

## 빅데이터를 활용한 북한 인식에 대한 연구 - 네이버 데이터랩, 구글 트렌드 활용 -

이종영  
강원과학기술센터

### <요약문>

최근 빅데이터는 의료, 축산, 오염, 마케팅 등 다양한 분야에서 활용되고 있고 우리의 생활에 밀접하게 들어와 있다. 빅데이터를 통해 남북한 당국의 화해 분위기와 불안 분위기가 조성될 때마다 북한에 대해 언제 검색이 많고 어떤 시기에 검색량이 많은지 알아보아 북한에 대한 인식의 변화를 알아보려고 하였다.

본 연구에서 빅데이터를 활용하는 대표적인 사이트인 네이버와 구글에서 북한에 대한 이슈가 있을 때 검색량이 어떻게 변하는지를 알아보려고 검색 빈도 분석을 통해 국민의 북한에 대한 인식을 알아보려고 하며, 이를 성별과 기기별로 얼마나 사용하는지를 분석하였다.

연구결과 네이버의 데이터랩과 구글의 구글트렌드를 활용해 '북한'이라는 키워드의 검색빈도를 알아보았는데, 네이버와 구글의 검색빈도에서는 차이가 있었지만 대체적으로 북한과 연관된 이슈가 있었을 때 검색빈도가 높게 나타나는 것을 알 수 있었다. 그러나 네이버와 구글을 사용하는 계층이 다르고, 사용되는 검색빈도를 추출하는 것에서 차이가 있는 것으로 보인다.

특히, 북한이라는 특수성에 따라 접경지역인 강원도와 서울에서 북한을 바라보는 시각이 높기 때문에 검색빈도가 높게 나타났고, 남성과 여성의 빈도를 보면 상대적으로 여성의 검색빈도가 높았으며, 네이버 데이터랩은 PC에서 구글 트렌드는 모바일에서 높은 검색 빈도가 나타났다. 이는 네이버의 경우 전연령층에서 고루 사용한 반면, 구글은 젊은층의 높은 사용이 있다고 추측되어 진다.

향후 네이버의 데이터랩과 구글의 트렌드는 검색 키워드에서 검색빈도가 차이가 있는 것으로 보여짐에 따라 향후 두 검색빈도의 차이를 규명하고 공통된 이슈가 될 수 있도록 조정이 필요하며, 검색량이 절대적인 것은 아니라는 것을 이해하고 해석해야 할 것이며, 연령대에 따라 북한에 대한 인식이 달라질 수 있음에 따라 향후 구체화된 연령과 키워드를 활용하여 검색빈도를 찾아낸다면 국민의 북한과 관련한 이슈에 대해 더 명확히 알 수 있을 것이라 보여 진다.

▣ 주제어 : 빅데이터, 북한, 북한인식, 네이버데이터랩, 구글트렌드

## I. 서론

최근 빅데이터는 의료, 축산, 오염, 마케팅 등 다양한 분야에서 활용하고 있고 우리의 생활에 밀접하게 들어와 있다.

특히, 스마트기기, 센서 등의 급속한 보급과 모바일 인터넷과 SNS의 확산으로 데이터 량이 기하급수적으로 증가하여 데이터가 경제적 자산이 될 수 있는 빅데이터의 시대가 도래하였고, 빅데이터는 엄청나게 많은 데이터로 양적인 의미를 벗어나 데이터의 분석과 활용을 포괄하는 개념으로(송태민, 2012) 이와 관련하여 남북한 당국의 화해 분위기와 불안 분위기가 조성될 때마다 국민들이 북한에 대한 검색의 변화를 파악하여 북한에 대해 언제 검색이 많고 어떤 시기에 검색량이 많은지 알아보고 북한에 대한 인식의 변화를 알아보려고 하였다.

