

일반논문 (Regular Paper)

방송공학회논문지 제23권 제4호, 2018년 7월 (JBE Vol. 23, No. 4, July 2018)

<https://doi.org/10.5909/JBE.2018.23.4.536>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

국내 공공데이터 관련 뉴스 의미망 분석을 통한 공공정책 연구

문혜정^{a*}, 이경서^{b)}

A Study on Public Policy through Semantic Network Analysis of Public Data related News in Korea

HyeJung Moon^{a*)} and Kyungseo Lee^{b)}

요 약

공공데이터는 정부 3.0을 기점으로 공급자 위주의 정보 개방에서 국민 개개인 중심의 맞춤형 정보 공유의 형태로 변모했다. 이에 따라 정부는 공공데이터의 가치를 극대화하고 재사용을 높이고자 관련 정책 및 사업을 펼치고 있다. 본 연구는 뉴스에 드러난 공공데이터 관련 이슈를 분석하여 이슈별 정부 기관 및 정부 사업 현황을 알아보고자 한다. 공공데이터를 포함하는 국내 온라인 뉴스와 공공기관 입찰정보를 대상으로 의미망 분석을 실시했으며 도출된 주요 주제어를 공공데이터에 내재된 사회적, 경제적 가치와 연결시키는 작업을 수행했다. 분석 결과 공공데이터 관련 주요 이슈는 공공데이터 접근성 확대, 신기술의 성장, 이해관계자들 간 협력과 갈등, 민간부문의 활용으로 나뉘었으며 이는 각각 투명성, 효율성, 참여성, 혁신 메커니즘과 밀접한 연관성을 보였다. 이슈별 주요 기관으로는 투명성이 기획재정부와 서울특별시, 효율성이 문화체육관광부와 경기도, 참여성이 산업통상자원부와 인천광역시, 혁신이 국토교통부와 경상북도가 해당되었으며 대부분의 이슈들이 정부주도로 형성되고 있었다.

Abstract

Public data has been transformed from provider-oriented information disclosure to a form of personalized information sharing centered on individual citizens since government 3.0. As a result, the government is implementing policies and projects to maximize the value of public data and increase reuse. This study analyzes the issues related to public data in the news and seeks the status of government agencies and government projects by issue. We conducted semantic analysis on domestic online news and public agency bidding information including public data and conducted the work of linking major key words derived with social and economic values inherent in public data. As a result, major issues related to public data were divided into broader access to public data, growth of new technology, cooperation and conflict among stakeholders, and utilization of the private sector, which were closely related to transparency, efficiency, participation, and innovation mechanisms. Also major agencies of four issues include the Ministry of Strategy and Finance and Seoul, Ministry of Culture, Sports and Tourism and Gyeonggi-do, Ministry of Trade, Industry and Energy and Incheon, and Ministry of Land, Infrastructure and Transport and Gyeongsangbuk-do. Most of the issues are being led by the government.

Keyword : Public Data, News Analysis, Bid Announcement, Big Data, Public Policy

I. 서론

정부 3.0의 도래로 정부는 공공정보를 적극 개방하고 일방적인 정보 공개에서 국민 개개인 맞춤형 서비스를 제공하기 위해 노력해 왔다¹⁾²⁾³⁾⁴⁾. 특히 공공기관이 보유한 데이터 개방에 관한 사항을 규정한 「공공데이터의 제공 및 이용 활성화에 관한 법률(공공데이터법)」의 시행은 공공데이터의 활용을 대폭 증가시켰다. 법 시행 이후 오픈API 신청 건수 및 데이터셋(dataset) 다운로드 수가 급격히 증가했으며 공공데이터를 활용해 새로운 서비스를 개발한 사례도 월평균 42개에서 244개로 약 5.8배 증가하였다¹⁾. 공공데이터 개방 추진체계는 공공데이터전략위원회의 발족이 토대가 되었다. 공공데이터전략위원회를 통해 공공데이터 관련 정책 시행 및 평가 체계가 마련되었으며 이를 중심으로 한 국정정보화진흥원, 행정안전부, 공공데이터 제공 분쟁조정위원회가 각각의 역할을 수행하고 있다. 최근에 발표한 정부혁신 종합 추진계획에 따르면 정부는 국민 중심의 공공데이터 개방 확대를 목표로 기존의 선택적 개방에서 전수조사를 통한 90%이상의 개방을 달성하겠다고 밝혔으며 경제적 가치뿐만 아니라 사회적 가치를 구현하기 위해 노력할 것을 약속했다²⁾. 이처럼 공공데이터는 특정 기관에만 한정된 것이 아니라 전 기관 차원에서 진행되는 사업이며 단기적인 프로젝트가 아닌 장기적인 정부 사업 중 하나로 인식되고 있다.

국가는 계속해서 축적되는 데이터를 제대로 관리하기 위

해 정부 안팎 상황을 살펴봄으로써 공공데이터 관련 정책의 실태를 파악하고 개선하기 위한 꾸준한 노력을 지속해야 할 필요가 있다. 본 연구는 언론에 드러난 국내 공공데이터 관련 주요 이슈가 무엇인지 알아보고 뉴스라는 대중매체를 통해 시기별 정책 및 정부 사업을 이해하고 분석하고자 했다³⁾. 또한 정부 기관별 입찰 공고를 통해 실질적인 정부 사업의 방향을 파악하고 국내 공공데이터 사업의 특징과 추후 주목해야 할 공공데이터 관련 이슈까지도 예측해 보았다.

II. 배경

1. 공공데이터 개념

국내 공공데이터 개념은 공공기관에 의해 생성되는 공공정보를 의미하는 공공부문 데이터(public data), 자유로운 사용과 수정이 가능한 자료를 의미하는 오픈 데이터(open data), 공개된 정부 자료를 말하는 오픈 정부 데이터(open government data), 공공재로서의 성격을 띠고 있는 시민데이터(citizen data) 등 다양한 의미가 혼동되어 사용되고 있다⁴⁾. 공공데이터법에서 명시하고 있는 공공데이터는 데이터베이스, 전자화된 파일 등 공공기관이 생성 또는 취득하여 관리하고 있는 광(光) 또는 전자적 방식으로 처리된 자료 또는 정보를 말한다. 우리나라에서 가장 보편적으로 인식되는 공공데이터는 공공데이터 포털을 통해 공개되는 공공 자료이다⁵⁾. 본 논문에서는 공공데이터법에서 정의하고 있는 개념에 따라 공개되지 않는 자료도 공공데이터로 포함시켰으며 영문 표기는 'public data'를 사용하였다.

2. 선행문헌검토

공공데이터를 분석한 연구로 각 지자체의 공공데이터 포털사이트의 현황을 조사한 연구가 있다⁶⁾. 해당 연구는 자

a) 서울과학기술대학교, 아이엘피(Seoul University of Technology & Science, Institute of Lottery Policy)

b) 한국외국어대학교 미디어커뮤니케이션학부(Hankuk University of Foreign Studies, Division of Media & Communication)

* Corresponding Author : 문혜정(HyeJung Moon)

E-mail: hyejung.moon@gmail.com

Tel: +82-2-738-3486

ORCID:https://orcid.org/0000-0001-8265-3256

※ 이 논문은 2015년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-과제번호)(NRF-2015S1A5B4A01036743).

