

## 메가시티에 대한 언론과 수용자 관심 차이 분석: 텍스트 마이닝 기법을 중심으로

이 우 철<sup>1</sup> · 이 장 석<sup>2\*</sup>

<sup>1</sup>중앙대학교 광고홍보학과 박사

<sup>2\*</sup>가천대학교 미디어커뮤니케이션학과 조교수

### Analysis of Differences in Media and Audience Interest in “Megacity”: Focusing on Text Mining Techniques

Woo-Chul Lee<sup>1</sup> · Jang-Suk Lee<sup>2\*</sup>

<sup>1</sup>Ph.D., Department of Advertising & Public Relations, Chung-Ang University, Seoul 06974, Korea

<sup>2\*</sup>Assistant Professor, Department of Media Communication, Gachon University, Seongnam 13120, Korea

#### [요 약]

본 연구는 메가시티에 관한 언론 매체의 보도 내용과 소셜 미디어상의 소비자 주도 토론 간의 차이점을 조명하기 위해 진행되었다. 토픽모델링 기법을 적용하여 언론기사와 소셜 네트워크 서비스(SNS) 콘텐츠를 분석한 결과, 정치적 이슈, 지역 균형발전, 부동산 및 경제적 가치 등이 주된 토픽으로 도출되었다. 언론기사는 국가적 차원의 프로젝트와 정책적 측면에 중점을 두는 경향이 있는 반면, SNS 콘텐츠는 메가시티가 개인의 경제적 이익과 밀접한 관계에 있음을 반영하였다. 이러한 발견은 메가시티 프로젝트가 다양한 이해관계자의 시각을 반영하는 다면적 영향을 고려해야 하며, 메가시티 정책의 설계와 소통 과정에서 국민의 경제적 관심사와 인식을 충분히 반영할 필요가 있음을 시사한다.

#### [Abstract]

This study verifies the differences in content between news articles and consumer-generated content on megacities. A comparison of main topics in news articles and on social networking service(SNS) using topic modelling analysis revealed political issues, regional balanced development, real estate, and economic value as key topics. While news articles mainly emphasized national-level projects and policy perspectives, SNS content highlighted aspects that were more closely tied to individuals' economic benefits. These findings suggest the need to consider the multidimensional impacts of megacity projects and the perspectives of various stakeholders. The design and communication of policies related to megacities should adequately reflect the economic interests and perceptions of the public.

**색인어** : 메가시티, 토픽모델링, 스킵그램, 텍스트 마이닝, 빅데이터

**Keyword** : Megacities, Topic Modeling, Skip-Gram, Text Mining, Big Data

<http://dx.doi.org/10.9728/dcs.2024.25.4.939>



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

Received 24 March 2024; Revised 12 April 2024

Accepted 17 April 2024

\*Corresponding Author, Jang-Suk Lee

Tel: +82-31-750-5756

E-mail: [jslee81016@gachon.ac.kr](mailto:jslee81016@gachon.ac.kr)

## I. 서론

도시 과밀화와 주택정책 문제, 대규모 행정도시 건설 등의 문제점이 가속화되면서 메가시티(megacity)가 이러한 문제의 대안으로 부상하고 있다. 에드워드 글레이저(Edward Glaeser) 하버드대 경제학 교수는 메가시티 리전(megacity region)의 출현을 ‘21세기의 필연적인 트렌드’라고 설교했다 [1]. 오늘날 세계 주요 국가에서는 각 국가의 경쟁력을 좌우하는 요소로 선정하고 메가시티를 육성하기 위해 많은 노력을 기울이고 있다[2].

국내에서도 메가시티 조성은 국가 균형발전을 위한 중요한 의제로서 ‘부울경(부산, 울산, 경남)’ 메가시티를 비롯한 여러 시·도의 광역 통합 시도가 활발하게 진행되어왔다. 이는 지방자치법 전면 개정과 국가균형발전 특별법 시행령 개정을 통한 법적 지원에 힘입은 결과다. 이러한 움직임은 지역 경쟁력을 제고하는 데 목적을 두고 있으며, 초광역권으로서의 메가시티 개념을 명확히 하고 해당 지역의 발전 및 협력 사업을 촉진하기 위해 마련되었다[3].

최근 여당인 국민의힘이 추진하는 경기도 김포시의 서울시 편입 계획을 발표하면서 메가시티에 대한 관심을 재점화했다. 이는 김포시민들의 삶의 질 향상과 서울의 경쟁력 강화라는 두 가지 목적을 달성하기 위한 전략으로 행정구역과 생활권을 일치시키려는 배경에서 시작되었다[4].

하지만 김포시의 서울 편입 논의가 정치적으로 활용되고 있음을 비판하는 목소리도 적지 않다. 일부 김포시민들은 이러한 움직임을 ‘총선용 정략’으로 보고 있으며, 실현 가능성에 대한 의문을 제기하고 있다[5]. 2024년 4월에 치러질 총선을 앞두고 김포시의 서울 편입이 선거와 관련된 이슈로만 활용되는 것이 아닌지에 대한 의문을 품는 것이다.

본 연구는 메가시티를 바라보는 언론과 수용자 관심의 차이에 주목하였다. 프레임(frame)이란 언론이 원하는 방식으로 특정 사건이나 현상에 독특한 해석이나 성격을 부여하는 과정을 의미하며[6], 특히 공공분야에서 언론의 프레임 구축은 전향이나 왜곡이 자주 관찰되는 영역이다[7]. 더불어 공공정책에 대한 언론의 프레임과 사회적 인식 간 차이를 밝힌 여러 연구 결과[8],[9]를 고려할 때, 메가시티 이슈에 대한 언론기사와 SNS상에서의 온도 차를 분석하는 것이 어느 때보다 시기적절하다 하겠다.

지금까지 메가시티와 관련된 연구는 주로 메가시티에서 나타날 수 있는 여러 위협과 그에 대한 대응 방안을 모색하거나 초광역권 구상의 공간으로서 거버넌스에 대한 함의를 탐색해왔다[2],[10]. 방법론적으로는 빅데이터가 활용되며 토픽모델링이나 의미연결망 분석을 통해 스마트시티에 대한 연구 동향을 분석하거나 언론의 반응을 조망하는 것도 가능해졌다 [11]-[13]. 하지만 메가시티와 관련된 다양한 이슈를 비롯해 메가시티가 직면한 과제와 기회에 대한 깊이 있는 분석을 제공하는 연구는 여전히 부족한 실정이다.

본 연구는 현대 사회에서 메가시티가 차지하는 비중과 영

향력이 점점 커지는 상황에서 메가시티의 다양한 측면을 빅데이터 분석을 통해 깊이 있게 탐구하는 것을 목적으로 한다. 이를 위해 메가시티 이슈에 대한 언론기사와 SNS 콘텐츠를 별도로 분석하여 도시의 사회적 복잡성을 보다 정밀하게 파악하고자 하였다. 본 연구는 메가시티 현상을 종합적으로 이해하고 도시 계획, 운영 및 정책 수립에 구체적인 지침을 제공함으로써 이론적, 실천적 시사점을 제공하고자 하였다.

## II. 이론적 배경

### 2-1 메가시티에 대한 논의

메가시티는 매우 큰 도시(very large city)를 의미하는 용어로 전 세계적으로 47개 정도로 보고되었다. 메가시티의 결정 기준은 도시의 인구가 1,000만 명 이상이 거주하는 도시를 메가시티로 지칭한다[3].