빅데이터를 활용하는 대표적인 사이트인 네이버와 구글에서 북한에 대한 이슈가 있을 때 검색량이 어떻게 변하는지를 알아보면 국민들의 북한에 대한 인식을 알 수 있고 다양한 의견을 수렴할 수 있는 역할을 할 수 있을 것이고, 향후 국민적 공감대를 수렴하는 기회가 될 수 있을 것이라 보았다. 서울시민 남북교류 협력 의식조사에서 서울시민 10명 중 6명은 남국관계에 부정적이었으나 10명 중 7명은 남북통일이 필요하다고 생각하는 것으로 조사되었듯이(한국일보, 2019/12/25) 국민들의 인식이 지속적으로 변화하고 있음에 따라 이슈가 되었을 때마다 변화하는 인식을 검토해 볼 필요가 있겠다.

본 연구에서는 빅데이터라고 할 수 있는 포털 사이트인 네이버와 구글의 데이터랩과 트렌드의 검색 빈도 분석을 통해 국민의 북한에 대한 인식을 알아보려고 하며, 이를 성별과 기기별로 얼마나 사용하는지를 분석하고자 한다. 다양하고 심층적인 통계를 적기에 제공 받고 빠르게 변화하는 구조를 미리 감지하고 선제적으로 대응하기 위해서는 빅데이터 기반이 중요하고(뉴시스, 2019/12.27) 빅데이터에서의 빈도분석은 포털 사이트에서 일어나는 현상을 기존과 다르게 파악할 수 있는 특징이 있음에 따라 이를 토대로 변화를 알아볼 수 있을 것이다.

기존의 연구는 소비자 가치와 예측을 빅데이터를 활용하여 연구하였고, 빅데이터를 활용하여 북한에 대한 인식에 대해서는 연구는 거의 없었던 점과 한 개의 포털 검색이 아닌 네이버와 구글 모두를 검색하여 설명하고 차이를 점검하고자 하였다.

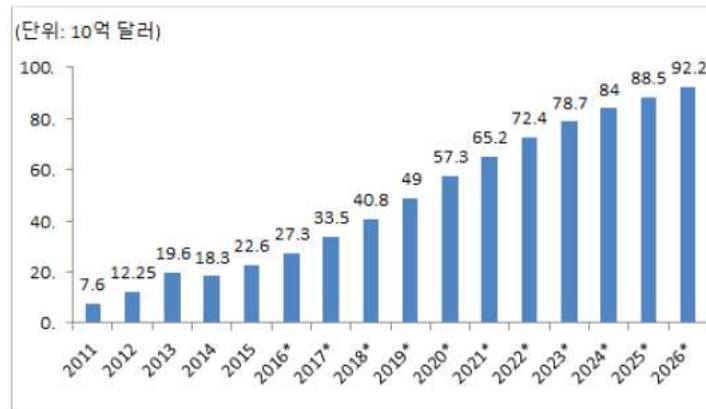
이러한 연구는 북한과의 관계가 급변하고 있어 관심도에 대한 변화를 선제적으로 파악하고 향후 통일에 대한 국민의 의식조사와 연관 될 수 있다. 그리고 향후 빅데이터를 활용하여 북한에 대한 인식에 대한 동향 분석이 필요하고 정책방안을 마련하는데 필요한 기초자료에 도움이 될 수 있을 것으로 판단한다.

## II. 선행연구

### 1. 빅데이터의 개념

빅데이터(Big Data)란 기존의 데이터 용량을 넘어서는 거대한 규모의 다양한 정형·비정형 데이터를 가리키며, 글로벌 시장조사 기관인 IDC에 의하면 2025년 전 세계의 데이터 생산량은 약 163ZB 정도의 크기가 될 것으로 예측되며, 이는 2016년에 생성된 16.1ZB보다 약 10배 정도 늘어난 규모이다. ICT기술의 발전과 함께 등장한 스마트 기기의 확산, 소셜네트워크서비스

스(SNS)의 활성화, 사물인터넷(IoT)의 보급 등으로 인해 빅데이터의 기반은 더욱 확대되고 있으며, 이들로부터 생성되는 방대한 양의 데이터들을 저장하고 처리하는 기술 또한 점점 진보해나가고 있다. 시장조사기관인 Wikibon에 따르면 소프트웨어, 하드웨어, 서비스를 모두 포함한 세계 빅데이터 시장은 앞으로 10년 후인 2026년에 총 922억 달러의 규모로 성장할 것으로 전망된다. 이는 2014년 기록했던 183억 달러에서 약 404% 증가한 수치이며, <그림 1>과 같이 2014년부터2026년까지의 연평균 성장률은 14.4%에 달한다(박선우, 2018).



