



‘빅데이터’ 관련 신문기사의 의미연결망 분석

A Semantic Network Analysis of the Newspaper Articles on Big data

저자
(Authors) 최윤정, 권상희
Choi, Yoon-jung, Kweon, Sang-Hee

출처
(Source) [사이버커뮤니케이션 학보 31\(1\)](#), 2014.3, 241-286 (46 pages)
[Journal of Cybercommunication Academic Society 31\(1\)](#),
2014.3, 241-286 (46 pages)

발행처
(Publisher) [사이버커뮤니케이션학회](#)
Cybercommunication Academic Society

URL <http://www.dbpia.co.kr/Article/NODE02380852>

APA Style 최윤정, 권상희 (2014). ‘빅데이터’ 관련 신문기사의 의미연결망 분석. 사이버커뮤니케이션 학보, 31(1), 241-286.

이용정보
(Accessed) 아주대학교
202.30.23.141
2016/03/08 10:17 (KST)

저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다.
이 자료를 원저작자와의 협의 없이 무단게재 할 경우, 저작권법 및 관련법령에 따라 민, 형사상의 책임을 질 수 있습니다.

Copyright Information

The copyright of all works provided by DBpia belongs to the original author(s).
Nurimedia is not responsible for contents of each work. Nor does it guarantee the contents.

You might take civil and criminal liabilities according to copyright and other relevant laws if you publish the contents without consultation with the original author(s).

‘빅데이터’ 관련 신문기사의 의미연결망 분석

최운정*

성균관대학교 신문방송학과 박사과정

권상희**

성균관대학교 신문방송학과 교수

요약

본 연구에서는 최근 사회적 화두가 되고 있는 빅데이터 관련 신문기사를 분석, 새로운 기술과 현상인 빅데이터에 대해 언론이 어떻게 보도하고 있는지를 파악하고자 하였다. 이를 위해 ‘빅데이터’라는 용어가 신문기사에 처음 등장한 2011년부터 2013년까지 경향신문, 동아일보, 조선일보, 한겨레신문 등 4개 전국종합일간지에 보도된 관련 기사를 수집하여 의미연결망 분석을 실시했다.

분석결과, 빅데이터 관련 신문보도의 의미구조는 2011년 중립적 성격에서 2012년부터 긍정적 방향으로 자리매김하고 있음을 알 수 있었고 대부분의 보도가 실용적 차원의 활용가치에 집중되어 있어 미래사회를 향한 경쟁을 촉구하고 그 결과로서 예측되는 장밋빛 정보사회를 기술하고 있다고 추론해볼 수 있었다. 진보성향의 매체인 경향신문·한겨레신문에서 빅데이터가 초래할 수 있는 부정적 측면을 제시하고 있기는 하지만 핵심적 이슈로 부각시키지는 못한 점에서 매체성향에 따른 의미구조의 차이가 두드러지지는 않았다. 또한 활용사례로서 기업의 마케팅이나 관련 업체들의 이슈만이 아니라 국민건강이나 교통정책과 같은 공공부문에서도 빈번하게 논의됨으로써 빅데이터 사회를 향한 기업 뿐 아니라 국가적 열망을 읽게 했다. 본 연구는 사회적 관심이 높아지고 있는

* emily0717@gmail.com, 제1저자

** skweon@skku.edu, 교신저자

빅데이터에 대해 신문이 어떻게 보도하고 있는가를 실증적으로 분석한 점, 그리고 정보통신기술의 발달로 나타난 새로운 기술 현상이 사회적 인식의 변화에 끼칠 수 있는 방향성을 살펴본 점에서 시사점을 찾을 수 있다.

키워드: 빅데이터, 신문보도, 단어, 의미연결망분석

1. 서론: 문제제기

‘빅데이터 현상’이 아니라 “빅데이터 소란”(이재현, 2013)이라고 표현되는 것도 무리는 아니다. 단지 정보통신 영역만이 아니라 정치, 경제, 사회 각 분야에서 ‘빅데이터’가 화두다. 세계적인 시장조사업체들이 혁신과 경쟁의 “넥스트 프린터”(McKinsey, 2011) 혹은 “21세기 원유”(Gartner, 2011) 등의 수식어와 함께 앞다퉀 ‘빅데이터’를 미래의 핵심기술이자 유망산업 분야로 주목하면서 빅데이터에 대한 관심이 기술을 넘어 사회 각 분야로 확대되고 있다.

사회적 관심이 커진 데에는 국내 언론의 뉴스보도도 예외가 아니다. ‘방대한 데이터’라는 표현 대신 2011년 처음으로 ‘빅데이터’라는 용어를 사용하기 시작한 언론은 기술뿐 아니라 사회의 새로운 트렌드로서 빅데이터를 소개하고 있다. 덕분에 빅데이터에 내재된 잠재력은 일반 공중에 게도 더 이상 낮설지 않다. 구글의 독감예보는 가장 흔하게 언급되는 활용 사례다. 구글은 감기와 관련된 주제를 검색하는 빈도가 늘어나면 독감환자가 늘어난다는 사실을 발견하고 이용자의 검색 데이터를 활용해 보건당국보다 먼저 독감유행 정보를 예측한 바 있다. 2012년 치러진 대통령 선거 또한 빅데이터가 왜 국내에서 사회적 열풍을 일으키고 있는지

를 일면 보여줬다. 소셜 정보를 토대로 사회 흐름을 읽고 나아가 이슈나 여론 분석, 선거결과 예측 등이 각광을 받았기 때문이다. 문화체육관광부가 SNS 데이터 분석을 토대로 발표한 ‘국민 라이프스타일’ 조사 보고서도 같은 맥락이다. 이에 따르면 “오늘 한국인은 현재와 일상을 중요하게 생각하고 작은 소비에 행복을 느끼며, 혼자 즐기는 데 익숙하다”고 총평하고 있다(주간동아, 2013.7.15).

그렇다면 언론은 새로운 관심사로 떠오른 빅데이터를 일반 공중에게 어떻게 의미화하고 어떠한 논리와 이슈를 구성하여 전달하고 있을까. 이미 수많은 프레임 연구나 담론 연구에서 밝혀진 것처럼 언론은 있는 그대로의 현실이 아니라 언론이 구성하는 현실을 보도한다. 이러한 측면에서 우리의 현실로서 자리잡기 시작한 빅데이터가 미디어를 통해서 어떻게 구성되고 있는지를 살펴보는 것은 중요하다. 일반 공중의 입장에서 새로운 기술과 현상에 대한 접근은 주로 미디어를 통해 이뤄지며 또한 미디어 보도가 어떻게 의미화하고 있는가는 일반 공중의 이슈 해석은 물론 사회적 이슈 구성에도 영향을 끼칠 수 있기 때문이다. 특히 선행 연구에서는 빅데이터 분석기술과 활용가치, 그리고 사회문화적 의미 분석에 초점을 맞추으로써 언론이 매개하는 빅데이터 이슈의 의미화 과정을 밝혀내지는 못했다.

이에 따라 본 연구는 빅데이터에 대한 언론의 뉴스보도 분석을 통해 빅데이터에 대한 이슈와 가치 가운데 무엇이 강조되고 있으며 새로운 기술과 현상에 대해 우리 사회가 어떻게 접근하고 있는가를 경험적으로 살펴보고자 한다. 이를 통해 빅데이터에 대한 사회적 인식의 방향이 어떻게 나아가고 있는가도 추론해볼 수 있을 것이다. 본 연구는 특히 메시지의 의미구조를 살펴보는 데에 유용한 의미연결망 분석 기법을 활용하고자 한다. 의미연결망 분석은 사회현상을 탐구하는 데 있어 행위자들의 관계 구조를 통해 특징을 밝혀내는 사회연결망 분석을 원용한 기법이다.

따라서 본 연구에서는 행위자 대신 신문기사에 보도된 뉴스언어들을 분석 대상으로 하여 미디어가 선택하는 특정 단어들의 현저성과 이들 단어들이 서로 관계를 맺는 상호작용의 구조에 주목하고자 한다. 각각의 뉴스언어들은 고정된 속성으로만 존재하는 것이 아니라 서로 연결되는 관계의 맥락 속에서 의미와 논리를 만들어내고 통제하는 역할을 할 수 있기 때문이다. 어떠한 단어가 상징적으로 사용되고 어떠한 단어가 빈번하게 사용되며 단어들 사이의 관계성이 어떠한지를 밝힘으로써 언론이 빅데이터 현상에 대해 전략적으로 구축하고자 하는 의미구조와 이슈를 찾아보고자 한다.

2. 이론적 논의

1) 빅데이터 이슈와 선행연구

개념적 정의가 아직 명확하지 않은 빅데이터는 세계적인 컨설팅 기업들이 2011년부터 미래의 유망기술로 선정하면서 정보통신기술 분야와 이의 활용분야 차원에서 주요한 이슈로 다뤄져 왔다. 미국의 시장조사 컨설팅업체인 가트너(Gartner)는 2011년 빅데이터의 등장과 관련, ‘신기술 발생단계’로 주목하면서 성장을 예측한 바 있다. 이처럼 IT나 기업 경영 컨설팅 기관들의 주도 아래 정의되어온 빅데이터는 “데이터베이스가 저장, 관리, 분석할 수 있는 범위를 초과한 규모의 데이터이자 기존의 방법이나 도구로는 수집, 저장, 검색, 분석, 시각화가 어려운 정형 또는 비정형 데이터”(McKinsey, 2011)를 의미한다. 또 데이터의 방대한 양(Volume), 데이터 유형의 다양한 형태(Variety), 빠르게 처리되는 속도(Velocity)라는 속성과 새로운 가치(Value)의 자원이 될 수 있다는

관점에서 4V로 요약되기도 하고 처리기술의 복잡성(Complexity)을 추가하여 정의되기도 한다(Gartner, 2012; SAS, 2012). 그밖에도 데이터에서 경제적 가치를 발굴하는 차세대 기술(IDC, 2012), 대규모 데이터와 관련된 기술(SERI, 2012) 또는 의미를 찾아내는 데이터 처리기술(LG경제연구원, 2012)로 정의되면서 ‘빅(Big)’이라는 뜻에서 연유하는 양의 개념이거나 통신기술을 지칭하는 데 머물지 않고 사회적 활용으로 확장하려는 새로운 기술적 의미가 강조되고 있다.

빅데이터의 출현 배경은 데이터의 생산과 축적이 가능해진 기술 환경의 변화에 있다. 스마트폰과 같은 스마트 기기의 확산, SNS 이용증가 등은 수많은 데이터 생산을 가능하게 했고 또 다른 한편으로는 이렇게 생성된 수많은 정보가 온라인 공간에 저장되고 있기 때문이다. 따라서 새로운 정보통신기술과 이를 바탕으로 한 커뮤니케이션 방식의 변화로 등장한 개념이라고 볼 수 있다. 정보사회의 패러다임 차원에서 보면 PC 시대, 인터넷시대, 그리고 모바일시대에 이어 나타난 스마트시대의 대표적인 IT이슈이다(한국정보화진흥원, 2013).

그러나 빅데이터가 사회적 관심사로 떠오른 것은 빅데이터의 실용적 활용 가치가 언급되면서부터다. 빅데이터의 성공적 활용사례로 빠지지 않는 구글의 독감트렌드가 이를 웅변적으로 보여준다. 구글은 2008년부터 검색 데이터를 통해 검색어, 시간대, 위치정보의 상관관계를 밝힘으로써 독감 유행 수준을 관련기관보다 빠르게 예측해왔다. 또 2008년 미국 대통령 선거에서 유권자 데이터베이스를 통해 맞춤형 선거전략을 펼쳤던 오바마 사례도 대표적인 예로 꼽힌다. 이후 트위터 등 다양한 SNS 데이터를 통한 여론이나 시장트렌드 분석, 의학정보 분석과 예측, 교통시스템과 같은 공공정책 수립 등 공공과 민간 영역 모두에서 활용이 급증한 것은 “데이터의 사회화”(김예란, 2013)를 통해 빅데이터에서 실용적 가치를 창출하려는 시도들이라고 할 수 있다.