최근 국내에서는 김포시의 서울 편입 발표 이후 메가시티와 관련된 이슈가 전국적으로 재점화하고 있는 가운데, 주로 지방의 경쟁력 강화와 국가 경쟁력과의 밀접한 연관성에 초점을 맞추고 있다. 충청권에서는 대전, 세종, 충남, 충북 등 4개 시·도가 단일 행정권역으로 통합하는 메가시티 추진을 선언하였으며[14], 대구·경북(TK)에서도 지역의 경제적 발전 가능성을 높이고, 인구나 자원의 수도권 집중을 완화하는 것을 목표로 메가시티 논의가 활발하게 진행되고 있다[15]. 이러한 움직임은 지방의 자립적인 경제 발전 가능성을 높이고 수도권에 집중된 인구나 자원을 분산시켜 균형적인 국토 발전을 도모하기 위한 것이라고 할 수 있다.

하지만 대한민국은 극심한 인구소멸위험 증가와 지방 인구의 축소, 그리고 공동화 현상 등 여러 문제에 직면하고 있으며, 이로 인해 지역 산업 인프라가 인구 밀집도가 높은 대도시로 이동하는 경향이 갈수록 심화되고 있다[2]. 이러한 변화는 대도시로의 인구 집중 현상을 가속화시키며, 결과적으로 지역 간 경제적 및 사회적 불균형을 더욱 크게 만들 수 있는 위험성을 수반한다. 지표상으로도 수도권과 비수도권 간 고용 경쟁력과 지역 총생산액 비중의 격차는 갈수록 커지고 있으며, 수도권 평일 교통혼잡비용은 17.8조로 전국의 53%에 육박하고 있다[3]. 특히, 코로나19 상황에서 나타나듯 ICT 기반 네트워크화된 스마트 도시로서 메가시티는 밀집된 인구나 복잡한 시스템은 현존, 잠재적, 초국가적, 비군사적 관점에서 새로운 위협을 초래할 수 있다[2]. 이와 같은 양극화를 효과적으로 관리하고 극복하기 위해 국가적 차원에서의 전략적이고 계획적인 접근이 필요한 시점이다.

### 2-2 TF-IDF 연관어 분석과 토픽모델링

본 연구에서는 추출된 키워드들을 빈도수에 기반하여 순위화하여, 가장 빈번하게 언급된 단어부터 분류하였다. 여기서

‘TF(Term Frequency)’는 한 문서 내 특정 단어가 출현하는 빈도를 나타내며, 해당 단어의 출현 빈도가 높을수록 그 단어는 그 문서에서 상대적으로 중요한 것으로 간주된다. 그러나 이러한 단어가 다른 문서에서도 일반적으로 자주 등장한다면 그 중요성은 상대적으로 감소할 수 있다. 이를 보정하기 위해 전체 문서 수를 그 단어가 출현하는 문서 수로 나눈 값인 ‘IDF (Inverse Document Frequency)’가 이용된다. 따라서 높은 IDF 값을 보이는 단어는 다른 문서에서는 드물게 등장하므로 해당 문서 내에서 더욱 중요한 의미를 지닌다고 할 수 있다. ‘TF-IDF’는 이 두 값을 곱하여 계산하는데 이는 특정 단어가 한 문서에서는 자주 등장하지만 전체 문서에서는 드물게 나타날 경우 높은 가중치를 부여하는 방식으로 작동한다. 이러한 메커니즘을 통해 문서 특이적인 중요 단어를 효과적으로 식별할 수 있다[16].

이에 반해 토픽모델링은 텍스트 기반의 문서 집합으로부터 토픽을 추출하는 기법으로 가장 많이 활용되는 것은 LDA (Latent Dirichlet Allocation) 비지도 학습모형이다[17], [18]. 토픽모델링 과정에서는 토픽의 수를 결정하는 것이 중요한 과제 중 하나이다[19]. 이를 위해 본 연구에서는 각 토픽 내에서 의미론적으로 유사한 단어들의 밀집도를 나타내는 지표인 Coherence Model을 사용하여 토픽의 의미론적 일관성을 평가하고 이를 기준으로 적절한 토픽의 수를 선정하였다. Coherence Model의 값이 높을수록 토픽 내 단어들 사이의 의미론적 연관성이 강하며, 이는 주제가 명확하고 일관된 정보를 포함하고 있음을 의미한다[16]. 이러한 접근 방식은 토픽모델링을 통해 추출된 주제들이 실제로 의미 있는 정보를 담고 있으며, 분석 대상 문서 집합을 효과적으로 대표한다는 것을 보장하는 데 도움을 준다.

### 2-3 Word2Vec

Word2Vec은 토픽별로 단어 간 유사도를 측정하는 기법이다. 이 연구에서는 대중매체와 소비자 버즈에서 언급된 상위 100개의 상업적 키워드를 TF-IDF 방법을 사용하여 추출하고, 이들 단어의 유사도를 평가했다. 단어 간의 유사도를 정밀하게 측정하기 위해 Word2vec라는 분산 표상 모델을 사용하여 각 단어를 벡터 공간에 임베딩(embedding)했다. 이 과정에서 각 단어는 문맥상 의미가 최대한 유지되도록 학습되기 때문에 결과적으로 단어들의 밀집도를 통해 대중매체와 소비자 버즈 사이의 의미론적 차이를 더욱 명확히 드러낼 수 있다. 이 방법론은 언론과 소비자 간의 의사소통 패턴과 관심사의 차이를 보다 구체적으로 이해하고 관련 분야에서 의미 있는 인사이트를 발견하는 기회를 얻을 수 있다.

임베딩은 자연어를 숫자 배열인 벡터로 변환하는 결과물이자 해당 변환 과정 전체를 아우르는 용어로, 단어나 문장을 각각 벡터로 전환하여 벡터 공간에 끼워 넣는 것(embed)을 의미한다[20]. 즉, 인간의 언어로 구성된 단어나 문장을 기계가 파악할 수 있는 벡터 형태로 전환하는 과정으로 자연어 처리

및 기계 학습 분야에서 활용되는 주요 기술이라고 할 수 있다. 이러한 임베딩 기술은 단어나 문장의 의미를 수치화하여 기계 학습 알고리즘이 인지하고 학습할 수 있도록 지원한다. 임베딩에는 Word2Vec, GloVe, FastText 등과 같은 알고리즘이 각각 고유의 방식으로 벡터를 생성한다. 본 연구에서도 유사한 의미를 가진 단어들을 동일한 공간에 배치함으로써, 두 데이터 집단의 키워드 집중도 차이를 비교 분석하였다.

Word2Vec의 Skip-gram 알고리즘은 단어 간 유사도를 파악하고, 의미적 연관성을 벡터 공간에 임베딩함으로써 식별하는 방법을 제공한다. 이 기법은 단어들의 의미를 벡터로 변환하여 표현하는 비지도 학습 기반의 분산 표상 방식이다. 이는 단어의 분산된 표현을 학습하기 위한 신경망 모델로 설명되며, 특정 차원에 단어를 임베딩하여 문맥상 의미가 보존되도록 학습시키는 과정을 포함한다[21]. 이를 통해 각 단어는 연속적인 값의 벡터로 표현되며, 의미적으로 유사한 단어들은 벡터 공간에서 서로 근접하게 배치된다[22]. Skip-gram 모델은 특정 문맥에서 단어의 출현 확률을 극대화하면서 학습하는 방식으로, 이는 단어 간의 유사도를 효과적으로 분석할 수 있는 최적화된 모델로 간주된다[19].

이번 연구 역시 언론에 나타난 메가시티에 대한 담론과 소비자 발신 채널인 SNS 피드를 크롤링한 텍스트 데이터를 담론으로 활용하여, TF-IDF를 활용한 연관어 분석과 워드클라우드, 토픽모델링, Word2Vec에서 중심단어로 주변단어를 예측하는 skip-gram 알고리즘을 이용하여 토픽별 키워드에 대한 집중도를 알아보려 하였다.