<그림 1> 빅데이터 세계시장 규모 예측 : 2011~2026

주 : \*는 예측치, 출처 : Statista

원출처 : Wikibon Big Data in the Public Cloud Forecast, 2016-2026(Wikibon)

빅데이터는 기존의 통상적으로 사용되는 데이터 수집, 관리 및 처리 SW의 수용한계를 넘어서는 거대한 규모의 데이터로 양(Volume), 속도(Velocity) 및 다양성(Varity) 측면에서 전통적인 데이터와 구분되는 정보자산을 의미하며, <표 1>과 같은 특성이 있다(김수연외 2, 2018).

<표 1> 전통적 데이터와 빅데이터의 특성 비교

구분	전통적 데이터	빅데이터
양	기가바이트(GB)	테라바이트(TB), 페타바이트(PB)
생산주기(속도)	시간, 일	실시간
구조	구조적	반구조 및 비정형
데이터 원천	중앙집중	분산
데이터 통합	쉬움	어려움
데이터 저장	RDBMS <sup>1)</sup>	HDFS <sup>2)</sup> , NoSQL <sup>3)</sup>

출처 : KISTEP 기술동향브리프, 2018-11

- 1) 관계형 데이터베이스 관리시스템(Relational Database Management System) : 모든 데이터가 테이블에 저장된 관계 모델에 의해 저장 관리되는 데이터베이스 시스템으로 전통적인 대부분의 데이터베이스는 이 모델을 기반으로 함
- 2) 하둡 분산 파일 시스템(Hadoop Distributed File System) : 범용 하드웨어에서 실행되도록 설계된 분산 파일 시스템으로 내결함성이 높으며 저가의 하드웨어에 배치되도록 설계되고, 응용 프로그램 데이터에 대한 높은 처리량의 액세스를 제공하여 대용량 데이터셋 응용 프로그램에 적합
- 3) '데이터를 저장하는 방식이 SQL만 있는 것은 아니다(Not Only SQL)'라는 뜻으로, 클러스터에서 실행할 목적으로 만들어져 전통적 데이터베이스시스템인 RDBMS에서 사용하는 관계형 모델을 사용하지 않고, 스키마 없이 동작하며 구조에 대한 정의를 변경할 필요 없이 DB 레코드에 필드를 자유롭게 추가할 수 있음

빅데이터는 금융, 마케팅, IT에서 많이 활용하고 있고, 이를 잘 활용하기 위해서는 지식과 경험을 갖춰야 하며 온라인과 오프라인의 연계가 더욱 중요해질 것이다.

그리고 해결하고자 하는 문제에 가장 적합한 자료를 선정하는 것이 중요한데 빅데이터의 경우에는 그 자체로 주요한 결론을 도출하기 보다는 기존 데이터를 보완하는 방식으로 활용하는 것이 바람직하고 문제 해결을 중심에 놓고 판단하면 효과적인 데이터를 찾는 데 도움이 될 수 있기 때문에 해결하고자 하는 문제에 대해 구체적으로 정의하는 작업이 우선되고 이를 위해서는 구체화될수록 데이터를 선별하는데 도움이 될 수 있다(유경원, 2019).

또한, 빅데이터의 80% 이상이 비정형 데이터로 추정되며, 이 가운데 특히 텍스트 데이터가 매우 중요한 위치를 차지하고 있으며, 빅데이터를 구성하는 텍스트 데이터의 주요 원천은 소셜 미디어이다. 이와 같은 현상은 온라인 공동체의 형태를 지닌 온라인 카페나 게시판에서도 관찰되며 나아가 포털 서비스의 포스팅, 뉴스 기사 등에 대한 댓글 또한 온라인 공간에서의 미디어 소비와 여론 생성 현장을 고스란히 기록하고 있다(이기황, 2016)고 할 수 있다.