Manuscript received May 16, 2018; Revised June 14, 2018; Accepted June 14, 2018.

1) 아시아뉴스통신(2012.07.31), 김을동의원, 공공데이터 제공 및 이용 활성화에 관한 법률 국회 발의.

2) 전자신문(2012.08.19), 말뿐인 공공정보 개방...사용자 눈높이 맞춰라.

3) 아이뉴스24(2013.03.14), '정부3.0' 시대 국가 신성장 동력은 '데이터'.

4) 아시아투데이(2018.03.20), 행안부, 정부혁신 BI '보다 나은 정부' 공개.

체 공공데이터 포털사이트를 보유한 지자체를 선정하여 관련 업무를 담당하는 실무자와의 상담 및 설문조사를 바탕으로 공공데이터 개방 현황을 평가하였다. 또한 여섯 개의 포털사이트 평가 기준을 수립하여 객관적인 척도로서 사용하였다. 하지만 공공데이터 포털사이트를 독립적으로 운영하고 있는 지자체가 여섯 군데에 그쳐 전반적인 공공데이터 현황을 파악하기에는 한계가 있다. 공공데이터의 가치를 분석한 연구로는 데이터 품질 평가 모델을 바탕으로 공공데이터 평가를 실시하고 데이터 개방정책이 창출하는 공공 가치를 분석하여 공공 가치 창출 메커니즘을 설계한 연구가 있다^[7]. 공공데이터 개방 정책을 공공 가치라는 관점에서 분석하였다는 점, 공공기관이 보유한 미공개 데이터까지 전수조사를 시행하였다는 점에서 의의를 가진다. 다만 해당 연구의 분석 대상은 공개된 데이터들의 속성인 데에 비해 본 연구는 데이터 관련 이슈들을 중점적으로 다룬다는 부분에서 차이를 보인다. 수요자 중심의 공공데이터 활용 방안을 조사한 연구로 수요자를 대상으로 설문 조사를 실시한 연구가 있다^[8]. 해당 연구는 공급자가 아닌 수요자 측면에서 공공데이터에 대한 실증적인 분석이 선행되어야 함을 강조하며 실제 이해관계자들인 민간부문의 실무자를 대상으로 공공데이터에 대한 기본 인식, 상업적 수요, 활용 장애 요인 및 활성화 요인에 대해 조사하였다. 하지만 조사 대상이 IT기업에 종사 중인 62명의 실무자로 한정되어 공공데이터의 다양한 가치 중 기업의 상업적 가치에 그쳤다는 점은 한계로 남는다. 또한 빅데이터(big data) 분석 기법을 통해 특정 지방자치단체의 공공데이터를 연구한 사례가 있다^[9]. 해당 논문은 공공데이터 포털을 분석하여 충북 지역의 공공데이터 개방 추진방안에 대하여 연구하였다. 공공데이터 포털의 데이터파일 중 충청북도과 관련된 데이터의 제목을 추출하고 빈도 분석을 실시하여 충북지역에 대한 공공데이터 주요 내용을 파악하였다. 공공데이터 간 연결성 분석을 통해 분야별 개방건수가 많은 단어들의 연결 관계를 보여주고 있는 점에서 의의를 갖지만 특정 지역에 한정된 데이터만을 분석했다는 한계가 있다.

본 논문은 뉴스 분석을 통해 공공데이터 가치를 연구한 논문을 발전시킨 것이다^[10]. 해당 논문은 뉴스 의미망 분석을 통해 공공데이터 관련 주요 가치를 도출하고 이를 시계열 분석하여 미래 주요 가치에 적합한 공공데이터 포털 개

방 정책에 대해 제언하였다. 하지만 공공데이터포털에 개방되어 있는 데이터가 정부의 개방정책을 대변한다고 판단하기 어렵다는 한계가 있었다. 따라서 본 연구에서는 정부 입찰공고 분석을 통해 공공데이터 관련 정부의 사업 방향을 파악하여 기존 연구를 보완하고자 하였다. 선행 연구들과 비교하여 본 연구가 가지는 차별성은 다음과 같다. 첫째, 온라인 뉴스 분석을 통해 공공데이터 관련 주요 이슈를 파악할 것이다. 이는 실용적 측면에서의 공공데이터 관련 주제를 도출하고 주제별 특성을 비교하는 데 도움이 될 것이다. 둘째, 정부 주요 기관별로 온라인 뉴스에 드러난 공공데이터 이슈가 무엇인지 구분하고 입찰 정보 분석을 통해 구체적으로 투자하고 있는 정부 사업은 무엇인지 살펴볼 것이다.

III. 연구 설계

1. 연구 질문 및 분석틀

연구 목적은 공공데이터로 인해 창출되는 이슈를 새롭게 분류하고 각 이슈에 해당하는 정부 사업을 파악하는 것이다. 먼저, 국내 공공데이터 이슈를 정의하기 위해 온라인 뉴스 의미망 분석을 실시하여 서로 다른 성격을 가진 단어들을 구분하고 새로운 주제어를 선정하였다. 이후 각 이슈에 대한 관심도가 높은 기관을 선별하여 해당 기관에서 게재한 데이터 관련 입찰공고와 공고별 배정예산을 살펴보았다. 연구 질문은 다음과 같다.

■ 국내 공공데이터 이슈는 어떻게 분류되는가?

■ 이슈별 주요 정부 기관과 관련 사업은 무엇인가?

공공데이터 가치 창출 프레임워크에서 정의한 공공데이터 가치 분류 체계를 토대로 공공데이터 이슈를 분류하였다^[11]. 가로의 x축 좌표는 이슈 창출에 대한 외부 이해관계자의 기여 정도에 따라 정부주도와 민간주도로 나뉜다. 기여 정도가 낮을수록 데이터 개방과 관리 측면이 강조되고 기여 정도가 높을수록 데이터 활용 측면이 강조된다. 세로의 y축은 사회적 이슈 창출 정도를 의미한다. y축이 낮을 경우 사회적 이슈 대신 경제적 이슈에 보다 집중하고 있다.

표 1. 공공데이터 이슈 분석틀

Table 1. Analysis Framework of Public Data related Issue

High	Transparency based issue 1. Main issue 2. Main central administrative agency 3. Main local government 4. Main government business	Participation based issue 1. Main issue 2. Main central administrative agency 3. Main local government 4. Main government business
	Efficiency based issue 1. Main issue 2. Main central administrative agency 3. Main local government 4. Main government business	Innovation based issue 1. Main issue 2. Main central administrative agency 3. Main local government 4. Main government business
Low	Focusing external stakeholder's role in issue generation	
	Low	High

출처 : Jetzek, et al(2014) 재구성

Source : Reconstruction of Jetzek, et al(2014)