이 과정에서 언론 발신과 소비자 발신에서 이야기하는 메가시티의 내용은 어떤 연관어로 구성되어 있으며 가장 연관이 높은 단어는 무엇임을 확인할 수 있다. 이와 함께 어떤 토픽으로 이 연관어들이 묶이는지를 살펴보고 묶인 개별 토픽들의 단어별 유사도를 시각화하여 유사도가 높은 토픽이 어떤 것인지 확인하여 메가시티와 가장 연관성이 높은 토픽과 그렇지 못한 토픽을 살펴보았다. 이를 통해 메가시티가 둘러싼 다양한 의미에 대해 기업과 소비자 측면에서 비교해 볼 수 있을 것이다.

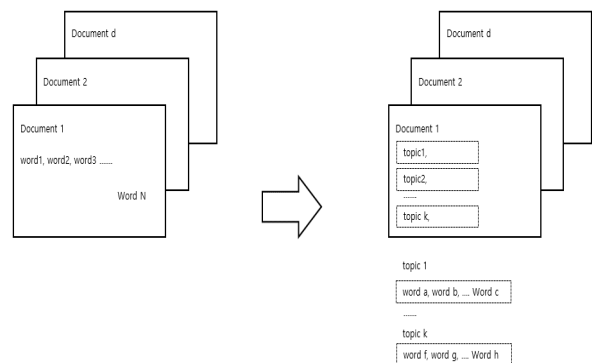


그림 1. 토픽모델링 과정  
Fig. 1. Process of topic modeling

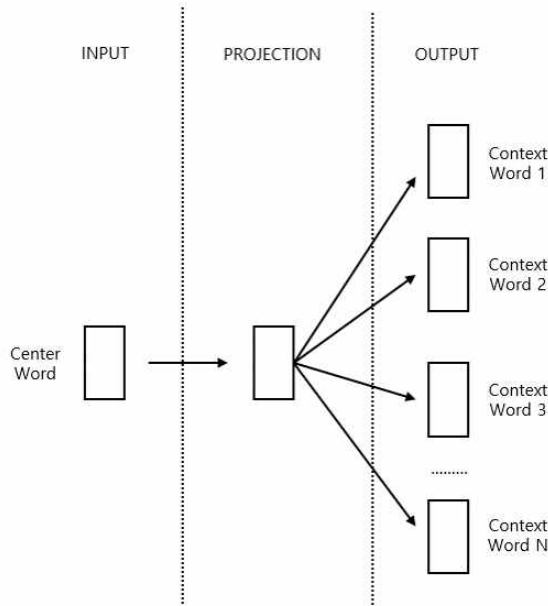


그림 2. 스킵그램 과정  
Fig. 2. Process of skip-gram

### 2-4 선행 연구

Word2Vec와 LDA와 같은 토픽모델링 기술은 온라인 리뷰나 뉴스 분석, 서비스 품질 평가 등 방대한 텍스트 데이터를 분석하여 관광객이나 이용자의 실제 경험과 인식을 보다 심도 있게 이해할 수 있는 혁신적인 방법론으로 부상하고 있다. 행정도시를 중심으로 분석했던 연구를 살펴보면, 서혜선은 LDA와 Word2vec 방법론을 이용하여 ‘의정부시’ 키워드 기반의 소셜 텍스트 데이터(네이버, 다음의 카페, 블로그, 지식인, 웹문서 등)에 대한 토픽모델링 분석 및 시각화를 실시하고, 선정된 토픽별 감성분석을 수행하였다[22]. 분석 결과, ‘입시교육’, ‘이사/수리’, ‘웨딩’, ‘취미/레저’, ‘행정/기관’, ‘음식/맛집’ 등 6개의 토픽이 도출되었다. ‘입시교육’ 토픽에서는 부정적, 중립적 단어들의 비율이 높았으며, ‘음식/맛집’ 토픽에서는 긍정적 어휘의 비율이 매우 높게 나타났다. 이해진과 강영옥은 SNS상의 부산 관련 텍스트 데이터에서 관광 토픽을 추출하기 위해 토픽모델링 기법을 활용하였고, 그 결과 ‘시장/음식거리’, ‘문화유산/명소’, ‘전망/조망’, ‘문화/축제’, ‘공원/자연경관’ 등의 9개 카테고리를 식별하였다. 이들 연구는 지역 사회 내에서 발생하는 다양한 대화와 이슈를 분석함으로써, 지역 사회의 관심사와 문제점을 파악할 수 있는 기반을 마련했다[23].

이민철과 김혜진은 대량의 뉴스 기사 데이터를 활용하여 주요 사건을 감지하고 사건 간의 관련성을 평가함으로써 복잡한 현시적 사건 정보를 요약하였다. 또한, Word2Vec과 LDA 토픽모델링 기법을 통해 한국어 텍스트 분석의 정확도를 높였을 뿐만 아니라 방대한 양의 데이터 속에서 사건 간의 숨겨진 관련성을 밝히고 있다[24].

광고 및 마케팅 영역에서도 활발하게 연구가 진행되었다. 차영란은 인플루언서의 뒷광고 논란 전후에 대한 댓글을 비교, 분석함으로써 대중의 인식 변화를 살펴보았으며[25], 나정희와 이우철은 1인 마켓의 인플루언서들이 인스타그램 피드에 올리는 상업적 메시지와 비상업적 메시지의 토픽모델링을 통해 이들 간에 어떤 차이가 있는지를 보는 탐색적 연구를 진행한 바 있다[16].

스마트시티와 관련하여 등은 Scopus에 게재된 스마트시티 관련 연구 논문 12,400개를 구조적 토픽모델링을 활용하여 분석하였다[11]. 15개의 주요 연구 토픽을 도출하였으며, 토픽 발견률에 따라 ‘Machine Learning’, ‘IoT & Security’ 등을 핫토픽으로, ‘Social Media’, ‘GIS’ 등을 콜드토픽으로 분류하여 현재까지의 스마트시티 연구 동향을 파악하고 향후 연구 및 정책 방향성을 제안하였다. 박건철과 이치형 역시 스마트시티에 대한 연구 경향을 파악하고 이를 바탕으로 스마트시티에 대한 정책적, 산업적, 학술적 방향성을 제시하였다[12]. 이들 역시 Scopus와 Springer DB에서 수집한 11,527건의 스마트시티 관련 학술논문을 LDA기반 토픽모델링으로 분석했다. 연구주제는 서비스, 기술, 시민·사용자 관점의 세 가지 범주와 8가지 세부주제로 분류되었으며, ‘시민중심 스마트시티 추진을 통한 지속가능성 확보’ 관련 주제가 가장 많이 언급되었다. 데이터와 프라이버시에 관련된 연구가 중심성이 높은 것으로 나타났다. 한편 황서이와 심지원은 2016년부터 2019년까지 국내 언론에 나타난 스마트시티 관련 보도기사의 시기별 경향을 파악하고, 언론을 통해 스마트시티가 어떻게 의미화되는지 의미연결망 분석을 활용해 확인하였다[13]. 연구 결과, 스마트시티는 다양한 영역과 연결되며, 사회적 발전 도구로 의미화되고 있었다. 특히, 2018년 스마트시티 혁신성장동력 프로젝트 시작 이후 스마트시티에 대한 논의가 활성화되었으나 여전히 기술 중심적 관점에서 논의되고 있다는 점을 지적하였다.

여러 연구의 결과처럼 본 연구도 토픽모델링과 같은 자연어 처리 기술 및 기계 학습 모델이 메가시티라는 담론에서 주요 이슈를 발견하는 데 충분히 적용될 수 있을 것으로 판단하였다.

## III. 연구설계

### 3-1 연구 프레임워크

이 연구는 빅데이터를 활용하여 메가시티라는 이슈에 대한 언론과 사회적 반응에 대한 담론을 분석하는 것을 목표로 하고 있으며, 총 네 단계의 절차를 통해 진행되었다.