빅데이터에서 새로운 복합체로서 양적인 변화뿐만 아니라, 그와 더불어 질적 변화까지 생긴다는 것이다. 그리고 이렇게 만들어진 자료는 미시 수준에서나 거시 수준에서나 종래에는 가능하지 않았던 즉 전통적인 사회과학의 틀에서는 할 수 없었던 새로운 형태와 내용의 분석을 가능하게 할 것이고 또 기술 환경의 변화추세가 앞으로 더 큰 규모로, 더 빠른 속도로 지속될 것이라는 점을 감안할 때 이렇게 생산되는 자료의 비중은 계속 커질 것이라는 점을 강조한다(Lazer et al., 2009). 그러므로 관련된 분야에 따라 또는, 다루는 문제에 따라 달라지기는 하겠지만 경험적 연구의 지평을 열어주는 기반으로 이런 자료가 만들어진다는 것은 자료의 질을 높여주는 것뿐만 아니라 새로운 방식의 연구를 시도해 볼 수 있는 기회를 제공하는 것이기도 하다(한신갑, 2015).

## 2. 빅데이터 및 북한인식 논문

빅데이터는 정보통신 분야에서의 화두로 정형 및 비정형 데이터들을 의미하고 빅데이터의 특징으로는 크기, 속도, 다양성을 들 수 있는데(국립중앙과학관, 2019) 이를 활용하여 대규모의 정보를 빠른 시간 안에 분석할 수 있다. 이를 잘 활용하면 고객의 행동을 미리 예측하고 대처방안을 마련할 수 있고, 생산성 향상, 시민이 요구하는 서비스 제공의 기회로 작용하고 향후 민간부문 뿐만 아니라 공공 부문의 혁신을 수반하는 패러다임의 변화를 가져올 수 있다(네이버 지식백과, 2019).

뉴스는 수많은 일들 속에 사람들의 관심사를 다루기도 하지만 새로운 관심을 불러오기도 하며, 사람들의 관심과 이슈를 살펴보는 방법은 다양한데 온라인 접촉이 일상화하면서 ‘실시간 급상승 검색어’가 대표적이다. 나와는 다른 의견을 가진 사람들의 다양한 목소리를 접하며, 다름에 대한 이해를 넓힐 수 있고, 같은 생각을 하는 사람들과 인식과 정보 공유를 통한 연대감을 느낄 수 있기(한국일보, 2019/12/21) 때문에 포털을 이용한 실시간 검색빈도는 매우 유용한 정보가 될 수 있다.

특히, 빅데이터 기술의 발전은 그동안 제한적 데이터로 예측해왔던 분석에 새로운 패러다임의 변화 가능성을 제시하고 있고, 더 많은 데이터의 양과 빈도 측정이 가능하다는 것 외에도 정형 데이터뿐만 아니라 비정형 데이터까지도 일정한 패턴과 흐름을 분석하고 변화를 예측하는데 사용 될 수 있는(이종민외 2, 2017) 장점이 있다.

국내외 인터넷 포털사이트와 빅데이터를 이용하는 연구가 있고(채영길·유용민, 2017; 이종민·이종아·정준호, 2017; 송해엽·양재훈, 2017; 이관섭·우종필, 2018; 왕린·이주경·황지현, 2019; 한선, 2009; 이궁희, 2017; 김형중·김진화, 2017), 포털의 검색지수를 활용한 실증분석의 연구와(박수지·신지영·정철, 2017) 빅데이터를 활용한 국민 통일 인식과 관련한 연구(송태민, 2015; 김상배, 2017)가 있다. 또한, 북한에 대한 인식과 관련한 연구(송샘·이재목, 2019; 이준희, 2019; 이성우·우준모, 2019)가 있다.

빅데이터 분석은 데이터를 활용한 분석을 대신하는 통상적인 표현으로 자리 잡고 있으며(최재경, 2016), 이는 4차 산업의 핵심적인 기술 중 하나로 크고 다양한 데이터를 분석하는 기법으로 핵심적인 기술로 등장하고 있으며, 일부 기업들은 빅데이터 분석을 경영에 활용하고 있다. 국내에서는 네이버 데이터랩(Naver DataLab)과 구글 트렌드(Google Trends) 등이 검색어의 검색 빈도 및 트렌드를 제공하고 있는데, 네이버 데이터랩은 특정 검색어의 검색 추이를 기간, 범위, 성별, 연령별로 제공하고 있고 구글 트렌드 역시 특정 검색어의 기간별, 지역별 검색 추이를 보여주며(김석영외 1, 2018), 박수지외 2(2017)인의 연구에서는 네이버 데이터랩을 활용한 관광 활성화 수요예측을 통해 효과적인 수용태세의 대비와 다각적인 전략을 제시하였다.