2. 분석방법

의미망 분석기법을 이용하여 공공데이터와 관련된 언론과 입찰공고의 주요 주제어를 도출하고자 한다. 뉴스 의미망 분석은 특정 이슈에 대한 대중 매체의 프레이밍(framing)을 이해하는 데 자주 사용된다^{[12][13]}. 의미망 분석은 정형 데이터와 비정형 데이터에서 중요 단어를 추출하고 구성요소들 간의 관계를 파악하여 의미를 도출해내는 연구방법을 말한다^[14]. 이러한 의미연결망 분석은 개별 노드(node)가 어떠한 연결 관계를 갖고 있는지 보여줌으로써 특정 이슈에 대해 어떤 연결 구조를 가지고 있는지 가시적으로 보여준다. 의미망 내에서 노드는 텍스트에 내포되어 있는 단어를 말하며, 이러한 단어를 연결시켜주는 관계를 엣지(edge)라고 한다^[15]. 따라서 노드와 엣지를 통해 분석하고자 하는 온라인 뉴스와 입찰공고에 내포된 주요 단어와 단어들 간의 연결 관계를 파악할 수 있다. 본 연구는 공공데이터 가치별 주요 노드를 비롯하여 각 노드에 대한 연결선 수(degree), 중심성(centrality), 동시 출현성(co-occurrence) 등을 분석함으로써 공공데이터 가치별 구체적인 특성을 분류하였다. 의미망 분석은 다음과 같이 진행되었다. 온라인 뉴스로부터 추출한 단어들을 대상으로 ‘애플리케이션’, ‘앱’, ‘Mobile App’처럼 동일한 의미를 가진 단어들을 하나의 단어로 통일시켰고 ‘공공 데이터’, ‘4차 산업 혁명’처럼 두 개

이상의 단어들로 이루어져있지만 특정 의미를 가진 한 단어로 사용되는 단어들은 복합어로 설정하였다. 전처리 과정이 완료된 뉴스 데이터는 검색 단어인 ‘공공데이터’와 ‘데이터’를 제외한 모든 명사 단어를 대상으로 동시 출현성 네트워크 분석(co-occurrence network)을 실시하였다. 해당 분석을 통해 주요 주제어를 도출하였으며 주제어별로 형성된 네트워크를 바탕으로 연관성이 높은 단어들끼리 도메인(domain)을 형성하였다. 마지막으로 각 도메인을 대표하는 단어를 테마(theme)로 선정하고 테마별로 같은 성격을 가진 단어들을 분석하였다. 분석 도구로는 의미망 분석 도구인 KH-Coder를 사용하였다.

3. 자료수집

뉴스는 특정 시대에 대중으로부터 주목받은 이슈를 이해할 수 있는 대표적인 매체이다^{[16][17]}. 본 연구는 공공데이터 관련 이슈와 정부 사업을 살펴보기 위해 국내 언론 이슈를 담당하는 온라인 뉴스를 선정하였고 정부 사업을 조사하기 위해 공공기관들의 사업 수요를 대표하는 입찰공고를 분석 대상으로 하였다. 빅카인즈(BigKinds)⁵⁾를 활용하여 모든 온라인 뉴스를 대상으로 ‘공공데이터’가 언급된 기사의 제목, 게재시기, 본문을 검색 단위로 하여 수집하였다^[18]. 전체 기간을 기준으로 했을 때 2008년부터 온라인 기사가 검

5) 빅카인즈는 1990년부터 현재까지 46개 매체의 약 4천만 건의 뉴스 기사를 저장하고 데이터 재가공이 가능하도록 정리하여 제공하는 사이트이다.

1. 공공데이터 이슈 분류

〈그림 1〉에서 같은 색상으로 표현된 노드들은 서로 밀접하게 연관된 네트워크이다^[20]. 노드의 분포를 보면 전체 노드가 크게 네 가지 색상으로 구분되어 있음을 확인할 수 있다. 노드 색상별 주요 주제어와 빈도수는 ‘혁신’(656회), ‘경제’(631회), ‘소통’(518회), ‘공유’(469회), ‘사회’(450회)가 중심이 되는 D1, ‘사업’(1,484회), ‘창업’(1,272회), ‘아이디어’(682회), ‘공모전’(664회), ‘대회’(858회)로 구성된 D2, ‘개방’(4,171회), ‘활용’(3,544회), ‘정부3.0’(1,992회), ‘정부’(2,170회), ‘행정’(1,822회)으로 이루어진 D3, ‘빅데이터’(1,542회), ‘분석’(657회), ‘비즈니스’(825회), ‘기술’(792회), ‘시스템’(615회)이 포함된 D4가 있다.