빅데이터에 대한 관심은 학문 분야에서도 점차 확대되고 있는 추세인데 빅데이터에 관한 선행연구는 크게 세 가지 방향에서 이뤄지고 있다. 첫 번째는 빅데이터 저장이나 분석 플랫폼에 관한 기술적 차원의 연구들(이동환·박정찬·유찬근·윤호상, 2013 등)이다. 따라서 정보통신공학 적 접근들이라고 할 수 있다. 첫 번째와 더불어 가장 많은 연구가 이뤄지고 있는 두 번째는 빅데이터 활용에 관한 실용적 차원의 연구들이다. 예를 들어 경제흐름을 예측하여 불확실성을 줄이거나(김유신·김남규·정승렬, 2012) 문화산업(윤홍근, 2013)에 접목하여 콘텐츠 활용을 최적화하고 정부정책 수립과 같은 공공부문(조영임, 2013)에서 활용될 수 있는 방안에 대한 논의가 이뤄지고 있다. 그리고 세 번째는 빅데이터에 대한 사회문화적 함의를 논의하는 연구들이다. 빅데이터 활용이 가져오는 결과 중에서 프라이버시 문제 등을 새로운 감시와 통제의 영역에서 비판하거나(김예란, 2013; 이광석, 2013) 빅데이터 형성 과정에서 은폐되는 이용자 활동의 가치 문제를 다루거나(백옥인, 2013) 혹은 빅데이터에 대한 연구 경향이 '컴퓨터연산 사회과학'으로의 전환을 가져오고 있는지에 대해 살펴본 연구(이재현, 2013) 등이 이에 속한다.

이처럼 빅데이터에는 양면성이 있다. 기술발전과 활용이라는 실용적 기대가와 동시에 개인 프라이버시 침해 가능성이라는 어두운 측면이 있기 때문이다. 이러한 정보사회의 두 측면은 조지 오웰의 소설 '1984'의 감시와 통제를 떠올리게 하는 점에서 '디지털 울타리(Digital Enclosure)'의 역설로 표현되기도 한다(Gordon & de Souza e Silva, 2011).

그러나 소설 '1984'의 비유 '빅브러더(Big Brother)'가 개인사생활에 대한 하향식 감시의 전형이라면 솔로브(Solove, 2004)는 정보사회에서 이뤄지는 감시는 한층 더 심각하다고 파악한다. 이용자들이 수많은 정보 축적 과정에서 정보 노출에 대한 위협을 인식하지 못하거나 덜 느끼며 이렇게 축적된 정보들이 기업경영과 같은 상업적인 차원에서 이용되면서

전통적인 감시모델보다 더 분권화되고 총망라적인 방식으로 변화하고 있기 때문이다. 특히 정보사회에서 데이터는 가치창출의 토대이지만 이를 활용하는 과정에서 데이터 생성 주체인 개개인은 소외되거나 프라이버시를 침해당할 수 있고 상업적으로 활용하는 경우에는 ‘경제적 감시(economic surveillance)’의 문제로 확대된다(Allmer, 2012, 백옥인, 2013 재인용). 결국 정보사회가 가져다주는 문제는 하향식 감시만이 아니라 누가 정보에 접근하고 정보가 어떻게 이용되는지를 파악하는 정보 접근권과 통제권의 불균형 측면에서 더 심각해질 수 있다(Gordon & de Souza e Silva, 2011). 마노비치(Manovich, 2011)는 빅데이터 사회를 정보의 생산능력, 수집능력, 그리고 분석능력에 따라 피라미드 구조를 형성하는 차별적 ‘데이터 계급(data-classes)’을 구분한 바 있다(김예란, 2013 재인용). 또한 이미 많은 논쟁을 일으킨 구글의 온라인 지도 서비스 스트리트 뷰(Google Street View)처럼, 공적 공간과 사적 공간의 경계 기준 등 전통적인 개념을 모호하게 만들고 있는 것도 데이터의 활용이 가져온 결과 중 하나이다(Gordon & de Souza e Silva, 2011).

따라서 빅데이터 현상은 실용적 차원의 활용방식과 가치창출의 이슈로 요약되는 장밋빛 정보사회에 대한 기대감만이 아니라 정치적 차원의 감시와 통제의 문제, 경제적 차원의 정보지식과 활용권력의 불균형 문제, 이러한 정보사회의 새로운 패러다임이 가져오는 사회적 인식의 변화 문제 등 다차원적으로 접근할 필요가 있는 것이다.

2) 기술 도입과 뉴스보도의 의미화

앞서 살펴본 빅데이터에 대한 정의들은 빅데이터를 사회적 기술로 접근하는 시각과 함께 빅데이터에 대한 사회적 인식과 태도를 보여준다(김

예란, 2013). 또한 이러한 새로운 기술과 현상에 대한 인식은 미디어의 보도를 거치면서 다시한번 재구성된다. 뉴스보도가 현실 그 자체가 아니라 미디어의 구성물이라는 것은 사회구성주의적 관점에서 수많은 연구가 이뤄져 왔는데 대표적으로 뉴스 프레임 연구가 이에 속한다. 이에 따르면 사람들의 이슈 인식에 영향을 주는 뉴스 현실은 미디어의 선택, 배제, 강조 등의 방식을 거쳐 매개된 현실이며 세상을 보는 창, 즉 '프레임'을 미디어가 설정한다는 것이다(Entman, 1991).

특히 새로운 이슈의 확산 초기에는 미디어 보도가 결정적 역할을 하는 경우가 많으며 갈등이슈나 과학기술과 같은 새로운 현상에 대한 보도의 경우에는 언론의 길잡이 역할이 더욱 중요해진다(Kitzinger & Reilly, 1997). 뉴스보도가 어떤 이슈나 현상에 의미를 부여하는 핵심적인 시각의 틀을 제공할 수 있기 때문이다(Tuchman, 1978; Pan & Kosicki, 1993). 쉘펠(Scheufele, 1999)은 또한 프레임 과정을 뉴스제작 과정에서 드러나는 프레임 형성과 구축, 수용자 차원의 프레임 효과 등으로 세분화하여 고찰했는데, 특정 이슈에 대한 미디어의 초기 보도 방향은 프레임의 첫 단계인 프레임 형성에 해당한다고 볼 수 있다. 최근 등장한 이슈이자 사회적 관심사로 떠오른 빅데이터에 대한 미디어의 보도가 어떠한 의미체계를 구성하고 있는지를 살펴보려는 것은 이러한 맥락에서다. 미디어 보도는 곧 빅데이터에 대한 사회적 인식의 형성에 영향을 끼칠 수 있기 때문이다.

새로운 과학이슈나 기술에 대한 미디어 보도와 관련하여 국내에서 이뤄진 연구는 그다지 많지 않은 편인데, 연구주제 측면에서는 윤리적 논쟁을 낳았던 인간배아복제, 줄기세포, 생명과학 관련 연구(김만재·전방욱, 2012; 김수정·조은희, 2005)를 비롯하여 유전자 연구(송해룡·김원제·조항민, 2005; 정재철, 2004), 원자력 발전(김원용·이동훈, 2005), 우주과학(박은선·이광형·김찬석, 2012) 등이 다뤄져 왔고 연구방법

측면에서는 프레임 분석이 주로 활용되었다.

인간배아복제 관련 신문기사를 분석한 김만재와 전방욱(2012)의 경우, 전통적인 프레임 분석기법과는 달리 언어네트워크분석 기법을 활용한 점이 특징적인데, 어떠한 단어들에 보도의 방향을 결정짓고 있는지를 밝혀냈다. 이에 따르면 ‘복제’라는 단어가 ‘치료’나 ‘성공’과 연결되느냐, 아니면 ‘윤리’나 ‘법안’과 연결되느냐에 따라 해석의 방향이 달라지며 2004년 황우석연구진 사례에서는 ‘국제적’, ‘성공’, ‘치료’라는 단어가 더 강조되면서 긍정적 프레임이 시도되었다고 분석했다.

과학기술과 마찬가지로 미디어기술이 사회의 변화로 직결된다는 관점에서 뉴미디어가 뉴스보도 경향성 연구의 분석대상이 되기도 했다. 예를 들어, 주재원과 나보라(2009)는 신문이 상대적인 뉴미디어 기술인 TV와 인터넷을 비윤리적이고 수준 낮은 미디어로 재현하는 반면 신문 스스로는 사회를 선도하는 엘리트적 미디어로 자리매김해온 뉴스보도 경향을 밝혀냈고, 권상희(2005)는 1989년부터 2004년까지를 대상으로 한 인터넷 기술 관련 신문뉴스 보도 프레임 연구에서 도입 초기에는 ‘정보화시대를 앞당길 첩병’ 프레임이 압도적이었고 상용화 이후에는 전체적으로는 호의적이지만 문제점을 다루는 비판적 프레임이 상대적으로 확대되어 왔다고 분석했다. SNS 관련 신문보도를 분석한 김형곤(2011)도 프라이버시 침해와 같은 부정적 측면의 보도는 배제되고 미래지향적이고 선구자적 이미지로만 묘사됨으로써 SNS의 확산에 영향을 끼치는 신문의 편향성을 찾아냈고 이수범과 강연곤(2013)은 보수성향의 신문은 트위터와 같은 SNS를 부정적으로 프레임하고 진보성향의 신문은 긍정적으로 프레임하고 있다고 밝힌 바 있다.

새로운 기술에 대한 미디어 보도 연구가 많지 않은 것과 관련, 김만재와 전방욱(2012)은 기술을 가치중립적으로 간주하는 데에 원인이 있다고 지적하며 기술적 이슈에 대한 미디어 보도는 사회가 기술적 이슈에

대해 허용할 수 있는 범위를 보여주고 사회적 인식 형성으로 이어진다고 주장한다. 빅데이터 또한 가치중립적으로 인식하거나 접근하기 쉽지만 총체적 변화의 출발점으로 작동할 수도 있고 사회전반에 대한 인식과도 맞닿을 수 있다(이재현, 2013).

앞서 서술했듯 뉴스 보도가 사회의 인식 변화에 영향을 미칠 수 있다는 것(Iyengar, 1991; Pan & Kosicki, 1993)은 뉴스 프레임 연구에서 꾸준히 입증되어 왔는데, 엔트만(Entman, 1991; 1993)의 경우, 프레임을 구성하는 요소 중 하나로 상징적인 뉴스언어에 주목했다. 기틀린(Gitlin, 1980)도 상징조작자가 일상적으로 언어적 담론을 조직하는 것을 프레임으로 보았다. 뉴스의 관점을 분석한 반 다이크(Van Dijk, 1998) 역시 특정 가치나 규범을 표현하는 의미구조화의 방법으로서 단어의 선택에 주목했다. 뉴스의 기본적인 구성요소가 단어이자 단어는 특정한 의도를 담고 있는 사회적 실천이기 때문이다(Schudson, 1995).

이를 종합하면, 뉴스보도에서 사용하는 단어와 단어들의 조합은 결코 우연이 아닌 것이다. 이런 관점에서 본 연구는 빅데이터 관련 신문기사에서 드러나는 단어들의 사용빈도와 주요 단어들 간의 의미적 연관구조를 파악함으로써 프레임의 출발점이 될 수 있는 미디어의 의미체계 구성 방식을 찾아보고자 한다. 또 미디어조직의 가치와 규범은 뉴스보도 태도에도 반영될 수 있기 때문에(Entman, 1991) 매체성향에 따라 뉴스단어를 통한 의미구조화가 서로 다르게 나타나고 있는지도 살펴보고자 한다.

