

# 빅데이터 분석을 통한 지방자치단체 정책이슈 도출 방법론

## Methodology of Local Government Policy Issues Through Big Data Analysis

김용진\*, 김도영\*\*

한국교통대학교 건설환경도시교통공학부\*, 수원시정연구원 도시공간연구실\*\*

Yong-Jin Kim(yj-kim@ut.ac.kr)\*, Do-Young Kim(suwon@suwon.re.kr)\*\*

### 요약

본 연구의 목적은 효율적이고 효과적인 정책 발굴 과정에서 빅데이터의 활용이 점차 중요해지는 현실에서 지방자치단체의 정책 이슈 발굴에 빅데이터 분석을 활용하는 방안을 제시하는 데 있다. 이를 위하여 본 연구에서는 수원시를 대상으로 지난 3년간의 수원시 약 18만 건의 기사를 분석하여 정책 이슈를 발굴하였으며, 이를 IPA 분석을 통해 정책의 우선순위를 평가하였다. 본 연구의 분석 결과는 신문 기사를 통한 반정형 빅데이터의 분석으로 전국의 주요 이슈와는 차별화된 지방자치단체의 차별화된 정책 이슈를 도출하는데 효과적임을 보였으며, 특히 도출된 정책 이슈들이 대부분 그 우선순위가 높은 것으로 평가되었다. 이처럼 본 연구에서 제시한 빅데이터 분석을 통한 정책 이슈 발굴의 방법론은 지방자치단체가 효율적인 정책 이슈를 도출하고 민의를 효과적으로 파악할 수 있음을 의미한다. 또한, 본 연구에서 제시한 방법론은 지방자치단체의 온라인 민원 자료, 주민 SNS 등 다양한 반정형, 비정형 빅데이터의 분석을 통한 정책 이슈 발굴에 적용이 가능할 것으로 기대된다.

■ 중심어 : | 빅데이터 | 지방자치단체 | 정책 | 형태소분석 | IPA 분석 |

### Abstract

The purpose of this study is to propose a method to utilize Big Data Analysis to find policy issues of local governments in the reality that utilization of big data becomes increasingly important in efficient and effective policy making process. For this purpose, this study analyzed the 180,000 articles of Suwon city for the past three years and identified policy issues and evaluated policy priorities through IPA analysis. The results of this study showed that the analysis of semi-formal big data through newspaper articles is effective in deriving the differentiated policy issues of different local autonomous bodies from the main issues in the nation. In this way, the methodology of finding policy issues through the analysis of big data suggested in this study means that local governments can effectively identify policy issues and effectively identify the people. In addition, the methodology proposed in this study is expected to be applicable to the policy issues through the analysis of various semi-formal and informal big data such as online civil complaint data of the local government, resident SNS.

■ keyword : | Big-Data | Local Government | Policy | Morphological Analysis | Importance-Performance Analysis |

\* 이 성과는 2016년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2016R1C1E2010619).

접수일자 : 2018년 08월 22일

심사완료일 : 2018년 09월 05일

수정일자 : 2018년 09월 05일

교신저자 : 김도영, e-mail : suwon@suwon.re.kr

## I. 서론

2012년 세계경제포럼(World Economic Forum)에서 빅데이터를 10대 기술 중 하나로 선정한 후, 빅데이터에 대한 논의와 활용이 활발해지고 있다. 특히, 빅데이터에 대한 활용은 기업에서 그 활용이 본격화 되었다. 구글과 같은 인터넷 기업들을 중심으로 빠르게 변화하는 소비자들의 심리를 파악해 기업의 정책에 반영하는 것이다[1]. 최근에는 미국, 일본 및 유럽 국가 등의 선진국을 중심으로 빅데이터를 활용한 정책 개발이 이뤄지고 있다. 교통, 안전, 보건, 복지 등의 분야를 중심으로 한 빅데이터 기반의 정책 개발은 국민들의 수요를 구체적인 자료에 기반하여 파악하고, 이를 바탕으로 최적화된 정책을 개발할 수 있다는 장점이 있다[2].

이에 우리나라에서도 2011년 ‘빅데이터를 활용한 스마트 정부 구현(안)’을 시작으로 2012년 방송통신위원회의 ‘빅데이터 서비스 활성화 방안’, 행정안전부의 ‘스마트 국가 실현을 위한 빅데이터 마스터플랜’ 등 빅데이터를 기반으로 한 정책 개발에 박차를 가하고 있다[3].