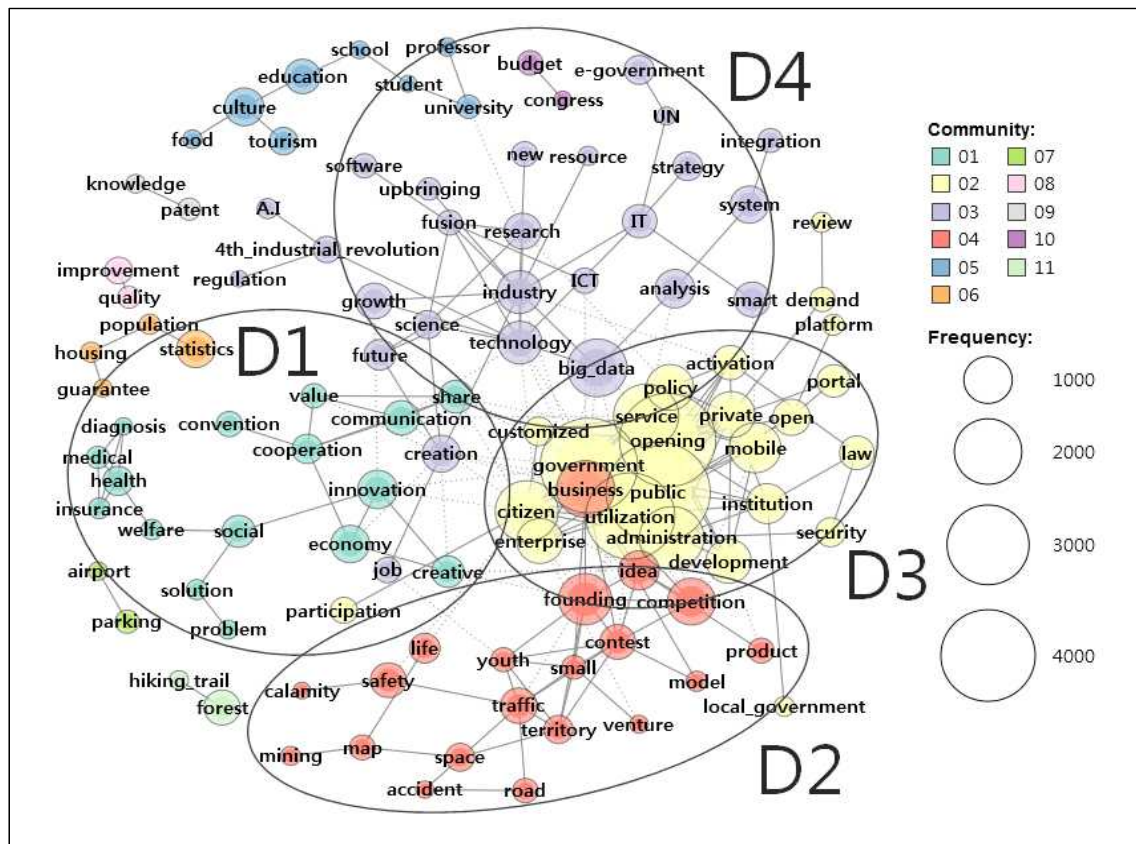


그림 1. 온라인 뉴스 동시 출현성 네트워크 분석
Fig 1. Co-Occurrence Network Analysis of Online News

위한 기관의 혁신과 산학연 협력 강화를 주제로 가지고 있었다. 주요 기사로는 ‘당진시, 열린 혁신으로 시민참여 강화6)’, ‘정부혁신, 시민 참여가 핵심이다7)’, ‘대전교육청, 정보 사전공개·시민참여 등 확대한다8)’, ‘공공데이터 품질 높여 민간 협력 강화9)’, ‘현대차-통계청, 위치기반서비스 협력 MOU 체결10)’ 등이 있었고 대부분이 정부 주도의 활동이었다. 최근에는 ‘보험사에 진료정보 넘긴 심평원11)’과 같이 소통의 부재로 인해 이해관계자 간 갈등이 드러난 기사도 등장하기 시작하였다. D2의 경우 공공데이터 활용 아이디어 공모전에 대한 기사가 가장 많았으며 각종 창업 지원 사업들이 포함되었다. 해당 기사로는 ‘신한은행, 빅데이터 센터 설립 기념 아이디어 공모전 개최12)’, ‘산림청, 산림 공공데이터서 창의적 아이디어 발굴13)’, ‘산업부, 제2회 공공데이터 활용 비즈니스 아이디어 공모전 개최14)’, ‘공공데이터로 창업, ‘부산 오픈스케어-D’ 문 열어15)’, ‘공공데이터 창업지원 1264억 투입 관련부처들 협업해 시너지도 높여16)’ 등이 있었다. 대다수의 공모전이 정부 기관 주도로 시행되었으나 활용 분야가 공공부문에서 민간부문까지 다양했고, 특히 비즈니스 모델 발굴을 통한 경제적 가치 창출에 주목하고 있었다. D3은 전체 기사 중 가장 높은 비중을 차지하고 있는 만큼 다른 도메인과 중복되는 단어들도 많았다. 따라서 빈도수는 낮지만 도메인의 특성을 잘 나타내는 단어인 ‘포털’, ‘활성화’, ‘애플리케이션’, ‘정책’을 중심

으로 뉴스를 살펴보았다. ‘공공데이터 포털로 유익한 공공 정보 얻는다17)’, ‘공공정보 개방 활성화 법제 개선부터 나서야18)’, ‘우수 공공앱 모바일 전자정부 이끈다19)’, ‘공공 정책정보 3분의 2 원문 그대로 공개20)’ 등이 이에 해당하였으며 공통적으로 데이터에 대한 접근성을 강화하고 데이터 활용을 높이는 사업 및 정책에 대한 내용을 담고 있었다. 반면 D4는 ‘공공데이터’와 더불어 ‘빅데이터’가 중심이 되는 기사가 주를 이루었다. 그 예로 ‘서울시-신한카드, 공공서비스 향상에 빅데이터 기술 접목21)’, ‘안행부, 올해 빅데이터 사업에 460억원 투입22)’, ‘공공 빅데이터 활용 지원 위한 ‘법정부 콘트롤타워’ 출범23)’과 ‘공공데이터로 융합 신산업 창출...행자부, 2차 공공데이터 기본계획 발표24)’, ‘기존 주력산업 고도화 4차 미래산업 육성만이 살길25)’ 등이 해당되었다.

온라인 뉴스 동시 출현성 네트워크 분석과 내용 분석을 통해 각 도메인들이 나타내는 이슈가 공공데이터 창출 가치와 매우 유사하다는 것을 발견할 수 있었다. Jetzek et al(2014)에 따르면 (1)투명성은 정부 데이터 개방과 정보의 비대칭성의 감소를 통해 가치를 창출하는 메커니즘, (2)효율성은 처리 비용 절감, 공공기관 간 연계 등의 과정을 통해 공공 자원의 활용도를 향상시키는 메커니즘, (3)참여성은 아이디어 및 자원 공유를 통해 시너지 효과를 창출하는 메커니즘, (4)혁신은 새로운 제품과 서비스, 자원, 생산 방법

6) 금강일보(2017.11.20), 당진시, 열린 혁신으로 시민참여 강화.

7) 디지털타임스(2017.03.30), 정부혁신, 시민 참여가 핵심이다, 22면.

8) 뉴스시(2013.08.01), 대전교육청, 정보 사전공개·시민참여 등 확대한다.

9) 서울경제(2016.05.25), 공공데이터 품질 높여 민간 협력 강화, 15면.

10) 이투데이(2013.11.19), 현대차-통계청, 위치기반서비스 협력 MOU 체결.

11) 메디컬투데이(2017.12.18), 보험사에 진료정보 넘긴 심평원.

12) 전자신문(2016.08.04), 신한은행, 빅데이터센터 설립 기념 아이디어 공모전 개최.

13) 디지털타임스(2017.03.16), 산림청, “산림 공공데이터서 창의적 아이디어 발굴”.

14) 전기신문(2014.10.17), 산업부, 제2회 공공데이터 활용 비즈니스 아이디어 공모전 개최.

15) MBN(2017.04.19), 공공데이터로 창업, ‘부산 오픈스케어-D’ 문 열어.

16) 파이낸셜뉴스(2015.04.02), 공공데이터 창업지원 1264억 투입 관련부처들 협업해 시너지도 높여, 18면.

17) 디지털타임스(2014.01.08), 공공데이터 포털로 유익한 공공정보 얻는다.

18) 디지털타임스(2013.12.23), 공공정보 개방 활성화 법제 개선부터 나서야, 9면.

19) 디지털타임스(2014.12.01), “우수 공공앱 모바일 전자정부 이끈다”, 9면.

20) 연합뉴스(2013.09.17), 공공정책정보 3분의 2 원문 그대로 공개.

21) 디지털타임스(2015.04.23), 서울시-신한카드, 공공서비스 향상에 빅데이터 기술 접목.

22) 전자신문(2014.01.07), 안행부, 올해 빅데이터 사업에 460억원 투입, 6면.

23) 디지털타임스(2016.12.08), 공공 빅데이터 활용 지원 위한 ‘법정부 콘트롤타워’ 출범.

24) 전자신문(2016.12.20), 공공데이터로 융합 신산업 창출...행자부, 2차 공공데이터 기본계획 발표, 10면.

25) 경향일보(2017.05.14), “기존 주력산업 고도화 4차 미래산업 육성만이 살길”.

의 조합을 통해 시장과 산업의 변화를 가져다주는 메커니즘을 말한다²²⁾. 해당 정의를 참고하여 온라인 뉴스에 등장한 단어들 중 공공데이터 가치를 설명하기에 적합한 단어들로 새로운 도메인을 생성하였다. 도메인별 구성 단어는 다음과 같다. 투명성은 ‘정부3.0’, ‘공공기관’, ‘공공데이터 포털’, ‘공공데이터 개방’, 효율성은 ‘전자정부’, ‘디지털’, ‘IT(information technology)’, ‘통합 시스템’, 참여성은 ‘시민 소통(시민들 간의 소통)’, ‘공유’, ‘협력’, ‘문제 해결’, 혁신은 ‘창업’, ‘비즈니스’, ‘제품’, ‘아이디어’이다.