3) 의미연결망 분석을 활용한 연구

의미연결망 분석(semantic network analysis)은 사회의 시스템 구조를 파악할 때 행위자 사이의 관계를 중시하는 사회연결망 분석(social

network analysis)을 커뮤니케이션 메시지에 적용시킨 방법이다. 사회 연결망 분석이 사회현상을 만들어내는 원동력을 각각의 개체가 아니라 개별 행위자가 연결되어 있는 관계성에 주목한다면 이를 원용한 의미연결망 분석은 행위자 대신 메시지의 구성요소인 단어들의 구조적 관계를 통해 의미를 분석한다(Wang & Rada, 1998; Wasserman & Faust, 1994).

전통적인 내용분석 방법이 노동집약적이고 연구자의 주관적 해석을 배제할 수 없다면 의미연결망 분석은 이러한 단점을 보완할 수 있는 컴퓨터 기반 메시지 분석 기법이라고 볼 수 있다(박한우 & Leydesdorff, 2004; 장하용, 2001) 나아가 개별 단어의 고정된 속성이나 개념이 아니라 상호 작용하는 관계의 맥락 속에서 역동적인 의미를 포착할 수 있는 장점이 있다(김상배, 2011).

이처럼 의미연결망 분석에서는 특정 유목을 가정하지 않은 상태에서 단어들의 빈도와 한 문장 안에서 동시에 사용되는 단어들의 관계를 통해 메시지(본 연구에서는 신문보도)의 의미화 패턴을 분석할 수 있다. 이때, 정보단위가 되는 단어나 구를 각각의 노드(node)를 형성하는 개념으로 놓고 개념 간의 연결 상태를 링크(link)로 나타내는데 여기서 링크로 드러나는 단어들의 공동출현(co-occurrence) 관계를 통해 의미를 해석한다(Wang & Rada, 1998). 단어들의 사용빈도와 관계는 텍스트에서 강조되는 상징성을 보여줄 수 있고 단어들의 결합을 통해 특정한 의미를 만들어내는 경향이 있기 때문이다(박한우 & Leydesdorff, 2004). 예컨대, ‘스마트’와 ‘사회’의 반복적인 동시 사용으로 ‘스마트 사회’라는 정보화 시대의 의미가 창출된다. 또 ‘실업’이라는 단어는 ‘도시화’와 연결될 때와 ‘구조조정’과 연결될 때 서로 상이한 맥락적 의미를 제공하며(장하용, 2001), 최근과 같은 상황에서는 ‘청년취업’과 강한 연결 상태를 보일 수 있기 때문에 사회적 상징성을 파악할 수 있게 한다.

의미연결망 분석은 개별 단어의 조합에만 주목하는 것은 아니다. 전체 연결망 안에서 어느 '위치'에 자리하고 있는지, 어떠한 '구조'를 형성하고 있는지에 주목함으로써(Wasserman & Faust, 1994) 메시지 안에 숨어 있는 커뮤니케이션 송신자의 의도를 계량화하고 구조적인 패턴을 추출할 수 있다(김상배, 2011; 박지영·김태호·박한우, 2013).

이런 배경에서 의미연결망 분석을 활용한 메시지 분석 연구들이 점차 늘어나고 있다. 장하용(2001)은 시민단체 보고서와 매체비평지의 기사를 대상으로 의미연결망 분석을 실시하여 시민단체가 바라보는 언론비평의 특징을 분석했고 박성희(2009)의 경우, 제17대 대통령 선거 후보 토론 텍스트를 분석하여 각 이슈에 따라 후보별로 논제구성의 논리와 일관성, 그리고 차이를 찾아냈다. 또 박지영 등(2013)은 아이돌 스타들의 SNS 메시지에 드러나는 단어들을 정체성 지표로 간주, 커뮤니케이션 방식을 규명했다. 그밖에도 심홍진 등(2011)은 스마트폰과 소셜미디어 이용자들의 미디어 이용행태를 분석함에 있어 인터뷰 텍스트를 바탕으로 의미연결망 분석을 실시했다.

신문기사를 분석 대상으로 한 의미연결망 분석은 남인용과 박한우(2007)의 연구와 김만재와 전방욱(2012)의 연구에서 찾아볼 수 있는데, 남인용과 박한우의 경우에는 대선 예비후보자에 대한 신문기사에서 사용된 단어와 연결망 구조를 통해 신문에서 만들어지는 후보 이미지를 추출했다. 또 김만재와 전방욱은 인간배아복제에 대한 신문기사를 분석대상으로 의미연결망 분석을 실시하여 핵심적 위치에 있는 이슈와 주변부 위치에 있는 이슈를 구분하고 이를 통해 뉴스의 보도 특성을 밝혔다.

이러한 선행연구를 통해 살펴볼 때, 의미연결망 분석은 신문기사에 등장하는 주요단어의 사용빈도, 주요단어(node) 사이의 연결(link) 상태, 그리고 네트워크의 연결 구조(structure)를 통해 신문기사 텍스트에 담긴 의도적인 의미의 흐름을 파악하는 데에 유용한 방법론으로 활용될 수

있다.

따라서 본 연구에서도 의미연결망 분석을 활용하여 빅데이터 관련 신문기사에 등장하는 특정 단어들의 현저성과 단어들 사이의 관계와 질서 구조를 통해 신문기사가 어떠한 의미와 이슈를 구성하고 있는지를 찾아 보고자 한다. 이는 곧 미디어 프레이밍이 어떻게 형성되는가를 이해할 수 있는 하나의 선결적 작업이 될 수 있기 때문이다.

지금까지의 이론적 논의를 바탕으로 본 연구에서 구체적으로 살펴보려는 연구문제는 다음과 같다.

- 연구문제 1: 보도시기별 빅데이터 관련 신문기사에 나타나는 주요 단어는 무엇인가?
- 연구문제 2: 보도시기별 빅데이터 관련 신문기사에 나타나는 주요 단어들의 연결망은 어떠한 구조적 특성을 보이는가? 이 연결망의 특성은 어떠한 의미를 구성하는가?
- 연구문제 3: 매체성향에 따라 주요 단어 사이의 의미 연결망은 어떠한 차이를 보이는가?

3. 연구방법

1) 분석대상과 자료수집

본 연구의 분석기간은 ‘빅데이터’라는 용어가 신문기사에 처음으로 등장한 시점부터 현 시점까지로 하였으며 구체적으로는 2011년 1월부터 2013년 12월까지를 분석기간으로 설정하였다. 분석대상은 경향신문, 동

아일보, 조선일보, 한겨레신문에 보도된 빅데이터 관련 신문기사의 제목 (부제 포함)이다. 조선일보·동아일보와 경향신문·한겨레신문을 선정 한 이유는 각각 보수성향과 진보성향을 대표하는 매체로 분류되어 온 선행 연구(이수범·강연곤, 2012 등)를 토대로 신문의 이념적 성향까지 고려 하는 이 연구의 목적에 부합하다고 판단하였기 때문이다. 또한 신문기사의 제목만을 분석대상으로 한 것은 제목은 기사의 내용을 압축적으로 전달하고 의제를 설정하는 역할을 하며 동시에 신문의 관점과 태도를 유추 할 수 있는 지표이기 때문이다. 즉, 기사의 함축적인 의미는 기사 제목에 서도 표출된다(김수정·조은희, 2005; de Vreese, 2004).

자료수집은 '빅데이터'를 키워드로 검색하여 이뤄졌으며 경향신문, 동아일보, 한겨레신문 기사의 경우 한국언론진흥재단의 기사데이터 베이스 카인즈(www.kinds.or.kr)를 통해 수집하였고 이곳에서 검색서비스가 제공되지 않는 조선일보는 신문사의 웹사이트 데이터베이스를 이용하였다. 이에 따라 중복기사를 제외하고 최종 분석대상으로 선정된 기사의 시기별, 매체별 검색결과는 <표1>에 정리되어 있다.

<표 1> 시기별·매체별 기사건수

매체	2011(1~12월)	2012(1~12월)	2013(1~12월)	총계
경향신문	0	8	45	53
동아일보	3	50	110	163
조선일보	1	47	61	109
한겨레신문	1	8	31	40
총계	5	113	247	365

시기별로 살펴보면 빅데이터 관련 기사가 2011년도에 미미하게 나타나기 시작해 2012년도에 총 기사건수가 20배 이상으로 폭증하고 있음을

알 수 있다. 2013년에도 2012년도보다 두 배 이상 늘어나며 증가세를 보였다. 매체별로는 각 신문마다 평균 지면수가 상이하기는 하지만 동아일보와 조선일보가 각각 누적기사건수 163건, 109건을 기록하며 상대적으로 보도량이 많았고 경향신문과 한겨레신문은 각각 53건, 40건이었다. 특히 동아일보는 해마다 관련 기사가 급증하며 가장 많은 보도량을 기록했다.

2) 분석방법

이러한 과정을 통하여 수집된 기사는 의미연결망분석 기법을 활용하여 분석되었다. 우선 한국어 텍스트 분석 소프트웨어인 KrKwic(박한우 & Leydesdorff, 2004)를 이용하여 개별 기사 제목(부제 포함)에 등장하는 단어들의 빈도순위에 따라 주요단어를 추출하였고 개별 기사제목 안에서 주요단어들이 공동으로 출현하는 빈도수를 파악하여 단어×단어의 행렬데이터를 구성하였다. 이때 관사, 접속사, 문장의 어미 등은 제외됐으며 유사한 맥락에서 사용된 단어들은 단일 단어로 일원화하는 과정을 거쳤다. 예컨대 ‘대선’, ‘총선’ 등은 ‘선거’로 통일시켜 빈도수를 산출하였다. 또 빅데이터의 예측력을 설명하기 위해 기사 안에서 사례로 언급된 ‘고객의 마음’, ‘욕망’, ‘감정’, ‘관심사’ 등은 낱말의 단어로 처리함으로써 주요단어 순위에서 누락되어 중요성이 간과되는 것을 피하기 위해 ‘개인성향’으로 균집화한 빈도수를 반영했다. 마찬가지로 ‘주가’, ‘기름값’ 등은 ‘시장동향’으로, ‘구글’, ‘야후’, ‘NHN’ 등은 ‘포털업체’로, ‘자살’, ‘범죄확률’, ‘광우병공포’ 등은 ‘사회이슈’로 균집화하여 빈도수를 산출하였다.

이러한 과정을 거쳐 만들어진 행렬데이터를 기반으로 단어들 사이의 연결구조와 특성을 파악하기 위해 사회연결망 분석 소프트웨어인 UCINET의 넷드로(NetDraw) 기능을 활용하여 공동출현 단어들의 연

결망을 시각화하였다. 이를 통해 연결망의 구조와 연결강도를 파악하였으며 각 노드(단어)가 연결망에서 차지하는 구체적인 역할을 가늠하기 위해 본 연구에서는 프리만(Freeman, 1979)의 중심성 개념 정리와 측정 방법을 토대로 계량화한 지표인 연결정도 중심성(degree centrality)과 매개중심성(betweenness centrality)을 살펴보았다.

프리만은 연결망 안에서 각 노드의 영향력을 뜻하는 중심성이 다양하게 나타날 수 있다고 보았다. 우선, 연결망 중심 위치에서 여러 노드들과 직접적으로 가능한 한 많이 연결된다면 이 노드의 핵심적 영향력을 쉽게 떠올릴 수 있다. 연결된 노드들의 합인 연결정도 중심성이 이에 속한다. 그러나 네트워크 크기에 제한되지 않는 상대적 값과 연결 강도를 고려한 활동성(activity)으로 파악한 점에서 링크수의 단순 합과는 다르다. 또한 특정 노드를 통하지 않으면 이와 연결된 노드들이 단절되어 의미가 끊어지는 경우도 있을 수 있다. 활동성과는 별개인 통제력을 의미한다. 노드와 노드 사이에서 다리 역할을 하는 중개 영향력이라고 할 수 있으며 프리만은 이것을 매개중심성으로 구분했다(Freeman, 1979).