1단계는 ‘연구자료 수집’ 단계로 빅데이터 분석 전문기업인 타파크로스의 ‘트렌드업(TrendUp4)’ 프로그램을 활용해 다양한 데이터를 수집, 정제, 분석하였다. 이 과정에서 사용된 데이터 수집기는 크게 언론 발신과 소비자 발신 두 가지 유형

으로 구분된다. 언론 발신 데이터 수집기는 네이버뉴스에 등록된 언론사 및 추가적으로 513개의 언론사를 대상으로 데이터를 수집하였다. 이를 통해 언론에서 보도된 다양한 뉴스와 정보들을 포괄적으로 수집할 수 있었다.

언론기사의 경우 원문이 아니라 원문의 요약물 사용하였다. 이는 언론기사 원문이 SNS 원문보다 길기 때문에 데이터 양의 차이에서 발생하는 오류를 방지하기 위함이었다. 원문 요약 기능은 원문에서 핵심내용을 포함한 특정 부분(200글자)을 잘라서 표시한다. 검색키워드가 원문 전체 중 여러 곳에 출현할 경우, 가장 많이 출현한 부분을 선택하여 잘라내는 방식이다. 요약문 추출방법은 원문에서 검색단어가 위치한 조건에 따라 점수를 부여한다. 점수가 가장 높은 검색단어를 중심으로 앞뒤로 350바이트 범위 내에서 요약문을 제공하며, 단어가 중간에 잘리지 않도록 설정되었다.

반면, 소비자 발신 데이터 수집기는 일반 사용자들이 소셜 미디어 플랫폼에 게시하는 콘텐츠를 대상으로 삼았다. 구체적으로 인스타그램, 유튜브, 블로그, 커뮤니티, X(트위터) 등의 SNS 채널에서 데이터를 수집하며, 이를 통해 일반 대중의 의견, 트렌드, 반응 등을 파악할 수 있었다. 이러한 양방향 접근은 보다 폭넓고 다양한 시각에서 정보를 수집하고 분석할 수 있도록 만든다. 따라서 본 연구에서는 ‘메가시티’ 처음 언급된 2023년 9월부터 2024년 1월 30일 동안의 언론 발신과 소비자 발신 담론을 대상으로 분석하였다. 언론 발신 데이터 분석을 위하여 총 16,228건의 원문 데이터를 찾았으며, 이중 최근 순으로 5,000개를 분석 데이터 세트로 활용하였다. 소비자 발신 데이터 분석을 위하여서는 총 18,170건의 원문 데이터를 찾았고 이 중 대중매체와 동일한 기준으로 최신순 5000개의 원문을 데이터 세트로 활용하였다.

2단계에서는 ‘수집된 자료의 데이터 전처리 작업’을 진행하였다. 이 과정에서 자연어 처리를 위하여 파이썬(Python)의 형태소 추출기 카이(khahii)를 이용하여 형태소 분석을 진행하였다. 형태소 분석은 수집된 텍스트 데이터에서 특수기호나 숫자 등을 제외하고 의미 분석이 가능한 명사, 동사, 수사, 대명사, 조사나 형용사, 감탄사 형태로 분류하는 작업과 함께 분석에 방해가 되는 불용어를 제거하는 작업을 수행하였다. 이러한 전처리 과정은 데이터의 질을 향상시키고 분석의 정확도를 높이는 데 중요한 역할을 한다.

3단계는 ‘자료 분석’ 과정으로 연관어 분석과 토픽모델링 분석을 주요 분석 방법으로 사용하였다. 연관어 분석을 통해 특정 키워드나 주제와 관련된 단어 사이의 연관성을 파악하였고, 토픽모델링을 통해서 대량의 텍스트 자료 속에서 숨겨진 주제들을 발견하고 이들 간의 관계를 이해하는 데 초점을 맞췄다. 이러한 분석을 통해 대규모 텍스트 데이터에서 의미 있는 패턴과 트렌드를 확인하였다. 연관어 분석을 위한 1차 정제 단계에서는 언론기사와 SNS 데이터에서 뽑아낸 키워드를 TF-IDF 분석을 통해 정렬했으며, 이를 빈도수가 높은 순으로 나열하여 가장 자주 언급된 단어부터 순서대로 분류하였다.

마지막 4단계는 ‘토픽명 부여 및 결과 해석’ 단계로 분석을 통해 도출된 토픽들에 명확한 이름을 부여하고, 이를 바탕으로 결과를 해석하는 작업을 수행하였다. 특히, 다양한 발신자 간의 의견이나 주제에 대한 접근 방식에서 어떠한 차이가 있는지를 분석함으로써 해당 주제에 대한 사회적 반응의 다양성을 이해하는 데 중점을 두었다.

### 3-2 연구문제

본 연구는 언론기사와 SNS에서 도출된 주요 키워드와 토픽 사이의 차이를 분석하는 것이다. 이 과정에서는 언론기사와 SNS 게시물을 각각 분석한 후, 두 데이터에서 나타나는 키워드와 토픽의 차이점을 밝힌다. 이를 통해 메가시티에 대한 공식적인 논의(언론기사)와 대중의 인식(SNS) 사이에 어떠한 의견 차이가 있는지, 특정 주제에 대한 사회적 관심의 차원이 어떻게 다른지 분석하고자 하였다. 이러한 비교 분석은 메가시티에 대한 다양한 시각과 인식을 이해하는 데 중요한 기초 자료를 제공할 것이다. 본 연구에서 설정한 연구문제는 다음과 같다.

- 연구문제 1. 메가시티 관련하여 언론기사와 SNS에 나타난 주요 키워드는 무엇인가?
- 연구문제 2. 메가시티 관련하여 언론기사와 SNS에 나타난 주요 토픽은 무엇인가?
- 연구문제 3. 메가시티 관련하여 언론기사와 SNS의 주요 키워드와 토픽의 차이는 무엇인가?

## IV. 분석결과

### 4-1 메가시티 관련 언론사 및 SNS 비교 분석결과

<연구문제 1>을 확인하기 위해 언론기사와 SNS에 나타난 주요 키워드를 분석하였으며, 빈도분석과 이를 시각화한 워드클라우드를 제시하였다(표 1, 그림 3 참조).

메가시티 관련 연관어를 분석한 결과, 1위부터 20위까지의 순위에서는 큰 차이가 발견되지 않았으나 20위를 넘어서면서 언론기사에서는 언급되지 않았던 키워드 두 개가 소비자 발신 데이터에서 등장하는 것을 확인할 수 있었다. 구체적으로 소비자 발신 연관어 분석 결과, 25위에서 40위 사이에서는 언론기사에서 등장하지 않았던 ‘부동산(26위)’과 ‘아파트(37위)’ 키워드가 발견됐다. 이는 언론기사와 소비자 발신 데이터 간의 차이를 보여주는 중요한 지표로 소비자들 실재 관심을 가지고 있는 주제가 언론의 보도 범위를 넘어설 수 있음을 의미한다. 즉, 이 두 키워드의 등장은 메가시티에 대한 일반 국민의 인식이 국가경쟁력이나 지역 균형 발전이라는 거시적 의미를 넘어 주로 경제적 문제, 특히 자신의 자산과 관련된 구체적인 경제적 이슈로 인식되고 있음을 시사한다.