표 2. 공공데이터 관련 이슈 및 가치 유사도 매트릭스
Table 2. Similarity Matrix of Public Data related Issue and Value

	*Transparency	*Efficiency	*Participation	*Innovation
*D1	0.154	0.086	0.316	0.110
*D2	0.132	0.079	0.064	0.779
*D3	0.461	0.116	0.100	0.216
*D4	0.135	0.257	0.073	0.175

온라인 뉴스를 분류한 D1, D2, D3, D4 도메인과 앞서 도출한 공공데이터 가치 도메인과의 연관성을 수치적으로 나타낸 결과가 <표 2>다. 뉴스 양이 많지 않은 것을 고려하여 자카드 계수(jaccard coefficient)²⁶⁾를 이용하여 빈도수와 관계없이 단어 간 동시 발생 여부를 계산하였다. 온라인 뉴스에서 추출한 이슈 도메인과 공공데이터 가치 도메인 간의 연관성을 측정된 결과 D1, D2, D3이 각각 참여성(0.316), 혁신(0.779), 투명성(0.461)과 서로 밀접하게 연관되어 있었다. 반면 D4는 효율성(0.257)과 비교적 낮은 연관성을 보였다. 결과적으로 D1, D2, D3는 참여성, 혁신, 투명성 중심 이슈라고 정의할 수 있으나 D4는 새롭게 정의해줄 필요가 있었다. D4의 중심 단어인 '데이터', '기술', '기업', '과학', '산업', '민간' 등을 고려하여 본래의 의미인 정부 기관 간 연계를 통한 효율성 창출에서 신기술 성장에 의한 효율성 증대로 그 의미를 변경하였다. 해당 정의는 혁신과 어느 정도 유사할 수 있으나 새로운 제품이나 서비스가 아니라 오로지 기술적인 측면을 강조한다는 점에서 다르다.

해당 결과를 토대로 국내 공공데이터 이슈 도메인을 시민 참여를 통한 공공부문의 협력을 뜻하는 참여성 중심 이슈, 경제 혁신을 통한 민간부문의 활용을 뜻하는 혁신 중심 이슈, 공공데이터 개방 및 활성화 정책을 의미하는 투명성 중심 이슈, 신기술 및 신산업의 성장을 의미하는 효율성 중심 이슈로 정의하였다.

2. 이슈별 주요 기관 분석

앞서 분류한 투명성, 효율성, 참여성, 혁신 중심 공공데이터 이슈에 해당하는 주요 기관을 파악하기 위해 온라인 뉴스에 등장하는 대명사를 중점적으로 살펴본 결과 정부 부처와 지역명이 빈번하게 등장하였다. 따라서 15개의 중앙행정기관과 17개의 지방자치단체를 분석 대상으로 선정하였고 이음동의어 처리 및 산하 기관 통합 등의 전처리 과정을 수행하였다. 또한 기관별로 시행하고 있는 정부 사업을 면밀히 살펴보기 위해 데이터 관련 입찰공고 분석을 추가 시행하였다. 입찰공고의 경우 같은 공고가 중복 분석되는 것을 방지하기 위해 재공고와 큰 변경사항이 없는 수정공고는 배제하였으며 매년 같은 내용의 입찰을 주기적으로 진행하는 공고는 의미가 있다고 판단하여 게재 횟수만큼 포함시켰다.

2012년부터 2017년까지 6년간 중앙행정기관이 뉴스에 언급된 횟수는 총 482회로 기관별로는 고용노동부, 중소벤처기업부, 교육부, 환경부, 외교부, 통일부, 보건복지부가 15회 미만으로 매우 적어 공공데이터 이슈 비중을 평가하기 적합하지 않았고 나머지 9개 기관은 15회와 105회 사이를 기록했다. 해당 기관들을 대상으로 투명성, 효율성, 참여성, 혁신 중심 이슈 비중을 비교한 결과 전반적으로 투명성과 혁신의 비율이 높았다. 투명성 비율이 가장 높은 기관은 기획재정부였고 효율성이 가장 높은 비율을 차지한 기관은 문화체육관광부였고 참여성은 산업통상자원부와 고용노동부, 혁신은 국토교통부, 국방부, 과학기술정보통신부에서 높은 비율을 차지했다. 기관별로 게재한 입찰공고 수와 공

26) 자카드 계수는 두 집합의 유사도를 측정하는 척도로서 특정 단어와 노드 단어 간의 결합을 나타낸다. 단어 w에 대해서 a가 노드 단어 전후의 w의 빈도이고 F1은 전체 데이터에서 노드 단어의 빈도, F2는 전체 데이터에서 특정 단어의 빈도일 때 자카드 계수를 계산하면 다음과 같다²³⁾.

$$\text{Jaccard} = \frac{a}{F1 + F2 - a}$$

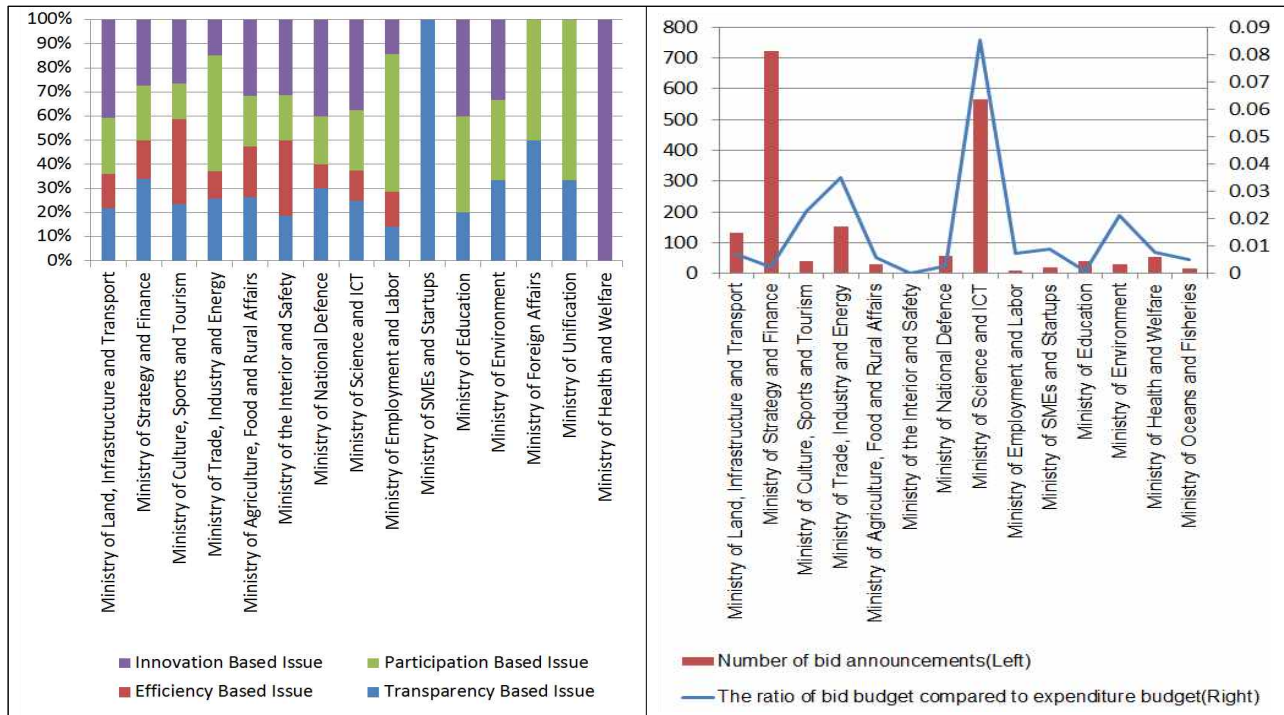


그림 3. 중앙행정기관별 공공데이터 이슈 비중 및 데이터 관련 입찰 정보

Fig 3. The Proportion of Public Data related Issue and Data related Bidding Information by Central Administrative Agency