4. 연구결과

1) 보도시기별 주요단어 비교분석

‘빅데이터’라는 용어가 신문에서 처음 언급된 2011년에는 기사보도량도 적었고 사용된 단어도 많지 않았다. 뚜렷한 특징을 파악하기에는 무리가 있으나 ‘분석’(6회)이 가장 많이 사용됐고 이어서 ‘빅데이터’, ‘개인정보’, ‘SNS’가 각각 4회씩 언급되며 SNS를 통한 개인정보가 빅데이터의 분석 대상으로서 활용되기 시작했음을 알 수 있다. 또한 정보통신과

관련한 기술용어(‘분석’, ‘빅데이터’, ‘통계’, ‘검색’, ‘기록’, ‘저장’, ‘경로’, ‘기계’)와 활용과 관련된 단어(‘SNS’, ‘독감’, ‘여론’, ‘기업’, ‘구글’, ‘파악’, ‘유튜브’, ‘K팝’, ‘경영’, ‘활용’, ‘행간’, ‘예측’)가 총 단어 출현빈도에서 68.5%를 기록하며 대다수를 차지하고 있다. 한편 가치를 함유한 것으로 추정할 수 있는 단어들은 ‘활용’, ‘행간’, ‘누군가’, ‘반대’, ‘유포’ 등이 각각 1회씩 사용되었다. 전체에서 차지하는 비중은 10.2%로 높지 않지만 부정적 성격의 단어가 이중 60%를 차지하고 있는 점이 눈여겨볼 만하다.

〈표 2〉 2011년 ‘빅데이터’ 관련 신문기사 제목에 나타나는 주요단어

사용 빈도	주요단어
6	분석
4	개인정보, 빅데이터, SNS
2	검색, 구글, 기록, 미래, 파악
1	경로, 경영, 기계, 기업, 누군가, 독감, 동반자, 반대, 사회, 세계, 시대, 여론, 예측, 유튜브, 유포, 저장, 조사, 통계, 행간, 활용, K팝
30	사용된 총 단어 수
49	총 단어의 출현 빈도 수
5	총 분석대상 기사 수

2012년에는 ‘빅데이터’(61회)가 가장 많이 언급됐고 ‘분석’, ‘시대’, ‘선거’, ‘SNS’, ‘예측’, ‘활용’, ‘시장동향’, ‘개인성향’, ‘개인정보’, ‘통신업체’, ‘포털업체’, ‘마케팅’, ‘사회이슈’, ‘소비자’, ‘정보’, ‘미래’, ‘조사’, ‘혁명’, ‘IT 업체’가 상위 순위에 포함됐다. 2011년과 비교할 때, 눈에 띄는 단어는 빈도순위 4위를 기록한 ‘선거’이다. 이것은 2012년에 치러진 제18대 대통령 선거와 관련하여 빅데이터가 본격적으로 활용되기 시작했음을 짐작케 한다. 또한 2012년 미국 대통령 선거에서 재선에 성공한 오바마 대통

령의 빅데이터 활용 전략이 잇따라 소개된 것도 한 이유로 추론해볼 수 있다.

〈표 3〉 2012년 ‘빅데이터’ 관련 신문기사 제목에 나타나는 주요단어

사용 빈도	주요단어
61	빅데이터
22	분석
13	시대
12	선거
10	예측, 활용, SNS
8	개인성향, 시장동향
7	개인정보, 통신업체, 포털업체
6	마케팅, 사회이슈, 소비자, 정보
5	미래, 조사, 혁명, IT업체
4	기술, 신무기, 신용카드, 제휴, 화두
3	가치, 경쟁력, 공공부문, 기업, 날씨정보, 도태, 변화, 수집, 신경영, 유용, 평가
2	개방, 검색, 경쟁, 공략, 글로벌시장, 대응, 맞춤형, 명암, 불확실성시대, 빅브러더, 사업, 산업, 성공, 서비스, 세상, 수익성, 여론, 유출, 의료정보, 정보실패, 컴퓨터, 파악, 패러다임, 해법, 효과, 힘, 21세기
1	개인, 관리, 데이터연금술사, 데이터큐레이션, 목표, 발굴, 벤치, 성장, 소문, 소통, 스마트폰, 승부, 신뢰, 실시간, 앱, 오바마, 의미, 의사결정, 자유, 정보상품화, 정치, 진출, 창출, 추적, 큐레이슈머, 프로스포츠, 희망
90	사용된 총 단어 수
353	총 단어의 출현 빈도 수
113	총 분석대상 기사 수

새롭게 등장한 신조어도 주목할 만하다. 예컨대, ‘큐레이슈머’, ‘데이터연금술사’, ‘데이터큐레이션’ 등 빅데이터의 활용 능력을 강조하는 단어들 이 나타나면서 빅데이터에 대한 뉴스의 긍정적 해석을 엿볼 수 있다. ‘빅

데이터’를 미래의 새로운 가치로 제시하는 단어들도 같은 맥락이다. 여기에는 ‘신무기’, ‘신경영’, ‘혁명’, ‘가치’와 같은 단어들이 속한다.

그러나 무엇보다 빅데이터 관련 2012년 뉴스보도에서 두드러지는 특징은 다양한 활용 사례가 소개되고 있는 점이다. 이에 따라 ‘개인성향’, ‘시장동향’, ‘사회이슈’, ‘마케팅’, ‘공공부문’, ‘날씨정보’, ‘의료정보’ 등이 본격적으로 등장하고 있으며 기업이나 경제적 측면과 관련된 ‘사업’, ‘마케팅’, ‘통신업체’, ‘포털업체’, ‘IT업체’, ‘신용카드’, ‘기술’, ‘소비자’, ‘벤처’ 등의 단어들이 대거 나타나고 있다. 또한 ‘활용’, ‘혁명’, ‘유용’, ‘성장’, ‘경쟁력’, ‘힘’, ‘승부’, ‘성공’, ‘가치’, ‘효과’, ‘수익성’ 등 빅데이터를 미래의 새로운 가치로 자리매김하기 위해 긍정적인 프레이밍을 시도한 단어들이 대거 등장하는 반면 빅데이터 활용의 부정적 측면을 직접적으로 언급하는 ‘빅브러더’, ‘명암’, ‘유출’, ‘추적’ 등의 단어는 전체에서 1.9%에 그치며 매우 미미하게 다뤄지고 있음을 알 수 있다.

또 다른 특징은 “빅데이터, 욕망을 통계내라”(조선일보, 2012.12.10), “감정까지 헤아려야 빅데이터 분석 쓸모”(동아일보, 2012.2.3), “방문자 관심사 정밀분석”(동아일보, 2012.2.14), “고객 마음 알고 싶은가? 빅데이터 분석하라”(조선일보, 2012.9.24) 등의 기사 제목에서처럼 논쟁적 의미를 함축할 수 있는 ‘개인정보’ 대신 ‘욕망’, ‘감정’, ‘관심사’, ‘마음’ 등으로 개인의 정보와 관련된 의미를 순화시키고 있음을 발견할 수 있다.

2013년에도 ‘빅데이터’가 가장 높은 빈도수(73회)를 기록한 가운데 ‘분석’, ‘예측’, ‘의료정보’, ‘활용’, ‘SNS’, ‘시장동향’, ‘스마트폰’, ‘고객’, ‘개인성향’, ‘맞춤형’, ‘교통정보’, ‘국민건강’, ‘기술’, ‘세상’, ‘신용카드’, ‘예방’, ‘정보’, ‘창조경제’, ‘과약’, ‘건보공단’이 상대적으로 높은 빈도수를 나타냈다. 특이할만한 점은 ‘의료정보’, ‘국민건강’, ‘건보공단’, ‘질병지도’ 등 의료 분야와 관련된 단어들이 대거 등장하며 2013년 들어 의료계에서 빅데이터 활용이 빈번하게 논의되었음을 짐작할 수 있다.

〈표 4〉 2013년 ‘빅데이터’ 관련 신문기사 제목에 나타나는 주요단어

사용 빈도	주요단어
73	빅데이터
33	분석
15	예측, 의료정보
11	활용, SNS
10	스마트폰
9	시장동향
8	고객
7	교통정보, 맞춤형
6	국민건강, 개인성향, 세상, 정보
5	건보공단, 기술, 신용카드, 예방, 창조경제, 파악
4	공개, 기업, 날씨정보, 담배소송, 사회이슈, 서비스, 스포츠, 시대, 정부, 통신업체, IT, NSA
3	개발, 개인정보, 검찰, 결정, 경영, 구글, 글로벌시장, 기억, 도구, 도약, 도청, 디지털, 선거, 시장동향, 앱, 온라인, 이용자, 재난, 전문가, 증거, 컴퓨터, 한국
2	격차, 공공기관, 교육, 국민, 국정원, 글, 내비게이션, 능력, 두뇌, 마케팅, 미래, 베이비부머, 변화, 보호, 스노든, 승리, 시스템, 시장, 영화, 융합, 이용, 일상, 전통시장, 정보지도, 정책, 지역, 지원, 질병지도, 추적, 쿠팡, 트렌드, 해킹, 혁신, 확보, 효과, 흥행, IT업체
1	가상재화시장, 가치, 감시망, 개입, 검색, 공포, 국가, 국정운영, 권력, 권리, 규제, 기록, 네트워크, 노출, 데이터사이언티스트, 데이터화, 도움, 도태, 모바일라이프, 무용지물, 미래병원, 비밀, 빅브러더, 빅테인먼트, 사업, 사용자, 산물, 성공, 성패, 소셜시청률, 소셜화, 수집, 스마트기기, 승자, 시민감시, 식별, 신권버전스, 실시간, 연관성, 영향력, 오바마스타일, 위법, 위치정보, 위협, 유용, 의학, 잊혀질권리, 자동차업계, 자원, 전략, 정보기관, 주권, 주도, 초감시사회, 클라우드, 통계, 통제자, 패러다임, 프리즘, 해법, IT강국, LTE, N스크린, TV
156	사용된 총 단어 수
505	총 단어의 출현 빈도 수
247	총 분석대상 기사 수

또 전년도까지 등장하지 않았던 ‘교통정보’와 ‘창조경제’도 비중 있게 언급되고 있으며 ‘예방’이라는 단어도 빈도 순위 10위를 기록하고 있는데, 이는 의료 분야에서 공공서비스 측면이 강조되면서 사용하기 시작한 단어로 유추해 볼 수 있다. ‘고객’과 ‘맞춤형’이 빈번하게 사용된 것도 개인정보를 활용하여 맞춤형 서비스 마케팅이 강조되고 있음을 엿보게 한다.

그밖에도 ‘데이터 사이언티스트’, ‘소셜시청률’, ‘빅테인먼트’ 등의 신조어가 등장한 것이 눈에 띄며 ‘빅브러더’를 비롯해 ‘도청’, ‘감시망’, ‘통제자’, ‘초감시사회’, ‘위협’, ‘위법’, ‘시민감시’, ‘추적’, ‘해킹’ 등 부정적 단어가 폭넓게 언급되고 있는 것도 특징이다. 나아가 ‘국정원’, ‘검찰’, ‘NSA(미국 국가안보국)’, ‘권력’, ‘규제’, ‘잊혀질권리’ 등 논쟁적 단어도 이전보다 다양하게 사용되고 있다. 물론 부정적 단어나 논쟁적 단어의 비중은 전체에서 7.7%를 기록해 여전히 미약한 수준에 머물고 있지만 2012년도(1.9%)보다 확대됐다고 볼 수 있다.