표 1. 언론기사와 SNS에 나타난 메가시티 연관어 및 버즈 수 분석 결과

Table 1. The results of the analysis on the associated terms and buzz volume of megacities appearing in news articles and social media

Rank	News Articles		SNS		Rank	News Articles		SNS	
	Related Term	Buzz Count	Related Term	Buzz Count		Related Term	Buzz Count	Related Term	Buzz Count
1	Inclusion	5,510	Inclusion	3,648	21	Nation	1,666	Administration	995
2	Citizen	5,316	Citizen	2,926	22	Busan-Ulsan-Gyeongnam	1,591	Balance	974
3	City	4,416	City	2,472	23	Problem	1,556	Issue	960
4	Metropolitan Area	3,884	Metropolitan Area	2,443	24	Politics	1,547	Competitiveness	940
5	Discussion	3,614	Local	2,217	25	Special Committee	1,528	Nation	917
6	Development	3,348	Busan	1,763	26	Gwangju	1,472	Real Estate	899
7	Concept	3,314	Concept	1,661	27	Issue	1,449	Center	898
8	Local	3,219	Discussion	1,610	28	Competitiveness	1,437	Democratic Party	882
9	Project	2,576	Development	1,584	29	Population	1,316	Sejong City	881
10	Busan	2,460	General Election	1,542	30	Criticism	1,262	Future	859
11	Balance	2,328	Policy	1,477	31	Guri City	1,261	Traffic	842
12	Gimpo City	2,288	Busan-Ulsan-Gyeongnam	1,443	32	Resident	1,249	Gwangju	785
13	Policy	2,221	Gimpo City	1,378	33	South Gyeongsang Province	1,207	Opposition	739
14	Chungcheong	2,219	Project	1,242	34	Center	1,193	Nationwide	736
15	Ruling Party	2,117	Chungcheong	1,234	35	Daejeon	1,153	Problem	736
16	General Election	2,091	Population	1,233	36	Governor	1,137	Hanam City	731
17	Administration	1,960	Gyeonggi Province	1,116	37	Opposition	1,127	Apartment	725
18	Democratic Party	1,755	Politics	1,091	38	Traffic	1,096	Cho Kyung-tae	723
19	Cho Kyung-tae	1,732	Ruling Party	1,067	39	Concentration	1,094	Daejeon	711
20	Gyeonggi Province	1,689	Guri City	997	40	Business	1,068	Business	703



\*This word cloud image was automatically derived from system.

그림 3. 메가시티에 대한 워드클라우드 분석 결과

Fig. 3. Word cloud analysis results on megacities

다음으로 <연구문제 2>를 확인하기 위해 언론기사와 SNS에 나타난 주요 토픽을 확인하였다. 먼저 언론기사에 대한 형태소 분석을 기반으로 TF-IDF 분석을 실시하여 연구문제 1에 대한 연관어들을 도출하였다. 이러한 연관어들의 유사성을 기준으로 하여 주제를 몇 개의 토픽으로 구분할 수 있는지 분석하기 위해 젠심(Gensim) 패키지를 활용하여 coherence 분석을 진행하였다(그림 4).

그래프에서는 4개의 토픽이 가장 높은 coherence 값을 보여주어 분석상 유효한 것으로 나타났으나, 주제의 다양성을 확보하기 위한 차원에서 두 번째로 높은 점수를 보인 6개의 토픽을 최종적으로 선택하였다. 이는 토픽모델링의 결과를 더

표 2. 언론기사에 나타난 메가시티에 대한 토픽모델링 결과

Table 2. Topic modeling results for 'megacity' shown in news articles

Rank	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
1	Project committee	Chungcheong	Resolution	Trigger	Emergency	Ulsan
2	Project Chairman	Megaregion	East-West	Main Point	Briefing Room	Gyeongnam
3	Project commi	Activation	Capital	Privilege	Pohang Office	Yangsan
4	Chairman Yesterday	Cornerstone	Specific	Hot Issue	New Year	Gimhae
5	as	Construction	Phenomenon	Gimpo-Seoul	Press Room	Gimhae-Yangsan
6	Special Committee	Completion	Northeast Daejeon	Push Through	Yun Sang-hyun	Gateway
7	Block	All Chungnam / Chungbuk	Gap	Others	Jo O-seop	Kick Away
8	Plan	Traffic Network	Improvement	Seoul Special City	Briefing Room	Hong Tae-yong
9	Stepping Stone	Daegu-Gyeongbuk	Enjoyment	Controversy	National Assembly Hall	Summon
10	Cho Gyeong-tae			Recent	Briefing	Busan Also

욱 풍부하게 하고 연구 대상의 다양한 측면을 탐색하는 데 기여할 수 있다[22].

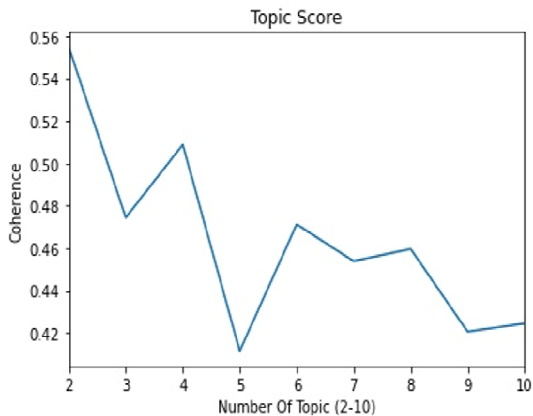
표 2에 제시된 토픽모델링 결과를 바탕으로 언론기사 데이터를 분석한 결과, 다음과 같은 토픽들이 도출되었다. 토픽 1은 '프로젝트', '조정태', '의원' 등의 키워드가 주로 등장하며 주로 <정치적> 관련 주제에 초점을 맞추고 있다. 이러한 키워

드 구성은 해당 토픽이 메가시티가 정치적 프로젝트나 정치인의 활동과 관련된 내용을 포함하고 있음을 보여준다. '충청권', '도시', '메가시' 등의 단어로 구성된 토픽 2는 <충청권 경제> 관련 주제를 다루고 있으며, 해당 토픽에서는 충청권의 경제적 상황이나 발전 및 도시화와 관련된 논의가 주를 이루는 것으로 보인다. 토픽 3에서는 '지역', '인구', '발전'과 같은 단어가 등장하면서 <지역균형발전>과 관련된 주제를 다루고 있다. 이 토픽은 인구 분포와 지역 발전의 상호작용에 대한 분석이 포함될 가능성이 높음을 시사한다. 토픽 4에서는 '서울', '편입', '메가시티' 등의 키워드를 포함하고 있으며, <서울편입>과 관련된 주제를 다루고 있다. 특히, 서울과 인접 지역 간의 관계 변화나 행정구역 조정과 같은 내용이 포함될 것으로 예상된다. 토픽 5는 '기자', '전북', '공식' 등의 단어가 포함된 토픽으로 <언론플레이>와 관련된 주제로 해석된다. 이는 언론의 역할과 그 영향력, 특히 지역 사회에서의 언론 활동과 관련된 논의를 포함할 가능성이 높다. 토픽 6은 '부산', '부울', '서울' 등의 키워드가 등장하며, <부울경>과 관련된 주제로 정의된다. 이는 부산, 울산, 경남 지역의 상호작용과 지역적 특성에 대한 분석이 포함될 것으로 보인다.

본 연구에서 분석한 여섯 가지 토픽으로 볼 때, 언론사는 메가시티에 대한 논의를 주로 정치적 관점과 지역 균형발전의 측면에서 다루었으며, 특히 '부울경'으로 일컬어지는 부산-울산-경남과 충청, 전라북도 등 기존에 논의가 있었거나 실행 가능성이 있는 지역에 초점을 맞춘 보도가 많았음을 확인할 수 있었다. 흥미로운 점은 메가시티 이슈에서 초기 논의의 대상이었던 '김포'에 대한 키워드가 추출되지 않았다는 것이다. 이에 대한 논의와 해석은 결론에 기술하였다.

다음으로는 SNS에 나타난 주요 토픽을 분석하기 위해 coherence 분석을 진행하였다(그림 5). <표 3>에 제시된 토

Text(0, 0.5, 'Coherence')



- Topic 1: [{'프로젝트', 208.93}, {'조정태', 131.67}, {'의원', 116.76}]
- Topic 2: [{'충청권', 273.75}, {'도시', 219.15}, {'메가시', 207.83}]
- Topic 3: [{'지역', 208.35}, {'인구', 187.68}, {'발전', 155.5}]
- Topic 4: [{'서울', 1162.42}, {'편입', 691.07}, {'메가시티', 522.63}]
- Topic 5: [{'기자', 208.32}, {'전북', 110.9}, {'공식', 109.08}]
- Topic 6: [{'부산', 211.19}, {'부울', 180.07}, {'서울', 173.9}]

\*This image was automatically derived from Gensim package.