고 배정예산 합계를 살펴보면 기획재정부와 과학기술정보통신부의 수치가 월등히 높았고 국토교통부, 산업통상자원부 또한 비교적 많은 수의 공고를 게재하고 있었다. 세출예산 대비 데이터 관련 사업에 많은 예산을 투입한 기관은 과학기술정보통신부였던 반면 기획재정부는 대부분의 공고에 배정 예산을 명시하지 않아 실제 예산 금액을 파악하기 어려웠다. 입찰 정부 내용을 분석해보면 주로 데이터 수집, 관리, 분석 기술에 관한 것으로 대다수가 효율성 중심 이슈를 포함하고 있었다. 그 중 투명성의 비중이 높은 기획재정부의 주요 국내 사업으로는 ‘온라인 메타데이터 DB 구축 시스템 확장개선 사업’, ‘교통정책과 투자계획 수립을 위한 데이터 구축’, ‘중요기록물 데이터베이스 구축 사업 용역’, ‘신용리스크 데이터 관리체계 고도화’, ‘소비자물가지수 생산을 위한 스캐너 데이터 활용방안 연구용역 사업’ 등이 있었다. 문화체육관광부의 효율성 중심 사업으로는 ‘빅데이터 과제 발굴 컨설팅 용역’, ‘빅데이터 분석을 통한 대체자료 서비스 전략 수립 연구용역’, ‘뉴미디어 기반 프로스포츠분야 소셜빅데이터 분석’ 등이 있었다. 전 기관을 대

상으로 참여성 이슈가 드러나는 공고를 분석한 결과 ‘맞춤형 문화데이터 개방 및 국민 공유 활성화 사업’, ‘정부 정책소통을 위한 빅데이터 분석’, ‘계산과학분야 확대 적용을 위한 데이터 공유형 연구협업 응용서비스 구현’ 등 소수의 입찰 정보를 발견할 수 있었다. 혁신이 높은 비중을 차지한 과학기술정보통신부의 경우 대부분의 입찰 공고가 한국정보화진흥원에서 게재한 연구 사업 및 시범사업에 관한 것이었다. 그 예로 ‘빅데이터 비즈니스 생태계 활성화 전략수립’, ‘한국의 데이터 활용기업 발굴 및 글로벌 홍보 연구’, ‘주요국 데이터 주도 비즈니스 현황 및 비즈니스 모델 비교 분석’ 등이 있었다. 지방자치단체가 온라인 뉴스에 언급된 횟수는 3,188회로 중앙행정기관보다 월등히 많았지만 단순히 위치나 장소로서 사용된 경우 역시 많았다. 하지만 이 경우에도 이슈 창출과 어느 정도 연관성을 가지고 있다고 판단하여 분석에 포함시켰다. 이슈별로 살펴보면 거의 모든 지역이 네 개의 이슈가 고루 분배되어 있었으나 그 중 참여성의 비중이 조금씩 높았다. 투명성이 중심이 되는 지역은 서울특별시였으며 효율성은 경기도, 참여성이 두드러

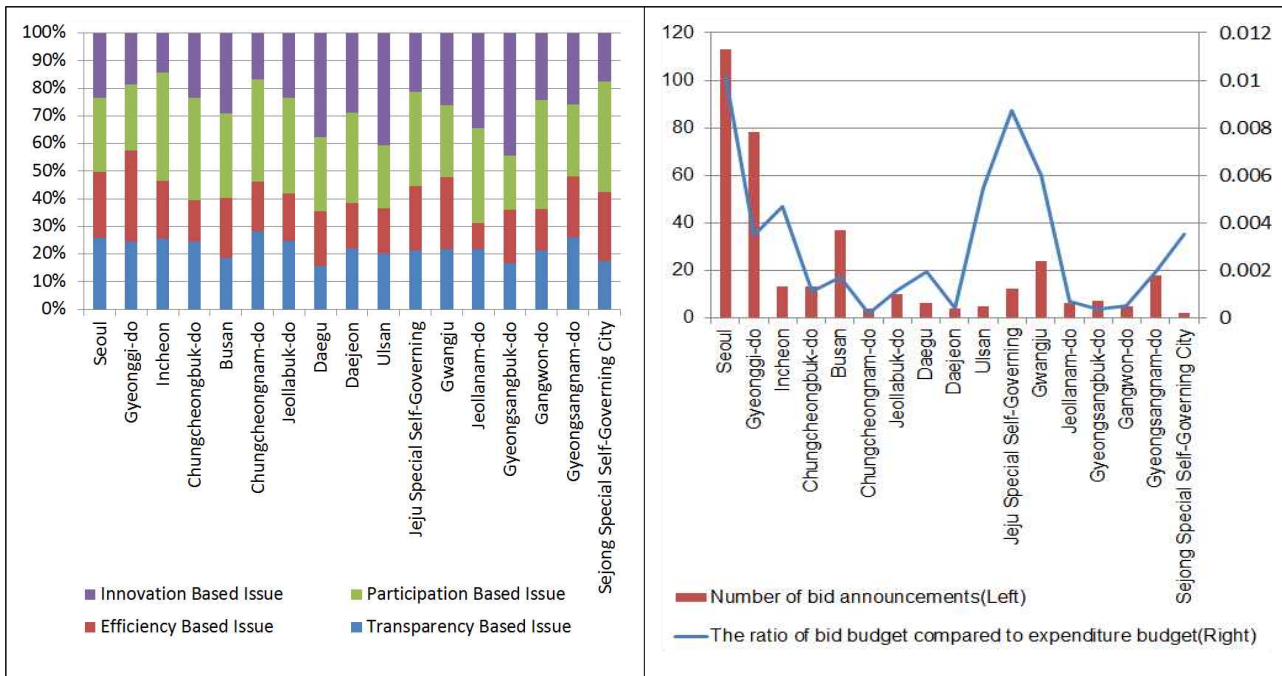


그림 4. 지방자치단체별 공공데이터 이슈 비중 및 데이터 관련 입찰 정보
Fig 4. The Proportion of Public Data related Issue and Data related Bidding Information by Local Government