이를 종합하면 <표 5>에서 볼 수 있듯이 보도시기에 따라 주요단어의 차이를 발견할 수 있다. 사용빈도가 2회 이상이면서 빈도 순위 10위안에 드는 단어들 가운데 2011년부터 2013년까지 공통적으로 등장하는 단어는 ‘빅데이터’를 제외했을 때 ‘분석’과 ‘SNS’, 두 단어뿐이다. 이는 빅데이터가 페이스북이나 트위터와 같은 SNS를 주요한 정보 출처로 삼고 있다는 점과 기술적인 분석을 통해 가공되는 정보라는 사실을 함축적으로 나타낸다고 볼 수 있다. 또 2012년부터 관련 기사가 급증한 것과 마찬가지로 빅데이터와 연관되어 사용된 단어 종류도 2012년부터 크게 확대되었다. 2011년에 사용된 주요단어들은 ‘빅데이터’의 근간이 되는 ‘개인정보’나 ‘SNS’, ‘검색’, ‘기록’ 등 정보통신기술 현황을 보여주는 기술이슈 구성 단어로 모아볼 수 있고 2012년의 경우, 활용이슈를 뒷받침하는 다양한 사례나 IT관련 업체와 마케팅 관련 단어들이 높은 빈도로 사용되었다. 반면 2013년에는 이전의 상업적 활용과 달리 의료나 교통과 같은 공공

서비스 이슈를 강조하는 단어들이 대체로 높은 사용빈도를 나타냈다.

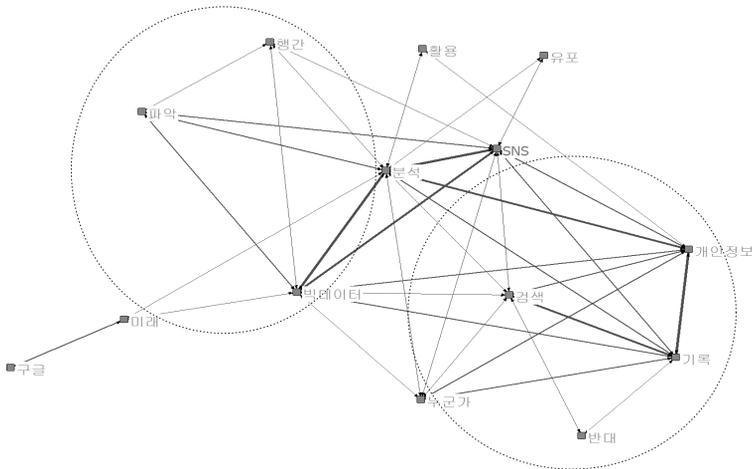
〈표 5〉 보도시기별 '빅데이터' 관련 신문기사 제목에 나타나는 주요단어 비교

	2011	2012	2013
분석기사 건수	5	113	247
총 단어 개수	30	90	156
최다빈도 단어	분석	빅데이터	빅데이터
상위빈도 단어	분석(6), 개인정보(4), 빅데이터(4), SNS(4), 검색(2), 구글(2), 기록(2), 미래(2), 파악(2)	빅데이터((61), 분석(22), 시대(13), 선거(12), 예측(10), 활용(10), SNS(10), 개인성향(8), 시장동향(8), 개인정보(7), 통신업체(7), 포털업체(7), 마케팅(6), 사회이슈(6), 소비자(6), 정보(6), 미래(5), 조사(5), 혁명(5), IT업체(5), 기술(4), 신무기(4), 신용카드(4), 제휴(4), 화두(4)	빅데이터(73), 분석(33), 예측(15), 의료정보(15), 활용(11), SNS(11), 스마트폰(10), 시장동향(9), 고객(8), 교통정보(7), 맞춤형(7), 국민건강(6), 개인성향(6), 세상(6), 정보(6), 건보공단(5), 기술(5), 신용카드(5), 예방(5), 창조경제(5), 파악(5)
단어집단 특성	기술이슈구성용어	상업적 활용이슈구성용어	공공부문이슈구성용어
상위빈도 내 공유단어	빅데이터, 분석, SNS		
총 단어 특성	· 정보통신 전문성 강조 (예: 경로, 분석, 통계)	· 신조어 등장 (예: 데이터 연금술사) · 트렌드 강조 (예: 신무기, 신경영) · 단어 대체 (예: 개인정보→마음)	· 신조어 등장 (예: 데이터 사이언티스트) · 서비스 강조 (예: 맞춤형, 예방) · 활용경고 단어 다양화 (예: 잊혀질권리)

2) 보도시기별 주요단어 의미연결망의 특성

(1) 2011년도 ‘빅데이터’ 관련 신문기사의 주요단어 의미연결망

〈그림 1〉은 2회 이상 사용된 상위 9개 단어들과 1회씩 언급된 단어가지만 의미구조를 파악하기 위하여 가치가 함유된 단어로 추정할 수 있는 ‘활용’, ‘행간’, ‘누군가’, ‘유포’, ‘반대’를 추가로 선정한 총 14개 단어의 의미지도이다. 다른 단어와 연결되는 관계 링크수(연결노드수)와 연결정도 중심성으로 살펴볼 때 ‘분석’, ‘빅데이터’, ‘SNS’, ‘개인정보’, ‘검색’, ‘기록’이 상대적으로 높게 나타나며 연결망 핵심부 위치에 자리하고 있다. 사용빈도에서는 ‘검색’이 ‘구글’과 동일한 2회를 기록했지만 다른 단어와 동시 출현하는 관계에서는 비중이 상이함을 보여준다. 이는 곧 의미연결망에서 ‘검색’이 ‘구글’보다 더 많은 단어나 이슈와 연결되고 있음을 시사한다.



〈그림 1〉 2011년 ‘빅데이터’ 관련 주요단어 연결망

의미연결망은 각 단어의 위치와 선의 두께로 연결강도(tie strength)

를 보여주는데, '분석'을 중심으로 '빅데이터', 'SNS', '개인정보'가 가장 강하게 연결되어 있음을 알 수 있다. 또 '분석' 왼쪽으로 '빅데이터', '행간', '파악' 등이 작은 관계망을 형성한 뒤 다소 거리감은 있지만 '미래'와 연결되어 있고 '분석' 오른쪽으로는 '개인정보', '누군가', '기록' 등이 또 다른 관계망을 만들고 있다. 중앙 위쪽을 살펴보면 '활용'과 '유포'가 각각 독립적으로 외곽 이슈를 구성하고 있음을 보여준다. '누군가'는 '개인정보'나 '기록'과 비교적 밀접하게 연결되면서 '빅데이터'와 관련한 개인정보 침해에 대해 우려의 시각이 공존하고 있음을 뜻한다. 그러나 단어와 단어를 중개하는 매개중심성 값이 0이라는 것은 의미의 흐름을 통제할 수 있는 핵심적인 단어로 사용되고 있지 않다는 것을 말해준다. <표6>은 UCINET으로 분석한 단어별 연결링크수와 중심성 값이다. 링크수가 같아도 의미를 주도하는 활동성과 의미를 중개하는 통제력이 반드시 일치하는 것은 아니다.

<표 6> 2011년 주요단어 연결지수와 중심성 값

단어	연결링크수	단어	연결정도 중심성	단어	매개중심성
분석	11	분석	40.385	분석	25.333
빅데이터	9	빅데이터	32.692	미래	12.000
SNS	9	SNS	32.692	빅데이터	11.333
기록	7	개인정보	30.769	SNS	7.333
검색	7	기록	30.769	기록	5.500
개인정보	7	검색	19.231	검색	5.500
누군가	6	누군가	15.385	개인정보	3.000
파악	4	파악	13.462	누군가	0.000
행간	4	미래	7.692	파악	0.000
미래	3	행간	7.692	행간	0.000
유포	2	활용	3.846	유포	0.000

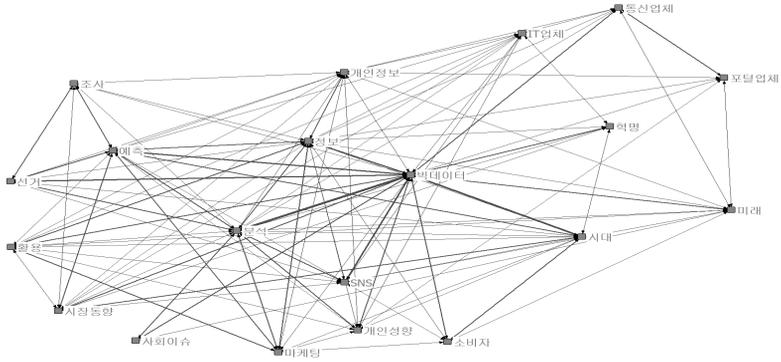
활용	2	구글	3.846	활용	0.000
반대	2	유포	3.846	반대	0.000
구글	1	반대	3.846	구글	0.000

*연결정도 중심성, 매개중심성: UCINET6 Freeman’s centrality 분석에서 구한 값.

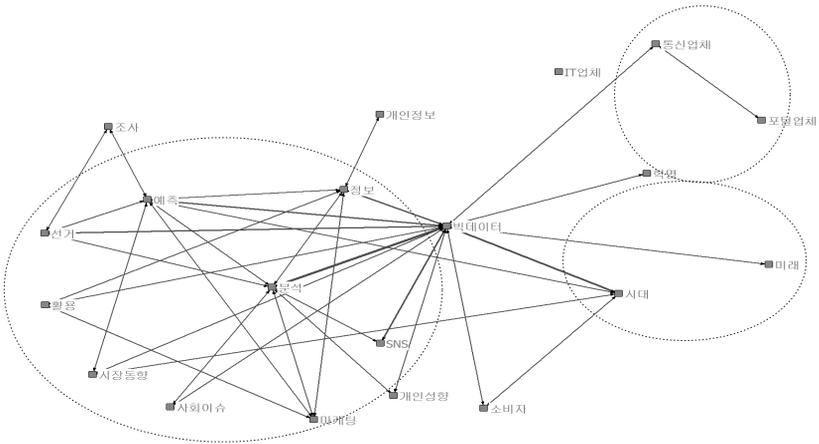
(2) 2012년도 ‘빅데이터’ 관련 신문기사의 주요단어 의미연결망

빅데이터와 관련한 2012년 신문기사의 주요단어 20개의 의미연결망은 좀 더 촘촘하고 복잡해진 격자형 구조를 형성하고 있다. 2011년과 달리 ‘빅데이터’가 의미지도의 중심 위치로 부상, 가장 광범위한 관계를 맺고 있음을 보여준다. ‘빅데이터’는 ‘분석’과 가장 두텁게 연결되어 있으며 이어서 ‘선거’, ‘예측’, ‘정보’, ‘시대’, ‘SNS’와 활발한 상호작용을 하는 하위집단(sub-group)으로 구별된다.