그림 4. Coherence 분석 결과와 6개의 토픽의 주요 키워드  
Fig. 4. Coherence analysis results and key words of 6 topics

표 3. SNS상에 나타난 메가시티에 대한 토픽모델링 결과

Table 3. Topic modeling results for 'megacity' shown on SNS

Rank	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
1	Decrease	Kim Tae-hum	Rent	Rent	Request	New York	Include	Self-sufficiency
2	Single City	Governor of Chungnam	Trade	Deposit	Significant	Beijing	Topic	Decrease
3	Indicate	Kim Moon-soo	Real Estate	Real Estate	Respondent	Connectivity Right	Start	Surrounding
4	Self-sufficiency	Province	Property	Increase Trend	Draw	London	Keep in Mind	Area
5	Beyond	Overambition	Lease	Cool Down	More	Giant	Report	Indicate
6	Tokyo	Kim Dong-yeon	Feature	Information	Democratic Party	Increase	Emerge	Small and Medium
7	More than	Nam Kyung-pil	Complex Name	Feature	Discard	Future Type	Among	Main
8	Influx	Affiliation	Record	Daebang	Naive	Equip	Heat Up	More than
9	Future Type	Chungnam Province	Villa	Trade	Park han-dae	International	Assign	Port
10	Area	During Tenure	Daebang	Megacity	Act	Single City	Alias	Giant

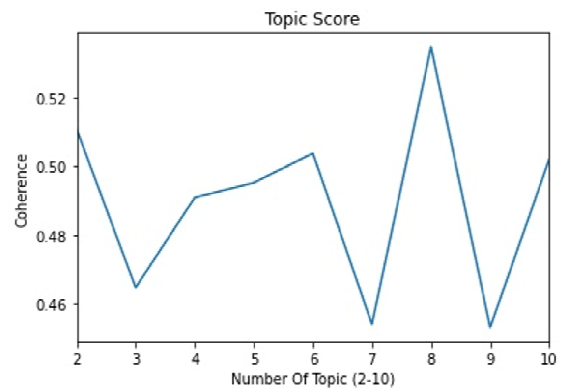
토픽모델링 결과를 기반으로 SNS 데이터를 분석한 결과, 총 8개의 토픽이 도출되었다.

토픽 1은 '인구', '보유', '커지' 등의 단어로 구성되어 있으며, <인구감소>와 관련된 주제를 나타낸다. 이는 지역 인구의 변화와 그에 따른 사회적, 경제적 영향을 반영하는 것으로 해석할 수 있다. 토픽 2는 '지사', '보이', '부울' 등의 단어를 포함하고 있으며, <정치적> 관점에서 메가시티에 대한 충청남도나 경기도 등 다양한 지역의 반응을 보여주는 주제이다. 토픽 3은 '정보', '아파트', '매매' 등의 단어로 이루어져 있으며, 부동산과 아파트 매매에 초점을 맞춘 <자산> 관련 주제를 나타낸다. 이는 메가시티가 주거 시장에 미치는 영향을 반영하는 것으로 볼 수 있다. 토픽 4는 '매물', '부동산', '천공' 등의 단어가 포함되어 있으며, <부동산> 시장의 현실적인 문제에 대한 주제를 다룬다. 이는 세 번째 토픽과 함께 부동산 시장의 동향과 그 변화를 중심으로 한 논의를 나타낸다. 토픽 5는 '생각', '내년' 등의 단어로 구성되어 있으며, <미래>와 관련된 주제를 나타낸다. 토픽 6은 '도쿄', '그렇', '사람' 등의 단어로 이루어져 있으며, 해외 도시 사례를 통해 <메가시티의 장점과 가능성>을 탐색하는 주제로 해석된다. 여기서 '그렇'의 키워드가 발견되었던 것은 한국의 메가시티도 다른 국가의 메가시티와 유사하다는 의미로 여러 차례 쓰였기 때문으로 보인다. 토픽 7은 '서울', '편입', '메가시티' 등의 단어를 포함하며, 현재 진행 중인 <서울편입>과 관련된 주제를 다루고 있다. 이는 메가시티 구상의 핵심 요소 중 하나로서의 서울 편입에 대한 논의를 반영한다. 토픽 8은 '도시', '서울', '메가시티' 등의 키워드가 포함된 <도시주변> 주제를 다루며, 도시 주변의 변화와 메가시티 구상이 주변 지역에 미치는 영향을 탐구하는 주제로 보인다.

토픽모델링 결과의 시각화 과정에서 각 토픽을 구성하는

단어들의 집중도와 분포는 중요한 분석 지표로 활용되었다. 이를 위해 단어 간의 거리를 사전에 정의하고 각 토픽 내에 동일한 수의 단어를 배치하여 토픽별 공간의 넓이를 비교하였다. 이러한 방식으로 키워드들이 서로 밀접하게 위치할수록, 즉 X축 상의 거리(inch)가 짧을수록 해당 토픽의 단어 집중도가 높다고 해석할 수 있다.

빈도분석을 통해 카이스퀘어 검증을 실시한 결과, 통계적으로 유의한 차이가 발견되었다( $\chi^2 = 1307.77, p < .001$ ).



- Topic 1: [('인구', 82.93), ('보유', 60.74), ('커지', 56.32)]
- Topic 2: [('지사', 89.87), ('보이', 74.89), ('부울', 70.96)]
- Topic 3: [('정보', 151.57), ('아파트', 133.65), ('매매', 125.46)]
- Topic 4: [('매물', 117.17), ('부동산', 102.14), ('천공', 94.81)]
- Topic 5: [('생각', 93.0), ('내년', 86.1), ('대구', 71.96)]
- Topic 6: [('도쿄', 77.49), ('그렇', 76.96), ('사람', 72.62)]
- Topic 7: [('서울', 558.79), ('편입', 362.5), ('메가시티', 278.6)]
- Topic 8: [('도시', 312.64), ('서울', 305.16), ('메가시티', 275.86)]

\*This image was automatically derived by a Gensim package.

그림 5. Coherence 분석 결과와 8개의 토픽의 주요 키워드  
Fig. 5. Coherence analysis results and key words of 8 topics



구체적으로 언론기사에서는 긍정적인 평가의 빈도가 11,574회로, 소비자 리뷰의 긍정적인 평가 빈도 10,350회에 비해 상대적으로 더 높게 나타났다. 반면, 부정적인 평가의 빈도는 언론기사에서 3,357회, SNS에서는 3,924회로, SNS가 언론기사에 비해 더 부정적인 경향을 보였다. 또한, 중립적인 평가의 빈도는 언론기사에서 1,297회였던 반면, 소비자 리뷰에서는 3,896회로 높았다. 이는 언론기사와 SNS에서 각기 다른 기준으로 미디어에서 메가시티 관련 이슈를 다루고 있으며, 미디어에 따른 인식에 차이가 있음을 의미한다.

## V. 결 론

하버드대의 에드워드 글레이저는 그의 저서 ‘도시의 승리 (Triumph of the city)’에서 도시에 대해 ‘인류를 더욱 풍요롭고 행복하게 만드는 가장 위대한 발명품’이라고 극찬했다 [1]. 도시가 인류의 잠재력을 끌어내 문명과 경제 발전의 원동력으로 작동하기 때문이다. 특히 성공한 도시들은 물리적 인프라가 아닌 인재와 기술, 아이디어와 같은 인적자본을 유인함으로써 혁신의 중심으로 부상했다는 점을 강조하면서, 교육과 신기술이 집적된 미국 실리콘밸리와 인도 벵갈루루 (Bengaluru)를 그 증거로 들었다.