지는 지역은 인천광역시, 충청북도, 세종특별자치시, 혁신은 경상북도, 울산광역시, 대구광역시가 높은 비중을 차지했다. 입찰 정보의 경우 서울특별시와 경기도가 가장 많은 공고를 게재했으며 부산광역시, 광주광역시, 경상남도 또한 높은 수치를 기록했다. 많은 예산을 투입한 지역은 서울특별시, 경기도, 인천광역시 순이었으나 공고별 배정예산이 가장 높은 지역은 인천광역시였고 세출예산 대비 총 배정예산이 높은 지역은 제주도였다. 서울특별시의 입찰공고로는 ‘행정데이터 통합 및 열린데이터광장 확장사업 용역’, ‘중요기록물 전산화 데이터 등록 및 전산화장비 유지보수 용역’, ‘통계자료 및 데이터 통합을 위한 공공데이터 플랫폼 구축 용역’ 등이 있었고 효율성이 높은 경기도는 ‘부천시교통정보센터 데이터베이스관리시스템 유지보수 용역업체 선정’, ‘2015년 빅데이터 분석컨설팅 용역’, ‘빅데이터 플랫폼 활용시스템 구축 용역’, ‘광명시 빅데이터 분석과제 전략 계획 수립 및 과제분석 용역’ 등과 같이 데이터베이스 구축과 빅데이터 분석에 관한 공고가 주를 이루었다. 참여성이 높은 공고인 ‘시민맞춤형 정책 빅데이터 구축’, ‘시민

과 소통강화를 위한 웹 소셜 빅데이터분석 용역’, ‘세종시민소리(VOC) 기반 민원 빅데이터 분석서비스’와 혁신이슈를 포함한 공고인 ‘2015년 빅데이터 정보 아이디어 및 디자인 공모전 행사 및 운영대행’, ‘빅데이터 활용 우수 비즈니스 모델 사업화 지원 용역’ 모두 빅데이터가 중심이 되는 효율성과 밀접하게 연결되어 있었다.

V. 결론 및 제언

정보(nodality)는 정부의 통치자원 중 하나로서 정책을 수행하는 데 있어 주요한 역할을 수행한다^{[24][25]}. 디지털 시대 정책도구의 변화로 정부의 정보력은 정책을 수행하고 평가하는 지표가 되고 있으며 국내 공공데이터 이슈의 대두는 이러한 흐름을 잘 나타내고 있다. <표 3>은 공공데이터 이슈를 가치 중심으로 분류하여 새롭게 정의한 것이다. 그 결과 (1)투명성 중심 이슈는 공공데이터 접근성 확대를 통한 정부3.0의 구현과 공공데이터 포털 구축 및 데이터

표 3. 공공데이터 이슈 매트릭스

Table 3. Public Data related Issue Matrix

High	Transparency based issue 1. Data accessibility expansion and activation policy 2. Ministry of Strategy and Finance 3. Seoul 4. Data management and integration platform construction	Participation based issue 1. Citizen participation, cooperation and conflict between stakeholders 2. Ministry of Trade, Industry and Energy, Ministry of Employment and Labor, 3. Incheon, Chungcheongbuk-do, Sejong 4. Expansion of citizen's customized data opening
	Efficiency based issue 1. Data related new technology and new industry growth 2. Ministry of Culture, Sports and Tourism 3. Gyeonggi-do 4. Big data analysis and system enhancement	Innovation based issue 1. Data utilization in the private sector 2. Ministry of Land, Infrastructure and Transport, Ministry of National Defence, Ministry of Science and ICT 3. Gyeongsangbuk-do, Ulsan, Daegu 4. Holding public data competition and finding business models
Low	Low Focusing external stakeholder's role in issue generation High	

표기 : 1. 주요 이슈, 2. 주요 중앙행정기관, 3. 주요 지방자치단체, 4. 주요 정부 사업

Composition : 1. Main issue 2. Main central administrative agency, 3. Main local government, 4. Main government business

활성화 정책, (2)효율성 중심 이슈는 빅데이터로 대표되는 4차 산업 육성 및 공공데이터 시스템 고도화 등 데이터 관련 기술의 성장, (3)참여성 중심 이슈는 생활형 데이터 개방으로 인한 시민 참여 증대 및 다양한 이해관계자들 간의 협력과 갈등, (4)혁신 중심 이슈는 아이디어 공모전 및 경진 대회 등을 통한 민간분야의 공공데이터 활용과 새로운 창업 모델 발굴을 의미한다. 이슈별 주요 행정 기관으로는 각각 기획재정부, 산업통상자원부, 문화체육관광부, 국토교통부가 해당되었으며 지역으로는 서울특별시, 인천광역시, 경기도, 경상북도가 각 이슈를 대표하였다. 기관별 입찰공고 수를 살펴보면 대부분의 사업들이 기획재정부, 과학기술정보통신부, 서울특별시, 경기도와 같은 주요 기관들에 집중되어 진행되고 있었는데 이는 세출예산을 고려해도 의미 있는 격차였다. 입찰공고를 통해 본 정부 사업 방향은 투명성이 데이터 관리 및 통합 플랫폼 구축, 효율성이 빅데이터 분석과 시스템 고도화, 참여성이 시민 맞춤형 데이터 개방 확대, 공공데이터 공모전 개최 및 비즈니스 모델 발굴이었으며 언론에서 드러난 이슈의 기본 틀로서 작용하고 있음을 알 수 있었다.

분석 결과를 토대로 살펴본 공공데이터 이슈의 특성은 다음과 같다. 첫째, 국내 공공데이터 관련 이슈에서 정부가 담당하는 역할이 매우 크다. 참여성과 혁신의 경우 외부 이

해관계자의 역할이 강조되는 이슈임에도 불구하고 우리나라는 주로 정부의 주도 하에 이루어지고 있다. 특히 혁신 중심 이슈의 핵심 단어가 ‘공모전’, ‘대회’일 정도로 민간분야의 데이터 활용이 정기적인 행사로 인해 발생되고 있었으며 ‘창업’ 또한 정부의 창업 지원 센터 및 대회로 인한 것이었다. 따라서 정부는 시민들의 자발적인 공공데이터 재사용을 도모하기 위해 진입장벽이 높은 활용 측면보다 데이터 수요 조사 및 오픈API 데이터 비율 확대 등의 개방 측면을 강화시킴으로써 공공데이터의 실용성을 높이는 데에 우선적으로 집중할 필요가 있다. 둘째, 참여성과 혁신이 중심이 되는 이슈가 계속해서 더욱 주목받을 것이다. 언론에 드러난 비율이 높은 이슈는 투명성과 관련된 주제들이었지만 정부 기관들과 함께 언급된 비중이 높은 이슈는 참여성과 혁신이었다. 그만큼 데이터 활용 부문에 대한 투자가 많이 이루어지고 있으며, 특히 일반 시민들의 참여를 유도하기 위해 노력하고 있음을 알 수 있었다. 이에 따라 참여성과 혁신은 국내 공공데이터 관련 이슈의 핵심이며 앞으로도 비중이 증가할 것으로 기대된다. 그러나 현 정부는 공공데이터 공개와 활용에 있어서 높은 비용과 복잡한 사용상의 절차로 인해 국민의 참여율이 저조한 형편이다. 공공데이터는 일종의 공공재로써 너무 과다한 비용을 요구하면 사실상 조건이 차단되기 때문에 비용을 대폭 낮춰서 누구

나 접근할 수 있게끔 해야 한다²⁷⁾. 정부는 이러한 이슈의 흐름이 실적을 위한 활동에 그치지 않도록 명확한 기준을 토대로 각 기관의 정책을 평가하고 가이드라인을 제시할 수 있는 방안에 대해 고민해야 할 것이다.