주요단어들의 연결강도가 4 이상이 되는 관계만 표시하고 있는 <그림 3>은 단어들의 의미구조를 더 요약적으로 보여주고 있는데, ‘빅데이터’가 크게 3가지 방향에서 범주화되고 있음을 알 수 있다. 첫 번째는 ‘정보’, ‘개인성향’, ‘SNS’를 바탕으로 ‘선거’, ‘시장동향’, ‘사회이슈’, ‘마케팅’ 부문에서 ‘분석’, ‘예측’ 등과 연결되면서 2011년보다 다양한 분야에서 활용되고 있는 사례를 보여주고 있다. 두 번째로는 ‘미래’, ‘혁명’이라는 미래형 가치 제시 단어들과 연결되어 ‘빅데이터’의 의미화 방향을 암시하고 있다. 그리고 세 번째로는 ‘통신업체’와 연결되어 있는데 ‘통신업체’는 다시 ‘포털업체’와 이어지면서 빅데이터를 둘러싼 정보통신 산업계의 변화를 추정해볼 수 있게 한다. 상위 20개 단어 안에는 부정적인 단어가 포함되어 있지 않으며 논쟁적 의미를 함축할 수 있는 ‘개인정보’의 경우에도 ‘정보’를 매개로 ‘마케팅’, ‘예측’, ‘분석’, ‘빅데이터’로만 연결됨으로써 긍정적인 ‘활용’ 측면만이 선택적으로 강조되고 있다.



〈그림 2〉 2012년 '빅데이터' 관련 주요단어 연결망



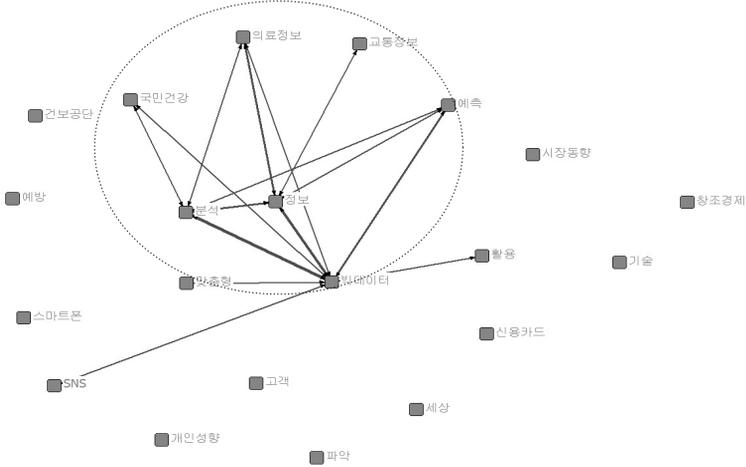
〈그림 3〉 연결강도 4이상인 2012년 '빅데이터' 관련 주요단어 연결망

또 의미연결망에서 각 단어가 다른 단어와 연결되는 링크수와 중심성 값을 비교한 <표 7>에 따르면, 단어빈도 순위에 비례하여 링크수나 중심성 값이 높은 것은 아니라는 사실을 재확인할 수 있으며 연결망 중심에 위치한 단어들이 높은 값을 나타내는 가운데 핵심 이슈를 구성하고 있음을 알 수 있다.

<표 7> 2012년 주요단어 연결지수와 중심성 값

단어	연결링크수	단어	연결정도 중심성	단어	매개중심성
빅데이터	19	빅데이터	30.100	빅데이터	57.524
정보	18	분석	14.883	분석	27.760
분석	17	정보	12.207	정보	26.161
개인정보	14	예측	9.532	개인정보	14.555
예측	13	시대	7.860	SNS	8.421
시대	12	선거	7.525	미래	7.606
활용	12	SNS	6.856	예측	6.662
시장동향	11	활용	6.355	시대	5.308
IT업체	11	마케팅	5.686	IT업체	4.822
SNS	11	시장동향	5.518	개인성향	3.913
개인성향	10	개인정보	5.184	포털업체	3.440
미래	10	개인성향	4.515	통신업체	2.848
마케팅	9	미래	4.181	활용	2.758
선거	8	조사	3.679	시장동향	2.614
통신업체	8	통신업체	3.177	선거	2.524
소비자	7	소비자	3.177	소비자	1.213
조사	7	포털업체	3.010	마케팅	1.021
포털업체	7	IT업체	2.843	혁명	0.944
혁명	6	사회이슈	2.341	조사	0.322
사회이슈	3	혁명	1.672	사회이슈	0.000

층형’과 밀접하게 연결되면서 의미의 흐름이 시장예측보다는 고객 서비스 측면을 강조하는 방향으로 바뀌었음을 알 수 있다.



〈그림 5〉 연결 강도 4이상인 2013년 ‘빅데이터’ 관련 주요단어 연결망

또한 2013년도에는 그 이전 어느 해보다 ‘빅데이터’의 문제적 측면을 꼬집는 부정적 단어나 논쟁적 단어가 다양하게 나타났지만 의미연결망에서는 찾아볼 수 없다. ‘빅데이터’와 지속적으로 연결되어 전략적이고 핵심적인 이슈를 구성하고 있지 않기 때문이다. 논쟁적 의미를 함축할 수 있는 ‘개인정보’ 또한 상위빈도 단어의 의미연결망에 포함되지 못했다. 새롭게 등장한 단어인 ‘창조경제’도 ‘기술’을 매개하지 않으면 의미의 흐름이 단절되면서 특별한 이슈를 만들지 못하고 있다.

〈표 8〉 2013년 주요단어 연결지수와 중심성 값

단어	연결링크수	단어	연결정도 중심성	단어	매개중심성
빅데이터	18	빅데이터	21.167	빅데이터	68.286
분석	16	정보	13.833	분석	36.914
정보	13	분석	13.167	기술	19.000
맞춤형	9	의료정보	6.333	정보	13.233
고객	8	예측	6.000	활용	6.843
예측	7	국민건강	3.667	고객	4.626
활용	7	맞춤형	3.333	국민건강	4.152
국민건강	7	활용	3.167	맞춤형	4.019
개인성향	5	고객	2.833	예측	2.000
스마트폰	5	교통정보	2.333	신용카드	1.117
신용카드	5	SNS	1.833	SNS	1.033
예방	5	예방	1.833	시장동향	0.676
의료정보	5	개인성향	1.667	개인성향	0.650
SNS	5	스마트폰	1.333	의료정보	0.250
교통정보	4	파악	1.167	파악	0.200
시장동향	4	신용카드	1.167	건보공단	0.000
파악	4	세상	1.000	교통정보	0.000
기술	3	시장동향	0.833	세상	0.000
세상	3	기술	0.500	스마트폰	0.000
건보공단	2	건보공단	0.333	예방	0.000
창조경제	1	창조경제	0.167	창조경제	0.000

3) 매체성향별 ‘빅데이터’ 관련 신문기사에 나타난 주요단어와 의미연결망

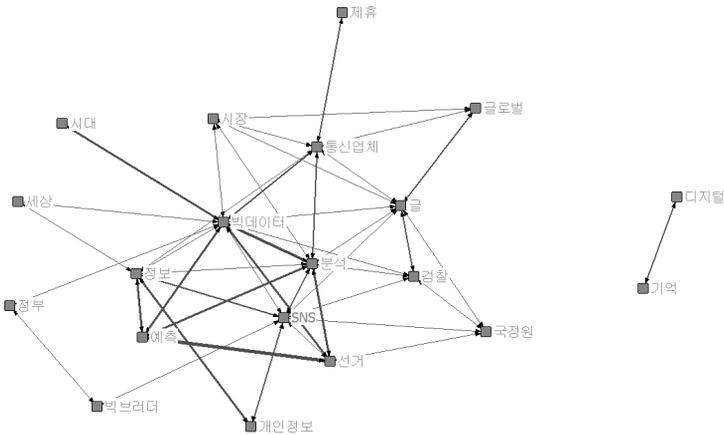
네 신문 모두에서 빈도순으로 상위 20개 단어에 공통적으로 나타나는 단어는 ‘빅데이터’, ‘분석’, ‘SNS’, ‘예측’, ‘선거’, ‘개인정보’, ‘시대’, ‘정보’이다. 이는 빅데이터가 SNS 등의 개인 정보를 바탕으로 분석되고 예측

되며 새로운 시대적 화두임을 드러내는 데에는 신문의 이념적 성향과 관계없이 동일함을 보여준다. 또한 2012년에 치러진 대통령 선거가 신문 보도에서도 차지하는 비중이 컸던 만큼 ‘선거’가 동일하게 주요단어로 등장하고 있음을 추측할 수 있다.

〈표 9〉 매체성향별 ‘빅데이터’ 관련 기사에 나타나는 주요단어

경향신문/한겨레신문			동아일보/조선일보		
단어	빈도	%	단어	빈도	%
빅데이터	20	26.7	빅데이터	118	31.6
SNS	6	8	분석	55	14.7
분석	5	6.7	예측	20	5.3
예측	5	6.7	SNS	19	5.1
통신업체	5	6.7	활용	17	4.5
선거	4	5.3	시대	16	4.3
세상	3	4	의료정보	16	4.3
활용	3	4	개인성향	13	3.5
개인정보	2	2.7	선거	11	2.9
검찰	2	2.7	개인정보	10	2.7
국정원	2	2.7	시장동향	10	2.7
글	2	2.7	기술	8	2.1
글로벌	2	2.7	마케팅	8	2.1
기억	2	2.7	맞춤형	8	2.1
디지털	2	2.7	미래	8	2.1
빅브러더	2	2.7	파악	8	2.1
시대	2	2.7	기업	8	2.1
시장	2	2.7	소비자	7	1.9
정보	2	2.7	신용카드	7	1.9
제휴	2	2.7	정보	7	1.9
	75	100		374	100

그러나 이들 8개의 공통 단어를 제외하면 매체성향에 따라 차이를 발견할 수 있다. 우선 경향신문과 한겨레신문에는 정보통신 산업계 변화를 시사하는 단어들(‘통신업체’, ‘글로벌’, ‘디지털’, ‘세상’, ‘시대’, ‘체휴’)과 부정적 또는 논쟁적 의미를 지닌 단어들(‘검찰’, ‘국정원’, ‘글’, ‘빅브러더’)이 크게 두 범주를 형성하고 있다. 반면 동아일보와 조선일보에 등장하는 ‘활용’, ‘의료정보’, ‘개인성향’, ‘시장동향’, ‘기술’, ‘마케팅’, ‘맞춤형’, ‘미래’, ‘과약’, ‘기업’, ‘소비자’, ‘신용카드’ 등은 모두 단일한 범주라고 볼 수 있는 활용분야와 관계돼 있다. 특히 ‘의료정보’를 제외하면 주로 기업의 새로운 마케팅 방식에 연결돼 있는 것이 특징으로 경향신문·한겨레신문과 대조를 이룬다.



〈그림 6〉 경향신문과 한겨레신문에 나타나는 주요단어 연결망

의미연결망 지도를 보면 경향신문과 한겨레신문이 ‘빅데이터’와 연관지어 무엇을 이슈화하고 있는지가 더 분명해진다. 경향신문·한겨레신문의 의미연결망에서는 ‘빅데이터’와 ‘분석’, 그리고 ‘선거’와 ‘예측’을 잇는 선이

두텁게 표시되며 두 단어집단이 가장 강한 연결을 보이는 관계(strong-ties)로 나타났고 동아일보·조선일보의 의미연결망에서는 ‘빅데이터’, ‘분석’, ‘예측’, ‘SNS’, ‘활용’, ‘시대’, ‘의료정보’, ‘정보’의 8개 단어로 구성된 관계가 가장 긴밀하게 연결되어 있음을 보여준다.

‘빅데이터’의 관계성 범위를 살펴보면 경향신문·한겨레신문에서는 ‘분석’과 동시출현 빈도가 가장 높고 이어서 ‘선거’, ‘시대’, ‘예측’, ‘통신업체’, ‘SNS’, ‘검찰’, ‘글’, ‘세상’, ‘시장’, ‘정보’, ‘정부’ 등 상위 20개 단어 중 12개 단어와만 직접적인 연결 구조를 드러낸다. ‘빅브러더’의 경우에는 ‘빅데이터’와 동시에 사용되고 있지는 않으며 연결망 외곽에서 ‘SNS’, ‘정부’와 각각 연결강도 1의 크기로 상호작용하고 있다.