이러한 움직임에도 불구하고 지방자치단체에서의 빅데이터 활용은 매우 부족한 것이 현실이다. 남양주시는 인구 이동 자료를 활용하여 인구 유입에 관한 정책을 수립하기 위해 2015년 10월부터 2개월간의 자료를 분석하여 활용한 바 있다[4]. 경기도는 CCTV의 추가 설치에 앞서 효율적이고 보다 안전한 환경 구축을 위해 기존 CCTV 자료를 분석하여 사각지대를 우선하여 설치하기 위한 기초자료로 활용한 바 있다[5]. 이처럼 지자체 수준에서도 빅데이터의 활용이 시작되고 있으나, 아직 본격적인 빅데이터의 분석 단계에 접어들지 못하였으며 그 사례 역시 많지 않은 것이 사실이다[6].

이에 본 연구에서는 지방자치단체에서 빅데이터를 활용한 정책 개발의 프로세스를 제시하고, 정책의 우선순위를 도출하고자 한다. 이를 위해 본 연구에서는 수원시를 대상으로 텍스트마이닝 기법을 활용하여 수원시의 우선 정책 과제를 도출하고 전문가 설문조사를 실시하여 정책의 우선 순위를 도출할 것이다.

본 연구의 결과는 이후 지방자치단체가 보다 다양한 빅데이터를 활용하여 민의를 보다 적극적으로 살피고

자료 기반의 구체적이고 효율적인 정책을 개발하고 추진하는데 중요한 기초자료로 활용될 것으로 기대된다.

## II. 이론적 배경

### 1. 빅데이터

빅데이터에 대해 학계에서 합의된 정의는 없으나, Manyika(2011)는 일반적인 데이터베이스가 저장, 관리, 분석할 수 있는 범위를 벗어나는 광대한 규모의 데이터라고 정의한 바 있으며[7], Gantz(2013)는 다양한 종류의 대규모 데이터로부터 가치 있는 정보를 추출할 수 있는 기술이라고 정의한 바 있다[8]. 또한, 위키피디아는 “기존 데이터베이스 관리도구로 데이터를 수집, 저장, 관리, 분석할 수 있는 역량을 넘어서는 대량의 정형 또는 비정형 데이터 집합 및 이러한 데이터로부터 가치를 추출하고 결과를 분석하는 기술”이라고 정의하고 있다. 이처럼 빅데이터를 정의하는 공통적인 요소는 데이터의 규모와 이러한 거대한 데이터 속에서 가치 있는 정보를 찾아내는 기술로 정리할 수 있다.

이러한 정의에서 나타나듯이 빅데이터에 대해 일반적으로 크기, 속도, 형태를 3대 요소로 보고 있다. 우선, 크기는 단순한 데이터 크기 외에도 이를 분석하거나 처리하는데 어려움이 따르는 데이터의 급속한 증가를 그 특징으로 보는 것이다. 다음으로 속도는 데이터의 생산과 유통, 수집, 분석 속도의 증가에 따른 실시간 처리와 오랜 시간에 걸친 데이터를 수집, 분석하는 특징이라 할 수 있다. 마지막으로 형태는 기존의 데이터의 주된 형태가 정형화되어 있었던 반면, 빅데이터는 XML, HTML 등과 같이 데이터베이스 스키마를 포함하는 반정형 데이터와 사진, 오디오, 비디오 형식 및 로그파일과 같은 비정형 데이터가 포함되는 특징을 의미한다[9].

### 2. 빅데이터 기반 정책 개발 사례

앞선 서론에서 언급한 바와 같이 미국, 일본, 유럽 등지의 선진국들을 중심으로 빅데이터를 활용한 연구와 정책 개발이 활발히 이뤄지고 있다. 미국 LA의 경우에는 범죄로부터 시민을 보호하고 범죄 관련 업무를 최적



1977년 개발하였다. IPA 분석은 다양한 제품, 서비스를 비롯하여 정책에 대한 강점과 약점을 분석하는데 광범위하게 사용되고 있으며, 특히 정책에 대한 중요도와 성취 정도를 평가하는데 활용된다. IPA 분석 기법의 특징은 중요도와 성취도의 비교 평가값에 의하여 의사결정에 활용할 수 있다는 점이다.



그림 2. IPA 분석 기법

[그림 2]와 같이 1사분면은 정책의 중요도가 높은 반면, 성취도가 낮으므로 특히 집중해서 개선해야 하며 많은 노력과 관심이 필요하다고 이해할 수 있다. 2사분면은 정책의 중요도가 높고 성취도 역시 높으므로 현재의 수준을 유지할 필요가 있다. 3사분면은 정책의 중요도가 낮고 성취도도 낮으므로 제한된 자원과 노력의 투입이 필요하다. 마지막으로 4사분면은 정책의 중요도가 낮는데 비해 성취도가 높은 것으로 홍보 또는 인식 개선이 필요한 것으로 이해할 수 있다.