본 연구는 빅카인즈에서 수집한 온라인 뉴스와 나라장터에서 수집한 정부 입찰공고만을 대상으로 공공데이터 관련 이슈를 분석하였다는 점에서 한계가 있다. 하지만 뉴스 분석을 통해 공공데이터 관련 이슈를 도출하고 이를 정부 입찰공고와 비교해가며 기관별 이슈와 사업을 분석했다는 점에서 의의를 가진다.

참 고 문 헌 (References)

- [1] Ministry of Public Administration and Security, Get the Data from the Public Data Portal, 2014.
- [2] Ministry of Public Administration and Security, Government Innovation Comprehensive Plan, 2018.
- [3] Man-Ki Moon, and Tae-Yong Kim, "Development of Digital Games Based on Historical Material and its Design Components - With History Based Games of 5 Countries", *Journal of Broadcast Engineering*, Vol.12, No.5, pp.460-479, 2007, <http://dx.doi.org/10.5909/JBE.2007.12.5.460>.
- [4] Gu Kim, "An Evaluation of Public Data Open Policy: Focused on Public Data Portal", *Chung-ang Public Administration Review*, Vol.31, No.2, pp.57-82, 2017, 06.
- [5] Open Data Portal, <http://data.go.kr>.
- [6] Hyungjun Seo, and Seunghwan Myeong, "Policy Alternatives for Public Data Disclosure by Local Government : Focusing on Operation of Public Data Portal Site", *Journal of Korean Association for Regional Information Society*, Vol.18, No.4 pp.1-27, 2015.
- [7] Junwon Lim, and Gyunghyun Choi, "The Influence of Open Data Policies on Public Innovation", *Journal of the Korean Institute of Industrial Engineers*, Vol.43, No.1, pp.19-29, 2017, <http://dx.doi.org/10.7232/JKIIIE.2017.43.1.019>.
- [8] Hyungjun Seo, and Seunghwan Myeong, "Policy Alternatives for User-oriented Public Data Utilization: Focusing on ICT Managers' Perception in Private Sector", *Journal of Korean Association for Regional Information Society*, Vol.17, No.3, pp.61-86, 2014.
- [9] Sang Yong Lee, "Promotion of Public Data Opening in Chungbuk through Analysis of Public Data Portal", *Chungbuk Focus*, No.103, pp.1-19, 2015, 02.
- [10] Kyungseo Lee, and HyeJung Moon, "A Study on the Change of Value of Public Data through News Semantic Network Analysis", *Winter Congress of the Lottery Society*, Seoul, Korea, pp.101-119, 2018, 02.
- [11] Thorihildur Jetzet, Michel Avital, and Niels Bjorn-Andersen, "Data-Driven Innovation through Open Government Data", *Journal of Theoretical and Applied Electronic Commerce Research*, Vol.9, No.2, pp.15-16, 2014, <http://dx.doi.org/10.4067/S0718-18762014000200008>.
- [12] Wouter van Atteveldt, *Semantic network analysis: Techniques for Extracting, Representing, and Querying Media Content*, BookSurge Publishers, Charleston SC, pp: 18-23, 2008.
- [13] HyeJung Moon, and SungKyung Kim, "Study on Perceptions through Big data Analysis on Gambling related News in Korea", *Journal of Broadcast Engineering*, Vol.22, No.4, pp.438-447, 2017, <http://dx.doi.org/10.5909/JBE.2017.22.4.438>.
- [14] Yoo-Soo Jung, "The Study of Composition of Medical Examination Service by Semantic Network Analysis", *The e-Business Studies*, Vol.18, No.4, pp.197-214, 2017, <http://dx.doi.org/10.20462/TeBS.2017.08.18.4.197>.
- [15] Doerfel M. L., "What Constitutes Semantic Network Analysis? A Comparison of Research and Methodologies", *Connections*, Vol.21, No.2, pp.16-26, 1998.
- [16] Chunsob Kwak, Ilsoo Kyung, and Hyunji Lee, "A Study on the Recognition of Information Acceptor about Civil Defence Alert Broadcasting", *Journal of Broadcast Engineering*, Vol.20, No.6, pp.827-836, 2015, <http://dx.doi.org/10.5909/JBE.2015.20.6.827>.
- [17] Eung Sook Kim, "Global Media Environments and Glocalism Contents as Alternatives for Cultural Diversity", *Journal of Broadcast Engineering*, Vol.12, No.5, pp. 480-490, 2007, <http://dx.doi.org/10.5909/JBE.2007.12.5.480>.
- [18] BigKinds, <http://www.kinds.or.kr>.
- [19] Korea ON-Line E-Procurement System, <http://g2b.go.kr>.
- [20] Clauset, A., M. E. J. Newman, and C. Moore, "Finding Community Structure in Very Large Networks", *Physical Review E*, Vol.70, No.6, 2004, <https://doi.org/10.1103/PhysRevE.70.066111>.
- [21] Koichi Higuchi, *KH Coder 3 Reference Manual*, pp.43-44, Aug. 3, 2017.
- [22] Thorihildur Jetzet, Michel Avital, and Niels Bjorn-Andersen, "Generating Value from Open Government Data", *Thirty Fourth International Conference on Information Systems*, Milan, Italy, 2013.
- [23] Scott, M., "Comparing Corpora and Identifying Key Words, Collocations, and Frequency Distributions through the WordSmith Tools Suite of Computer Programs", *Small Corpus Studies and ELT: Theory and Practice*, pp.47-67.
- [24] Christopher Hood, *The Tools of Government*, London, Mac millan, pp.4-18, 1983.
- [25] Christopher Hood, and Helen Z. Margetts, *The Tools of Government in the Digital Age*, Palgrave Macmillan, 2007.
- [26] SangRak Kim, and ManMo Kang, "Today and Tomorrow of the Technology in Big data Analysis", *Communications of the Korean Institute of Information Scientists and Engineers*, Vol. 32, No. 1, pp. 8-17. 2014.

27) 뉴시스(2018.02.19), 文대통령, 2주만에 수보회의 주재...공공데이터 개방 혁신 논의

저 자 소 개



문 혜 정

- 1995년 2월 : 서울과학기술대학교 전자계산학과 공학사
- 1998년 2월 : 서울과학기술대학교 전자계산학과 공학석사
- 2013년 2월 : 서울과학기술대학교 공공정책학 박사
- 2018년 3월 : 동방문화대학원대학교 미래예측학과 박사과정
- 1995년 1월 ~ 2009년 3월: 대상정보기술, KCC정보통신, 한솔CSN, KCC솔루션즈, 연합복권 전산개발 및 설계
- 2009년 8월 ~ 현재 : 아이엘피 대표, 복권학회합동조합 이사장
- ORCID : <https://orcid.org/0000-0001-8265-3256>
- 주관심분야 : 데이터 분석, IT정책, 사행산업, 유가(儒家), 제자백가



이 경 서

- 2018년 8월 : 한국외국어대학교 미디어커뮤니케이션학사 졸업예정
- ORCID : <https://orcid.org/0000-0003-2969-6032>
- 주관심분야 : 데이터 마이닝, 의미망분석, 미디어 콘텐츠