반면 동아일보·조선일보의 경우, ‘빅데이터’는 상위 20위 안에 드는 모든 단어와 연결되어 더 큰 관계성의 범위를 보여준다. 이 중 눈에 띄는 점은 연결강도 5의 ‘개인정보’보다는 ‘개인성향’이 연결강도 10을 나타내며 ‘빅데이터’와 ‘개인성향’, 두 단어 사이의 교류빈도가 더 높다는 것을 알 수 있다. 또 연결강도가 10이 넘는 상위순위에 ‘분석(50)’, ‘정보(31)’, ‘예측(22)’, ‘SNS(20)’, ‘시대(19)’, ‘활용(15)’, ‘의료정보(12)’, ‘개인성향(10)’, ‘선거(10)’, ‘시장동향(10)’이 포진돼 있어 부정적이거나 논쟁적 단어는 핵심적인 이슈 구성에서 제외되어 있음을 알 수 있다.

중심성 값을 비교한 <표 10>을 살펴보면, 연결정도 중심성 측면에서는 경향신문·한겨레신문의 경우, ‘빅데이터’, ‘분석’, ‘예측’, ‘선거’, ‘SNS’, ‘글’, ‘검찰’ 등이 더 많은 단어들과 연결되어 핵심적으로 사용되고 있고 동아일보·조선일보의 경우에는 ‘빅데이터’, ‘분석’, ‘정보’, ‘예측’, ‘활용’, ‘마케팅’, ‘의료정보’ 등이 다른 단어와의 연결수가 많은 단어로 나타났다. 매개중심성을 살펴보면 경향신문·한겨레신문에서는 ‘빅데이터’, ‘SNS’, ‘통신업체’, ‘글’, ‘정보’가 이슈를 아우르는 전략적 역할을 주도적으로 하고 있으며 동아일보·조선일보의 경우에는 ‘빅데이터’, ‘분석’, ‘활용’, ‘정

SNS	12.281	활용	6.526	정보	12.276	시대	6.794
통신업체	8.772	SNS	5.895	분석	9.404	SNS	5.133
글	8.772	시대	5.789	선거	2.600	개인정보	3.498
개인정보	5.263	마케팅	4.526	정부	2.183	기업	2.882
검찰	5.263	의료정보	4.316	시장	2.119	미래	2.096
시장	4.386	개인정보	4.211	검찰	1.150	개인성향	2.066

5. 결론

본 연구는 빅데이터 현상에 대한 신문의 뉴스보도가 어떻게 이뤄지고 있는가를 살펴보기 위해 의미연결망 분석 기법을 활용하여 관련 기사의 제목(부제 포함)을 분석하였다. 특히 언론의 뉴스보도는 현실 그 자체가 아니라 언론이 재구성하는 현실이라는 전제에서 일반 공중은 물론 사회적 인식의 형성에도 영향을 끼칠 수 있는 빅데이터 관련 신문기사의 언어적 상징과 의미구조를 찾아보고자 했다.

이를 위해 경향신문, 동아일보, 조선일보, 한겨레신문을 분석대상으로 한 결과에 따르면, 첫째 ‘빅데이터’라는 용어는 2011년 신문기사에 처음 등장한 이래 2011년 총 기사건수 5건에서 2012년 113건, 2013년 247건으로 관련 기사가 폭증하며 언론의 관심 또한 지속적으로 커져 왔다.

둘째, 시기별로 신문기사에서 나타나는 주요단어를 살펴보면 2011년에는 ‘빅데이터’를 비롯해 ‘통계’, ‘검색’, ‘저장’, ‘경로’와 같은 정보통신기술 용어들과 활용과 관련된 단어들(‘구글’, ‘여론’, ‘예측’, ‘기업’, ‘SNS’)이 다수를 차지하고 있지만 빅데이터 현상의 문제점을 경고하는 ‘누군가’, ‘반대’, ‘유포’와 같은 부정적 가치를 포함한 논쟁적 단어들도 함께 등장하고 있다. 그러나 빅데이터를 새로운 정보통신 기술로서 파악, 그 특

징적 성격을 보도하는 점에서 단어 선택을 통해 특정 방향으로 프레이밍 하려는 의도는 나타나지 않았다.

하지만 2012년 빅데이터 관련 신문기사에는 긍정적으로 추론할 수 있는 단어들이 대거 등장한다. 빅데이터의 실용적 활용 사례를 예시하는 ‘선거’, ‘마케팅’, ‘사회이슈’, ‘시장동향’ 등의 단어는 물론 ‘혁명’, ‘미래’, ‘시대’라는 단어 사용으로 새로운 미래 사회를 제시하고 있고 ‘크레이슈머’, ‘데이터 연금술사’, ‘데이터큐레이션’과 같은 신조어를 통해 빅데이터 분석 능력을 강조하고 있다. ‘개인정보’라는 단어 대신 ‘욕망’, ‘관심사’, ‘감정’, ‘마음’ 등의 단어로 대체하여 사용함으로써 논쟁적 의미를 피하려는 전략도 엿볼 수 있다. 또한 ‘빅브러더’ 등의 부정적 단어도 눈에 띄지만 전체 단어에서 차지하는 비중은 1.9%에 그치고 있어 주요 이슈를 형성하지는 못했다.

한국과 미국의 대선이 치러진 2012년에는 ‘선거’가 활용분야로 가장 많이 언급된 반면 2013년으로 넘어가면 ‘의료정보’, ‘국민건강’, ‘질병지도’, ‘교통정보’ 등의 단어가 상위 빈도순위를 차지하면서 의료와 교통 분야와 같은 공공 부문에서 적극적으로 활용되고 있음을 짐작하게 했다. 또 ‘분석’이나 ‘예측’ 이외에도 ‘맞춤형’이나 ‘예방’과 같이 이용자 또는 국민에 대한 서비스를 뜻하는 단어를 반복 사용함으로써 빅데이터의 이점을 강조하고 있는 것도 발견할 수 있다. ‘데이터 사이언티스트’나 ‘소셜시청률’과 같이 빅데이터 분석과 관련된 전문가나 새로운 영역을 뜻하는 신조어들이 계속해서 등장하고 있는 것도 눈에 띈다. 2013년에는 부정적이거나 논쟁적 단어가 그 어느 때보다 많이 등장한 것도 특징이다. 예컨대 ‘빅브러더’, ‘초감시사회’, ‘위협’, ‘위법’, ‘시민감시’, ‘감시망’, ‘통제자’, ‘국정원’, ‘검찰’, ‘NSA(미국 국가안보국)’, ‘권력’ 등으로 단어의 폭이 넓어졌다. 그러나 부정적 단어들은 여전히 적은 비중(7.7%)을 차지하고 있어 중요하게 다뤄지고 있다고 보기는 어렵다. 이러한 단어들이 등장한

데에는 여러 이유가 있을 수 있지만 2013년에 불거진 미국 정보기관의 감청 실체 폭로 사건과 국정원 트위터 선거개입 사건에 대한 연이은 보도를 한 원인으로 유추해볼 수 있다. 또한 부정적 단어들은 주로 경향신문·한겨레신문에서 빅데이터와 관련한 외곽 이슈를 구성하는 단어로 사용되고 있지만 빅데이터 현상의 문제점을 다루는 핵심적인 논제로 부상하지는 못했다.

셋째, 주요단어들의 의미 구조를 보여주는 의미연결망 분석결과에 따르면 2011년에는 ‘분석’이 가장 높은 연결정도 중심성을 나타냈다. 또 이를 둘러싸고 빅데이터 분석을 통한 미래가치 제시 측면이 주요 의제를 형성했고 누군가 개인정보와 검색기록을 분석하는 데 따른 부정적 의제가 주변부 이슈로 자리하고 있음을 확인할 수 있었다. 2012년부터는 ‘빅데이터’가 가장 높은 연결정도 중심성과 매개중심성을 기록하며 연결망 핵심 위치에서 영향력을 드러내고 관계성 범위도 점차 확대하고 있음을 알 수 있다. 그러나 의제는 단일하게 활용가치로 집중되어 나타났다. 다만 활용의 분야로서 사살예보와 같은 ‘사회이슈’, 추가예측과 같은 ‘시장동향’, 고객의 관심사를 분석하여 마케팅에 활용하는 ‘개인성향’ 파악이 주요 이슈를 구성하고 있으며 관련 업체인 IT기업과 포털업체, 통신업체가 연결망 주변부에서 빅데이터 이슈를 지원하고 있다. 2013년에도 활용에 치우친 단일 의제가 형성된 가운데 세부 분야로서 ‘의료정보’, ‘교통정보’, ‘국민건강’이 새롭게 대두되었으며 여전히 ‘스마트폰’과 ‘SNS’를 통한 정보 활용가치가 강조되고 있음을 보여준다. 2012년에는 ‘신문기’나 ‘화두’와 같은 단어가 주변부에 위치한 상태에서 빅데이터의 중요성을 암시하고 있다면 2013년에는 새로운 패러다임을 제시하는 단어보다는 ‘기술’, ‘신용카드’와 같은 구체적인 분석방법이나 대상이 의미연결망에 나타나고 있다. ‘빅데이터’와 동시 출현하는 관계단어의 범위도 해마다 확대되고 있는 것이 특징이지만 단어들의 의미는 긍정적인 방향으로 치우쳐

있다.

넷째, 경향신문·한겨레신문 그리고 동아일보·조선일보로 나누어 매체의 성향에 따른 의미연결망 차이를 살펴본 결과에 따르면, ‘빅데이터’가 핵심부 위치에서 가장 큰 영향력을 발휘하고 있는 점은 동일하다. 그러나 경향신문·한겨레신문은 ‘빅브러더’와 같은 부정적 이슈를 주변부에 위치시키고 있는 반면 동아일보·조선일보의 경우에는 의미연결망이 더 복잡하게 펼쳐지지만 단어들이 형성하는 이슈의 성격은 모두 활용가치로 모아져 있다. 또 경향신문·한겨레신문은 활용사례나 문제점 제기 모두에서 ‘선거’로 의미가 편중되어 빅데이터의 관계성 범주가 작게 나타나는 반면 동아일보·조선일보에서는 관계성이 폭넓게 드러나는 것도 다른 점이다.

결국 빅데이터 관련 신문기사의 의미구조는 2011년 중립적 성격에서 2012년부터 데이터의 사회적 활용가치를 강조하며 긍정적 방향으로 의미화하고 있음을 알 수 있다. 대부분의 보도가 실용적 차원의 활용가치에 집중되어 있어 미래사회를 향한 경쟁을 촉구하고 그 결과로서 예측되는 장밋빛 정보사회를 묘사하고 있다고 추론해볼 수 있다. 진보성향의 매체인 경향신문·한겨레신문에서 빅데이터가 초래할 수 있는 부정적 측면을 제시하고 있기는 하지만 핵심적 이슈로 부각시키지는 못하는 점에서 매체성향에 따른 의미구조의 차이가 두드러지지는 않았다. 또한 활용사례로서 기업의 마케팅이나 관련 업체들의 상업적 이슈만이 아니라 국민건강이나 교통정책과 같은 공공부문에서도 빈번하게 논의됨으로써 빅데이터 사회를 향한 기업뿐 아니라 국가적 열망을 읽게 했다.