#### IV. 분석 결과

본 연구에서는 빅데이터 분석을 통한 수원시의 주요 정책 이슈를 발굴하기 위해 2015년부터 2017년까지의 주요 일간지 및 주요 지역 신문을 분석하였다. 우선, 주요 일간지의 해당 기간 동안 수원 관련 기사는 128,303 개이고, 지역 신문의 수원 관련 기사는 53,911건으로 나타났다. 수원 관련 기사들에 대한 형태소 분석으로 수원과 관련된 주요 키워드들을 도출하여 워드클라우드로 표현한 결과는 [그림 3]과 같다.

주요 일간지 및 지역 신문의 182,304건의 기사를 분

석한 결과에서는 2015년 발생한 수원시의 살인 사건과 메르스가 가장 주된 이슈로 나타났다.



그림 3. 2015~2017년 수원시 주요 이슈

이를 연도별로 보다 상세하게 살펴보면, 2015년에는 수원시 토막살인 사건으로 인한 강력범죄의 이슈가 가장 빈도수가 높았으며, 전국적인 이슈로 떠올랐던 메르스 관련 이슈가 도출되었다. 2016년의 경우에는 2015년 대비 보다 다양한 이슈들이 도출되었으며, 국회의원선거와 화성방문의해, 지진 등의 이슈가 나타났다. 마지막으로 2017년의 경우, 대통령선거, 청소년, 창업, 파복하우스, 마복하우스 등의 보다 다양한 키워드들이 나타났다. 이러한 이슈들은 수원과 관련되어 있는 이슈들로 메르스, 국회의원선거, 대통령선거와 같은 전국적인 이슈들이 함께 자리 잡기도 하였으나, 탄핵, 촛불시위 등 2017년 전국적인 주요 이슈는 수원시와 차이가 있는 것으로 나타났다.

다음으로 수원시의 주요 정책 이슈를 보다 구체적으로 발굴하기 위해, 도출된 키워드들을 그 특징별로 나누어 분류하였다. 분류한 결과는 1.소외계층 및 복지, 2.구도심 및 도시균형개발, 3.재난·안전, 4.고용 및 취·창업, 5.첨단산업, 6.환경, 7.농업, 8.문화·역사·관광의 총 8가지 분야 총 110여개의 이슈로 정리할 수 있다. 소외계층 및 복지는 청소년, 장애인, 여성, 외국인, 저소득층, 독거노인 등의 키워드를 포함하며, 구도심 및 도시균형개발은 수원의 도시재생, 마을만들기를 비롯하여 재개발, 재건축 등을 포함한다. 재난·안전은 살인사건, 성범죄를 비롯하여 메르스, 지진 등의 이슈를 담고 있으며 고용 및 취·창업은 취업, 창업, 일자리, 기업, 고용, 실업자 등의 이슈를 포함하고 있다. 첨단산업은 정보화, 스

마트, 산업혁명, 첨단산업단지, 반도체 등의 키워드를 포함하고 있고 환경에는 친환경, 미세먼지 등의 시민 관심이 담겨 있으며, 농업에는 수원의 농업기술센터, 청년농업인, 정조 등의 키워드가 포함되어 있다. 마지막으로 문화·역사·관광에는 관광, 능행, 세계문화유산, 문화제 등의 이슈가 담겨 있다.

이러한 정책 이슈들에 대한 IPA 분석을 실시하기 위하여, 관련 분야 교수, 석·박사급 연구원, 공무원, 시민 활동가 등을 대상으로 전문가 설문조사를 실시하였다. 설문조사는 2017년 12월 1일부터 2017년 12월 10일까지 온라인 설문조사를 실시하였으며, 총 설문 응답자는 70명이며, 이 중 일관성이 떨어지는 설문 부수를 제외한 총 66부의 응답을 분석하였다. 설문항목은 도출된 정책 이슈들에 대해 수원시에서 중요하게 다뤄야 하는 정도와 현재 대응 정도(성취도)에 대한 7점 척도의 질문으로 구성되어 있다.