본 연구는 ‘메가시티 서울’에 대한 국민적 관심이 고조되는 가운데 언론 발신 메시지와 소비자 발신 메시지 간 차이를 밝힘으로써 메가시티에 대한 담론과 의미를 조망하고자 하였다. 이를 위해 언론기사와 SNS에서 다뤄진 메가시티에 대해 토픽모델링을 통해 분석하였다.

주요 연구 결과는 다음과 같다. 우선 언론기사와 SNS 두 채널 모두에서 ‘정치적’, ‘서울편입’이라는 공통된 토픽이 발견되었다. 이는 2024년 4월 총선을 앞두고 국가적 혹은 지역적 차원의 정책이나 정치적 변화가 국민의 일상생활에 지대한 영향을 미치고 있음을 반영한다.

다음으로 언론기사와 SNS에서 메가시티 이슈를 다루는 방식에 대한 토픽모델링 분석 결과, 두 플랫폼 간 주요 토픽과 키워드에서 상당한 차이가 발견되었다. 언론기사에서는 주로 ‘정치인’이나 ‘프로젝트 위원회’와 같이 메가시티 프로젝트의 정책적이고 기획적인 측면이 강조되었다. 이는 언론이 메가시티 관련 논의를 정책 결정 과정, 관련 정치인의 역할 및 프로젝트의 공식적인 진행 상황에 초점을 맞추고 있음을 나타낸다. 즉, 언론기사에서는 국가 경영 시스템의 관점에서 지방 균형발전이나 전국적 차원의 메가시티 구축 같은 대규모 국가 프로젝트의 장점과 비전을 중심으로 한 논조가 주를 이루었다. 반면, SNS에서는 ‘아파트’나 ‘부동산’과 같은 키워드가 발견되어, 언론기사에서는 다루어지지 않았던 주제들이 SNS 사용자들 사이에서 논의되고 있었다. 이는 SNS 사용자들이 메가시티 이슈를 논의할 때, 프로젝트가 개인의 생활과 직접적으로 연관된 부동산 가격이나 주거 환경의 변화와 같은 실생활에 미치는 영향에 더 많은 관심을 기울이고 있음을 보여

준다. 이러한 차이는 언론이 제시하는 메시지가 대체로 거시적인 국가 발전 관점을 반영하는 데 비해 일반 소비자들은 국가적 프로젝트에 대한 주목뿐만 아니라 자신의 삶과 밀접한 경제적 영향에도 큰 관심을 기울이고 있음을 시사한다.

한편 이번 연구에서 발견한 흥미로운 점은 ‘김포’ 토픽이 발견되지 않았다는 점이다. 김포를 서울에 편입시키겠다는 발표가 기존의 메가시티 이슈를 재점화했음에도 불구하고 메가시티에 대한 토픽모델링 분석에서 김포에 대한 키워드 순위가 상대적으로 높지 않았다는 사실은 몇 가지 해석을 가능케 한다. 첫째, 메가시티와 관련된 논의가 김포의 서울 편입 문제보다 더 넓은 범위의 주제를 포괄하고 있을 수 있다. 즉, 메가시티의 구상과 계획이 김포에 국한되지 않고 다양한 지역과 관련된 쟁점들을 포함하고 있을 가능성이 크다. 둘째, 메가시티에 대한 대중의 관심이나 언론의 보도 방향이 김포의 서울 편입과 같은 구체적 사안보다는 메가시티의 전반적인 비전이나 다른 지역적 요소에 더 초점을 맞추고 있음을 보여준다. 이러한 결과는 메가시티에 대한 논의의 다양성과 복잡성을 보여주며, 이러한 이슈가 단일 지역의 문제가 아닌 보다 광범위한 도시 계획과 발전 전략의 일부임을 시사한다.

마지막으로 메가시티에 대한 언론기사와 SNS 간 긍정, 부정, 중립 평가 차이를 확인하였을 때, 언론기사에 나타난 메가시티 이슈가 SNS상에 나타난 메가시티 이슈에 비해 긍정적인 내용을 더 많이 포함하는 반면, SNS상의 메가시티 이슈는 부정적인 내용과 중립적인 내용을 더 많이 담고 있었다. 이는 SNS상의 메가시티 이슈가 언론기사에서 다루는 메가시티 이슈보다 개인의 다양한 경험과 주관적인 평가를 더 많이 반영하며, 그로 인해 더 폭넓은 감정 스펙트럼을 포함할 가능성을 나타낸다. 언론기사는 상대적으로 일관된 기준에 따라 긍정적인 정보를 전달하는 반면, SNS상의 메가시티 이슈는 개인의 경험에 기반한 생생한 반응을 제공하기 때문에 나타난 차이로 판단된다. 따라서 향후 언론기사와 SNS상의 메가시티 이슈의 내용을 분석하고 활용할 때 각각의 특성을 고려해야 할 필요가 있다.

본 연구의 이론적 시사점으로는 토픽모델링을 통해 얻은 결과가 메가시티 관련 정책 개발과 실행에 있어서 실질적인 통찰력을 제공할 수 있음을 보여준다는 점이다. 또한, 언론기사와 SNS라는 두 채널의 데이터 소스로부터 획득한 정보를 통합적으로 분석하는 데 있어 토픽모델링이 어떻게 활용될 수 있는지에 대한 이해를 심화시켰다. 이는 향후 메가시티와 같은 대규모 도시 개발 프로젝트의 기획과 평가에 있어 중요한 방법론적 기여를 할 수 있을 것이다.

이를 토대로 한 실무적, 정책적 시사점은 메가시티 구축이 단순히 행정구역의 재편을 넘어 지역의 사회적 인프라 개선이나 경제적 자립성 강화, 그리고 궁극적으로는 국가경쟁력 제고에 기여할 수 있는 실효성 있는 전략이 되어야 한다는 점을 밝혔다. 특히, 메가시티와 같이 다양한 이해관계가 엮인 국가적 프로젝트를 위해서는 언론과 SNS에서의 광범위한 의견과 반응을 면밀히 관찰하고 이를 정책에 반영하는 것이 무엇

보다 중요할 것으로 사료된다. 그래야만 정책 결정자와 일반 대중 모두의 관심사와 우려를 충족할 수 있는 방안을 수립할 수 있기 때문이다.

이 연구는 메가시티의 다양한 이슈를 분석하기 위해 토픽 모델링을 활용하였지만, 몇 가지 한계점을 가지고 있다. 첫째, 토픽모델링은 대량의 텍스트 데이터에서 통계적인 패턴을 분석하여 토픽을 도출하기 때문에 추출한 토픽이 실제 메가시티에 대한 의미와 중요성을 정확히 반영하지 못할 수 있다. 이는 토픽의 수와 같은 파라미터(parameter)에 대한 사전 지식과 조정이 필요함을 시사한다.

둘째, 본 연구는 주로 언론기사와 개방형 SNS 데이터에 국한되어 수행되었기 때문에 두 매체에서 제기되지 않은 이슈나 의견이 포함되지 않을 수 있다. 특히, 후속 연구에서 다양한 이해관계자의 의견을 폭넓게 수집하고 이를 분석하기 위해서는 폐쇄형 SNS 데이터 등 더 넓은 범위의 데이터 소스를 포함하는 것이 필요하다.

셋째, 메가시티 프로젝트와 같은 복잡한 사회적 현상을 완전히 이해하기 위해서는 텍스트 데이터 분석뿐만 아니라 경제적, 사회문화적, 환경적 요인들을 포함한 다학제적 접근이 필요하다. 토픽모델링 결과를 다른 형태의 데이터와 통합하는 것은 이러한 현상에 대한 보다 깊은 이해를 제공할 수 있을 것이다. 이러한 한계점들을 인지하고 해결하기 위한 노력은 향후 연구에서 메가시티 관련 이슈에 대한 더 깊이 있는 분석과 이해를 도모하는 데 중요할 것으로 사료된다.