이러한 연구결과를 토대로 몇 가지 시사점을 도출해본다면, 우선 빅데이터는 결코 가치중립적이지 않다는 것이 언론의 보도에서도 확인된다는 사실이다. 언론은 빅데이터의 이점만을 선택, 부각시키는 긍정적 프레임िंग을 시도하고 있지만 여전히 ‘빅브러더’와 같은 부정적인 단어들이 등장

하고 있는 점은 빅데이터의 이면에 해당하는 감시사회의 문제를 인정하는 보도라고 할 수 있다. 그러나 언론은 누가 왜 무엇을 위해 욕망을 통제하는가에 대한 논의보다는 ‘욕망’을 ‘통계’냄으로써 드러나는 데이터의 유용한 잠재력에 더 주목하고 있다. 즉, 빅데이터라는 새로운 현상에 대한 언론보도가 긍정적인 방향으로 의미의 흐름을 통제하고 있다고 해석해볼 수 있다. 언론 보도가 일반 공중의 인식과 사회적 인식 형성에 미치는 영향을 고려하면 우리 사회가 빅데이터 현상에 어떻게 접근하고 있는가를 간접적으로 보여주는 셈이다.

두 번째로는 의료나 교통 등의 공공부문이든 맞춤형 고객 마케팅 등의 기업 차원이든 빅데이터 분석을 토대로 이뤄지는 활용 사례들이 개별 상황과 관계없이 일반화되어 보도됨으로써 데이터에 대한 맹신을 부추기고 있는 점이다. 이는 곧 빅데이터를 기술적 패러다임이 아니라 사회적 패러다임의 문제로 전환시켜 활용의 필요성과 정당성만을 강조하는 결과로 이어지게 할 수 있다.

세 번째로는 빅데이터 관련 단어들에서 발견되는 활용의 주체 대다수가 관련 기업들인 것은 빅데이터 현상에 대한 논의가 시장화로 집중되어 있음을 함축한다. 그 결과, 빅데이터를 ‘21세기 원유’로 기정사실화하여 상업적 열망을 가속화시키고 이로 인해 야기될 수 있는 정보 접근권과 통제권의 문제, 새로운 디지털 정보사회의 격차 등의 논의는 줄어들 수밖에 없다. 다시 말해 언론의 논의가 경제적 시장화로 확대될수록 정치나 사회 영역에서 이뤄질 수 있는 논의의 장은 상대적으로 축소될 여지가 있는 것이다.

본 연구는 새로운 기술현상이자 이슈로서 사회 각 분야에서 일고 있는 빅데이터 열풍에 대해 신문들이 어떻게 보도하고 있는지를 실증적으로 분석했다는 점에서 일차적 의의가 있다. 특히 정보통신기술의 발전으로 나타난 새로운 현상이 기술 부문은 물론 사회인식의 변화에 끼칠 수 있

는 방향성을 검증했다는 데에서도 연구의 함의를 찾을 수 있다. 또한 과학적인 의미연결망 분석 방법을 통해 신문기사에서 드러나는 의미의 흐름을 객관적으로 계량화하고 미디어의 선택과 통제 전략을 찾아낸 연구 방법도 보도 프레임 연구의 지형을 확장하는 데 기여했다고 생각된다. 그러나 4개 종합일간지 분석으로 한정된 것과 관련기사의 제목만을 분석함으로써 보다 다양한 논의로 이어지지 못했으며 나아가 본 연구 결과만으로 언론의 보도 특성을 일반화할 수는 없다는 점에서 한계가 있다. 또한 가능하면 연구자의 주관적 해석의 여지를 없애기 위해 의미연결망 분석 방법을 활용했으나, 이 분석만으로 미디어가 재구성하는 의미체계를 완벽하게 파악할 수는 없었으며 질적 해석이 수반되는 점에서 연구자의 주관적 해석을 완전히 배제시켰다고 할 수는 없다. 향후 연구에서는 뉴스미디어가 구성하는 프레임을 명확히 추출하기 위해 기사제목 외에도 기사내용, 기사주제, 기사작성자, 정보원 등을 좀 더 세밀하게 분석하고 실제 공중의 인식과 비교하는 논의를 통해 우리 사회의 인식이 어떻게 변화하고 있는지를 입체적으로 파악하는 것이 필요할 것이다.

■ 참고 문헌 ■

- 권상희 (2005). 인터넷 뉴스프레임. 『한국언론정보학보』, 30호, 35~87.
- 김만재 · 전방욱 (2012). 언어네트워크 분석 기법을 활용한 인간배아복제 신문보도 분석. 『한국생명윤리학회』, 13권 2호, 19~34.
- 김상배 (2011). 네트워크로 보는 중견국 외교전략: 구조적 공백과 위치권력 이론의 원용. 『국제정치논총』, 51권 3호, 51~77.
- 김수정 · 조은희 (2005). 생명과학에 대한 한국과 미국의 뉴스프레임 비교연구. 『한국언론학보』, 49권 6호, 109~139.
- 김예란 (2013). 빅데이터의 문화론적 비판. 『커뮤니케이션 이론』, 9권 3호, 166~203.
- 김원용 · 이동훈 (2005). 언론보도의 프레임 유형화 연구: 국내원자력 관련 신문보도를 중심으로. 『한국언론학보』, 49권 6호, 166~197.
- 김유신 · 김남규 · 정승렬 (2012). 뉴스와 주가: 빅데이터 감성분석을 통한 지능형 투자자의사결정모형. 『지능정보연구』, 18권 2호, 143~156.
- 김형곤 (2011). 언론의 소셜미디어 담론의 주체분석. 『한국콘텐츠학회논문지』, 11권 6호, 242~251.
- 남인용 · 박한우 (2007). 대권 예비후보자 관련 신문기사의 네트워크 분석과 홍보전략. 『한국정당학회보』, 6권 1호, 79~107.
- 박성희 (2009). 제17대 대통령 후보 합동 토론 언어네트워크 분석. 『한국언론정보학보』, 45호, 220~254.
- 박은선 · 이광형 · 김찬석 (2012). 과학기술 실패 보도 프레임 연구. 『한국언론학보』, 56권 3호, 213~237.
- 박지영 · 김태호 · 박한우 (2013). 의미연결망 분석을 통한 셀러브리티의 SNS 메시지 탐구. 『방송통신연구』, 82호, 36~74.
- 박한우 · Leydesdorff (2004). 한국어의 내용분석을 위한 KrKwic 프로그램의 이해와 적용. 『Journal of The Korean Data Analysis Society』, 6권 5호, 1377~1388.
- 백옥인 (2013). 빅데이터의 형성과 전유체제 비판. 『동향과 전망』, 87호, 304~331.
- 송해룡 · 김원재 · 조항민 (2005). 과학기술 위험보도에 관한 수용자 인식 연구. 『한국언론학보』, 49권 3호, 105~128.

- 심홍진 · 김용찬 · 손혜영 · 임지영 (2011). 언어 네트워크 분석을 통한 스마트폰과 소셜미디어 이용자의 미디어 이용행태에 관한 탐색적 연구. 『한국방송학보』, 25권 4호, 82~138.
- 윤홍근 (2013). 문화산업에서 빅데이터의 활용방안에 관한 연구. 『글로벌 문화콘텐츠』, 10호, 157~179.
- 이광석 (2013). 지배양식의 국면변화와 빅데이터 감시의 형성. 『사이버커뮤니케이션학보』, 30권 2호, 191~231.
- 이동환 · 박정찬 · 유찬곤 · 윤호상 (2013). 빅데이터 기반의 실시간 네트워크 트래픽 분석 플랫폼 설계. 『정보보호학회논문지』, 23권 4호, 721~728.
- 이수범 · 강연곤 (2013). 국내 일간지의 트위터 이슈에 관한 보도 프레임 분석. 『한국언론학보』, 57권 1호, 28~53.
- 이재현 (2013). 빅데이터와 사회과학. 『커뮤니케이션 이론』, 9권 3호, 127~165.
- 장하용 (2001). 언론보도와 비평의 구조: 신문보도의 비평에 대한 네트워크 분석. 『한국언론정보학보』, 16권, 108~133.
- 정재철 (2004). 한국 신문의 유전자 연구 프레임 비교분석. 『한국언론정보학보』, 25호, 135~162.
- 조영임 (2013). 빅데이터의 이해와 주요 이슈들. 『한국지역정보화학회지』, 16권 3호, 43~65.
- 주재원 · 나보라 (2009). 올드미디어는 뉴미디어를 어떻게 재현하는가. 『언론과 사회』, 17권 2호, 2~48.
- LG경제연구원 (2012). 빅데이터 시대의 한국.
- SERI (2012). 빅데이터: 산업 지각변동의 진원.
- de Vreese, C. H. (2004). The effects of frames in political television news on issue interpretation and frame salience. *Journalism & Mass Communication*, 81(1), 36~52.
- Freeman, L. C. (1979). Centrality in Social Networks: Conceptual Clarification. *Social Networks*, 1(3), 215~239.
- Gartner (2012). The importance of Big Data: A definition.
- Gitlin, T. (1980). *The whole world is watching*. Berdeley: University of California Press.

- Gordon, E. & de Souza e Silva, A. (2011). *Net Locality*. Oxford: Wiley Blackwell.
- Entman, R.M. (1991). Framing U.S. Coverage of International News: Contrasts in Narratives of the KAL and Iran Air Incidents. *Journal of Communication*, 41(4), 6~27.
- Entman, R.M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication*, 43, 51~58.
- IDC (2012). Worldwide Big data Technology and Services 2012~2015 Forecast.
- Iyengar, S. (1991). *Is Anyone Responsible: How Television Frames Political Issues*. Chicago: University of Chicago Press.
- Kitzinger, J., & Reilly, J. (1997). The rise and fall of risk reporting, *The European Journal of Communication*, 12(3), 319~350.
- McKinsey (2011). Big Data: The next Frontier for Innovation, Competition, and Productivity.
- Pan, Z., & Kosicki, G. M. (1993). Framing analysis: An approach to news discourse. *Political Communication*, 10, 55~75.
- SAS (2012). Data equity : Unlocking the value of big data.
- Scheufele, D. A. (1999). Framing as a theory of media effects. *Journal of Communication*, 49, 103~122.
- Schudson, M. (1995). *The power of news*. MA: Harvard University press.
- Solove, D. (2004). *The digital person: Technology and privacy in the information age*. New York and London: New York University Press.
- Tuchman, G. (1978). *Making News: A study in the Construction of Reality*. New York: The Free Press.
- van Dijk, T. (1998). *Opinions and ideologies in the press*. In A. Bell. & P. Garrett(Eds.). *Approaches to media discourse*(pp. 21~63). London: Blackwell.
- Wang, W. & Rada, R.(1998). Structured hypertext with domain semantics. *ACM Trans. Inform. Syst.*, 16, 372~412.
- Wasserman, S. & Faust, K. (1994). *Social network analysis: methods and applications*. Cambridge: Cambridge University Press.

논문 투고일 2014년 1월 15일
논문 수정일 2014년 2월 15일
게재 확정일 2014년 2월 20일

ABSTRACT

A Semantic Network Analysis of the Newspaper Articles on Big data

Choi, Yoon-jung

Ph.D. student, Dept. of Journalism and Mass Communication, Sungkyunkwan
University

Kweon, Sang-Hee

Professor, Dept. of Journalism and Mass Communication, Sungkyunkwan
University

The purpose of this paper is to grasp the media frames about big data which has become a new technological and social phenomenon. For this purpose, related articles of the four major daily publications were gathered, from the year 2011 when the term ‘big data’ appeared for the first time on the newspaper to the year 2013, and then analyzed through a semantic network analysis.

As a result, this study found the change of the media frames concerning big data from the neutral attitudes in 2011 to the positive one started in 2012. Most of the articles emphasized the beneficial aspects of big data since they focused on practical uses. Although two progressive newspapers dealt with harmful consequences that big data might cause, the difference among the media was not distinctive. In addition, this study showed that not only business areas but also public sectors such as health care services or traffic policies are eager to participate in the big data market. This paper also verified the

influence of a new phenomenon caused by ICT(information and communication technology) upon social awareness.

Keywords: Big Data, Semantic Network Analysis, Newspaper Articles