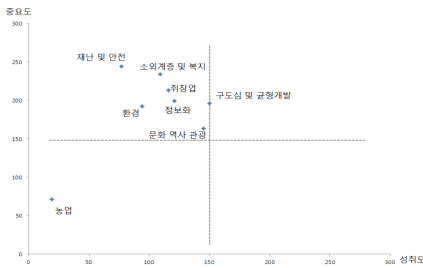


그림 4. 수원시 주요 IPA 분석 결과

IPA 분석 결과는 [그림 4]와 같으며, 여기에서 x축은 성취도를 y축은 중요도를 의미한다. 재난 및 안전, 소외계층 및 복지, 고용 및 취·창업, 첨단산업, 환경, 문화·역사·관광, 구도심 및 균형개발 등의 7개 이슈가 모두 2사분면에 위치하는 것으로 나타났다. 이는 이러한 정책 이슈들이 전문가들이 판단하기에 정책의 중요도가 높는데 비해 수원시의 대응 정도는 미흡한 것을 의미한다. 다만, 그 중 구도심 및 균형개발과 문화·역사·관광의 정책 이슈들은 상대적으로 높은 평가를 받았다.

그에 비해 수원시의 농업 관련 정책 이슈는 그 중요도와 대응 정도가 모두 낮은 평가를 받았다. 이는 농업 관련 정책 이슈는 정책적 중요성과 시급성이 다소 떨어지는 것을 의미한다.

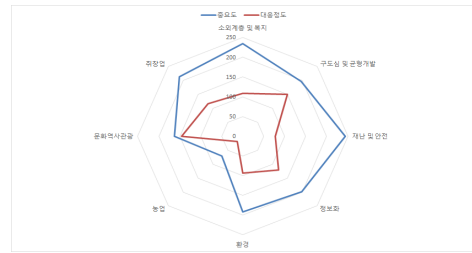


그림 5. 수원시 대내 이슈 중요도 대비 성취도 분석 결과

이를 방사형 그래프로 나타내면, [그림 5]와 같다. 정책 이슈의 중요도 대비 수원시의 대응 정도를 살펴보면, 재난 및 안전, 소외계층 및 복지는 수원시가 정책적으로 우선순위에 두고 노력할 필요가 있다. 반면, 문화·역사·관광은 수원시의 대응 정도가 중요도에 일정 수준 근접해 있는 것으로 나타나는 것을 알 수 있다.

## V. 결론

본 연구는 빅데이터의 활용을 통한 정책 발굴의 필요성을 인식하고, 빅데이터 중 신문 기사의 반정형 데이터를 통해 지방자치단체의 정책 이슈를 발굴하고자 하였다.

연구의 공간적 대상인 수원시를 중심으로 수원시 관련 주요 일간지 및 지역 신문의 지난 3년간의 기사 약 18만 건을 분석하였다. 분석 결과, 수원시의 주요 정책 이슈는 1.소외계층 및 복지, 2.구도심 및 도시균형개발, 3.재난·안전, 4.고용 및 취·창업, 5.첨단산업, 6.환경, 7.농업, 8.문화·역사·관광의 총 8가지 분야 총 110여개의 이슈로 나타났다.

정책 이슈의 중요도와 수원시가 우선하여 추진해야 하는 정책을 도출하기 위해 본 연구에서는 수원시 관련 석·박사급 연구원, 공무원, 시민단체 및 관련학과 교수 총 70명을 대상으로 전문가 설문조사를 실시하여, 이를 기반으로 IPA 분석을 실시하였다. IPA 분석 결과, 8개 정책 이슈 중 농업을 제외한 소외계층 및 복지, 구도심 및 균형개발, 재난·안전, 고용 및 취·창업, 첨단산업, 환경, 문화·역사·관광이 모두 정책적 중요성이 높은 것으로 나타났다. 이 중 수원시가 일정 수준으로 대응하고

있는 문화·역사·관광을 제외한 나머지 정책 이슈들의 우선순위가 높았으며, 그 중 특히 재난 및 안전과 소외 계층 및 복지의 우선순위가 높았다.

본 연구에서 분석한 신문 기사를 통한 빅데이터 분석은 해당 기간 동안 탄핵, 촛불 시위, 대통령 선거 등 전국 주요 신문의 대형 이슈에도 불구하고 수원에서 주요하게 다루지는 정책 이슈들을 도출하는데 효과적임을 보였다. 또한, 분석을 통해 도출한 8가지 정책 이슈들의 경우 전문가 설문조사를 통한 IPA분석 결과, 대부분이 수원시에 그 중요성이 높은 것으로 평가되었다.

본 연구는 지방자치단체의 정책 발굴의 효율성을 높이고 민의를 효과적으로 파악하기 위한 빅데이터 분석의 활용 방안을 제시하였다는 점에서 의미를 지닌다. 즉, 빅데이터를 활용한 정책 이슈의 발굴과 정책의 개발은 기존의 답다운 방식의 일방적인 정책 추진과 비교되며, 숨은 민의를 파악할 수 있으며 효과적이고 효율적이라는 장점을 지닌다. 특히, 빅데이터에 관한 실질적인 연구가 많지 않은 국내 현실에서 빅데이터를 활용하여 구체적인 지방자치단체의 정책 이슈를 발굴하는 방법론을 제시하였다는 점에서 의미를 지닌다. 그러나 본 연구는 수원시만을 대상으로 하고 있다는 지역적인 한계와 함께 신문기사만을 분석하고 있다는 점에서 한계를 지닌다. 향후에는 신문기사와 주민 민원, SNS 등 다양한 빅데이터의 분석을 통한 정책 발굴의 연구가 필요할 것으로 예상된다.

