것으로 분석된다. 더하여 주요 키워드를 살펴봤을 때, TF-IDF 값이 국내에서는 '시민참여', '사용자 중심', '서비스' 등과 같이 시민과 사용자를 중심으로 한 서비스 제공을 기반으로 한 키워드들이 높게 형성되었다. 반면, 해외에서는 'Data', 'innovation' 'research' 등의 단어가 높게 형성되는 것을 확인하였다. 이는 국내 연구가 시민참여와 사용자 중심 서비스 제공에 초점을 맞추고 있는 반면, 해외 연구는 기술과 데이터를 바탕으로 시민의 역할을 더욱 명시적으로 연구를 통해 강조하고자 하는 것을 의미한다. 전반적으로 21세기 도시 발전의 핵심 패러다임으로 자리 잡은 이러한 스마트도시 리빙랩 연구는 도시의 복합적·다차원적인 문제를 효율적으로 해결하기 위한 실효성이 높은 혁신적인 방법론으로 자리 잡아가고 있다. 따라서 미래에는 기술과 사회적 맥락이 긴밀하게 연결된 융합적·통합적인 도시의 관리 전략과 중요한 프레임워크로 지속해서 발전에 기여할 것으로 전망된다.

주제어: 스마트도시 리빙랩(Smart City-Living Lab), 텍스트 마이닝, TF-IDF, 시민참여, 데이터