## 참고문헌

[1] G. Edward, *Triumph of the City*, New York, NY: Penguin Books, 2012.

[2] S. K. Cho and D. Cha, "Study on Threats and Defense Policy According to Appearing Mega City in the Future Korea," *The Journal of Social Convergence Studies*, Vol. 4, No. 3, pp. 11-16, June 2020. <https://doi.org/10.37181/JSCS.2020.4.3.011>

[3] Gyeonggi Research Institute. A Megacity Strategy for Living in the Capital Region and the Provinces [Internet]. Available: <https://www.gri.re.kr/web/contents/issdiag.do?schM=view&page=11&viewCount=10&schPrjType=ISS&schStartYear=&schEndYear=&schSubj1=&schSubj2=&schProjectNo=20200433&schBookResultNo=14564>.

[4] Business Post. People Power Party's 'Seoul Mega City' Party's Opinion Promotes, Will It Be a Game of Thrones in the General Election? [Internet]. Available: [https://www.businesspost.co.kr/BP?command=article\\_view&num=331573](https://www.businesspost.co.kr/BP?command=article_view&num=331573).

[5] Yonhapnewstv. "According to the Will of the Citizens" Vs. "Strategy for the General Election"... Opposition and Opposition Parties Battle over 'Mega City Seoul'. [Internet]. Available: <https://www.yonhapnewstv.co.kr/news/MYH202>

31106018200641.

[6] E. Goffman, *Frame Analysis: An Essay on the Organization of Experience*, New York: NY, Harper & Row, 1974.

[7] Y. J. Lee, G. Park, and M. O. Yim, "A Study on Framing Effect Focusing on Empathy as the Mediator: How to Influence The Relationship between Framing on the Problem of the Aged and the Perceived Generation Gap, the Perceived Threat and Policy Acceptability," *Journal of Communication Science*, Vol. 15, No. 4, 271-306, December 2015. <https://doi.org/10.14696/jcs.2015.12.15.4.271>

[8] J. M. Lee, "Comparison of Social Awareness and Media Frames on the Government Budget during Conservative and Progressive Governments," *Journal of Governmental Studies*, Vol. 27, No. 3, pp. 165-191, December 2021.

[9] J. Y. Min, "The Amplifying Aspects of SNS Comments: An Exploratory Study through the Sentiment Comparison between News Site Comments and SNS Comments," *Information Systems Review*, Vol. 22, No. 4, pp. 163-184, November 2020. <https://doi.org/10.14329/isr.2020.22.4.163>

[10] C. H. Lee and Y. J. Kim, "Mega-City Discourses and Transformation of Nation-State: Theoretical Implications of the Megaregional Integration Project for the Spatial Governance," *21st Century Political Science Review*, Vol. 32, No. 3, pp. 1-24, September 2022. <https://doi.org/10.17937/topsr.32.3.202209.1>

[11] H. Park, D.-H. Kim, and S. Jang, "Research Trend Analysis on Smart City based on Structural Topic Modeling(STM)," *Journal of Digital Contents Society*, Vol. 20, No. 9, pp. 1839-1846, September 2019. <https://doi.org/10.9728/dcs.2019.20.9.1839>

[12] G. C. Park and C. H. Lee, "A Study on the Research Trends for Smart City using Topic Modeling," *Journal of Internet Computing and Services*, Vol. 20, No. 3, pp. 119-128, June 2019. <http://dx.doi.org/10.7472/jksii.2019.20.3.119>

[13] S. Hwang and J. Shim, "Semantic Network Analysis of 'Smart City' in Newspaper Articles: From 2016 to 2019," *Journal of Digital Contents Society*, Vol. 21, No. 5, pp. 941-950, May 2020. <http://dx.doi.org/10.9728/dcs.2020.21.5.941>

[14] The Joongang, "Rather than Concentrating on the Metropolitan Area, We Need to Save the Provinces"...Four Cities and Inces in Chungcheong Area Declare Mega-City Promotion [Internet]. Available: <https://www.joongang.co.kr/article/25206815#home>.

[15] Daily Newspaper. 'Mega City' Discussion, what Happens in TK?... Daegu City and Gyeongbuk Province Have 'Temperature Difference' [Internet]. Available: <https://www.imaeil.com/page/view/2023110816384483396>

[16] J. H. Na and W. C. Lee, "What Makes Top Influencers

- Different?: Content Analysis and Text Mining of 1-person Market Instagram,” *Journal of Practical Research in Advertising and Public Relations*, Vol. 16, No. 1, pp. 64-96, February 2023. <https://doi.org/10.21331/jprapr.2023.16.1.003>
- [17] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent Dirichlet Allocation,” *Journal of Machine Learning Research*, Vol. 3, pp. 993-1022, March 2003.
- [18] C. Kang, K. K. Kim, and S. Choi, “A Topic Analysis of Abstracts in Journal of Korean Data Analysis Society,” *Journal of the Korean Data Analysis Society*, Vol. 20, No. 6, pp. 2907-2915, December 2018. <https://doi.org/10.37727/jkdas.2018.20.6.2907>
- [19] S. H. Yoon and K. H. Kim, “Expansion of Topic Modeling with Word2Vec and Case Analysis,” *The Journal of Information Systems*, Vol. 30, No. 1, pp. 45-64, March 2021. <http://dx.doi.org/10.5859/KAIS.2021.30.1.45>
- [20] K. C. Lee, *Korean Embedding*, Seoul: Acon Publishing, 2019.
- [21] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient Estimation of Word Representations in Vector Space,” arXiv:1301.3781, September 2013. <https://doi.org/10.48550/arXiv.1301.3781>
- [22] H. Seo, “Topic Modeling and Visualization of Uijeongbu City SNS data using LDA & Word2vec Methodology,” *Journal of Korean Data Analysis Society*, Vol. 22, No. 6, pp. 2391-2403, December 2020. <https://doi.org/10.37727/jkdas.2020.22.6.2391>
- [23] H. Lee and Y. Kang, “Understanding Tourist’s Region of Attraction and Image of City through Social Network Data Analysis,” *Journal of the Korean Urban Geographical Society*, Vol. 23, No. 1, pp. 101-114, April 2020. <https://doi.org/10.21189/JKUGS.23.1.8>
- [24] M. Lee and H.-J. Kim, “Construction of Event Networks from Large News Data Using Text Mining Techniques,” *Journal of Intelligence and Information Systems*, Vol. 24, No. 1, pp. 183-203, March 2018. <https://doi.org/10.13088/jiis.2018.24.1.183>
- [25] Y.-R. Cha, “A Comparative Analysis of Comments Before and After the Controversy Over the 'Back Advertising' of Influencers: Focused on LDA and Word2vec,” *Journal of the Korea Contents Association*, Vol. 20, No. 10, pp. 119-133, October, 2020. <https://doi.org/10.5392/jkca.2020.20.10.119>



**이우철(Woo-Chul Lee)**

2002년 : 중앙대학교 대학원  
(언론학석사)  
2014년 : 중앙대학교 대학원  
(광고홍보학박사)

1997년~2008년: Lee & DDB  
2010년~2012년: DDB KOREA  
2020년~2023년: 지우컴퍼니 상무  
2023년~현 재: 타파크로스 부사장  
※관심분야 : 광고/매체 창의성, 소셜 빅데이터 분석, 데이터 기획/거래 등



**이장석(Jang-Suk Lee)**

2016년 : 중앙대학교 대학원  
(광고홍보학석사)  
2020년 : 중앙대학교 대학원  
(언론학박사)

2022년~현 재: 가천대학교 미디어커뮤니케이션학과 조교수  
※관심분야 : 뉴미디어, HCI, 미디어 정책 등