#### 참 고 문 헌

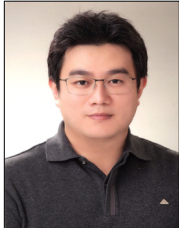
- [1] 성지은, 박기량, “빅데이터를 활용한 정책 사례 분석과 시사점,” 과학기술정책, 제24권, 제2호, pp.1-6, 2014.
- [2] 김재생, “빅데이터 분석 기술과 활용사례,” 한국콘텐츠학회지, 제12권, 제1호, pp.14-20, 2014.
- [3] 김정경, “국내·외 빅데이터 동향 및 성공사례,” ie 매거진, 제23권, 제1호, pp.47-52, 2016.
- [4] 유영성, 빈미영, 옥진아, 최준소, 천영석, *지자체의 공공 빅데이터 정책 사례연구*, 경기개발연구원, 2014.
- [5] 김동완, “빅데이터의 분야별 활용사례,” 경영논총, 제34권, pp.39-52, 2013.
- [6] 송진희, 김정숙, “빅데이터 서비스의 선진사례 분석,” 한국콘텐츠학회논문지, 제12권, 제1호, pp.32-37, 2014.
- [7] J. Manyika, M. Chui, and B. Brown, “Big Data: The Next Frontier for Innovation, Competition, and Productivity,” Mckinsey Global Institute, 2011.
- [8] J. Gantz and D. Reinsel, “The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East - United States,” IDC Country Brief, 2013.
- [9] 차승은, 주현택, “Big-Data 활용 동향 및 활성화에 대한 고찰,” ie 매거진, 제22권, 제1호, pp.41-45, 2015.
- [10] 김용진, *수원시정연구원 데이터베이스 구축 및 활용·관리 방안 연구*, 수원시정연구원, 2015.
- [11] 김배현, “해외 주요국가의 빅데이터 정책 비교 분석,” 한국콘텐츠학회논문지, 제12권, 제1호, pp.38-40, 2014.
- [12] 한영광 “지방자치단체 빅데이터 추진 현황과 정책적 시사점,” 한국지역정보학회 학술발표대회 논문집 2016, 제5호, pp.69-88, 2016.
- [13] 김신곤, 조재희, “지방자치단체의 빅데이터 도입을 위한 제언,” 한국지역정보학회지, 제16권, 제3호, pp.13-41, 2013.
- [14] 김병일, “빅데이터 분석과 데이터마이닝을 위한 저작권 제한,” 저작권, 제30권, 제1호, pp.29-61, 2017.
- [15] 최재원, 이호, 김정민, 송주호, “토크모델링을 활용한 소프트웨어 분야 대학 교과과정 분석,” 한국전자거래학회지, 제22권, 제4호, pp.193-214, 2017.
- [16] 서동민, 정한민, “빅데이터 분석 서비스 지원을 위한 지능형 웹크롤러,” 한국콘텐츠학회논문지 제13권, 제12호, pp.575-584, 2013.
- [17] 김도영, *4차 산업혁명시대 수원시 혁신방향과 과제 도출*, 수원시정연구원, 2017.

[18] 최승범, “지방자치단체 빅데이터의 활용트렌드와 활성화 조건에 관한 연구,” 사회경제와 정책연구, 제6권, 제1호, pp.177-205, 2016.

저 자 소 개

김 용 진(Yong-Jin Kim)

정회원



- 2005년 7월 : 서울대학교 건축학과 건축학 전공(공학사)
- 2012년 2월 : 서울대학교 건설환경공학부 도시계획/설계 전공(공학박사)
- 2015년 9월 ~ 현재 : 한국교통대학교 건설환경도시교통공학부 조교수

<관심분야> : 도시개발, 도시정보, 공간복지

김 도 영(Do-Young Kim)

정회원



- 2014년 2월 : 협성대학교 도시건축공학과 도시공학 전공(박사수료)
- 2013년 3월 ~ 현재 : 수원시정연구원 연구위원
- 2017년 9월 ~ 현재 : 경기블록체인연구소 연구위원

<관심분야> : 스마트 거버넌스, 생태도시, 도시재생