북한에 대한 인식과 관련하여 송샘·이재목(2019)의 연구는 가치관이 형성되는 시기에 한 세대에 집단적으로 경험한 것은 그 세대의 정치적 성향과 인식에 지속적으로 영향을 미치며, 주변국에 대한 세대별 인식 차이가 외교·안보·대북(통일) 정책에 대한 의견 차이로 이어져 갈등의 핵심축이 될 수 있다는 점을 주목해야 한다고 하였다.

특히, 이준희(2019)의 연구는 우리나라 정권의 북한에 대한 이미지와 남북관계에 대해 우호적 행위가 많아지면 공세지향의 비우호적 행위가 점차 줄어들어야 하는데 우호적 이미지와 행위 외에도 비우호적 행위 또한 많아지고 있고, 남과 북의 행위에 있어 우호적 관계가 많아진다고 해서 비우호적 관계가 줄어드는 제로섬 관계가 아니라는 것을 말했다.

설진아(2009)의 연구는 인터넷 검색 포털 사이트인 구글과 네이버에 대해 어떻게 이용, 평가하고 있는지를 실증 분석하였는데 두 사이트 역할에 대해서는 차이가 있지만 최신정보제공과 전문적이고 가치 있는 정보를 제공한다고 하였으며, 포털 사이트는 이제 정치, 경제, 사회, 문화 분야 등 일상생활에서 무시하지 못할 영향력을 행사하게 된다고 하였다.

채영길·유용민(2017) 연구는 포털의 모바일 뉴스사이트가 배열한 선거 뉴스를 분석함으로써 정치 커뮤니케이션에서 포털의 역할과 그 함의를 고찰하였는데, 포털의 선거 보도가 뉴스 중심어 의미연결망 구조에서 차이가 있는 것으로 나타나 새로운 차원의 정책적 접근과 사회적 관심을 갖게 한다고 하였다.

이는 빅데이터로 대변되는 온라인 포털의 데이터 플랫폼이 다양한 사회현상을 설명할 수 있다는 의미로 해석할 수 있을 것이다.

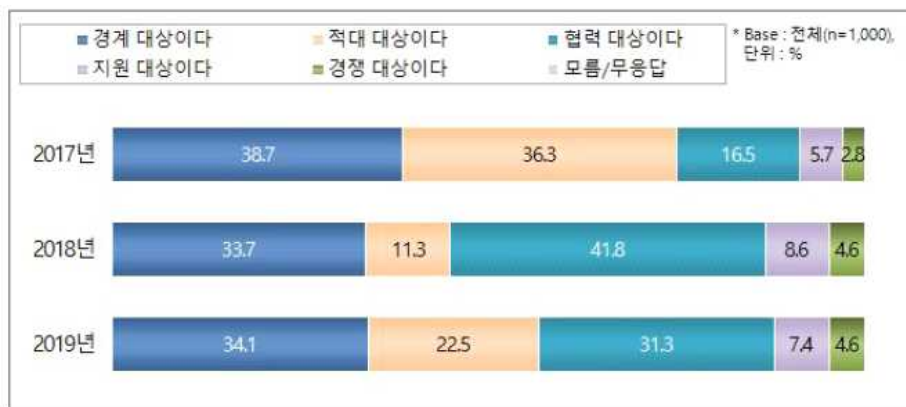
이종민외 2(2017)인의 연구에서는 네이버와 구글의 포털 검색 지수를 활용한 전세 가격 예측 실증분석으로 포털 검색 지수가 경제지표로서의 유용성이 있는지의 여부를 검증한 선행연구와 유사하게 포털 검색지수를 활용할 경우 전세시장 예측 정확도 향상에 유의미한 결과를 나타내고 있음을 검증하였고, 전세가격 예측 시 포털 검색 지수를 활용한다면 보다 효과적이고 선제적인 정책 입안과 시장 대응이 가능할 수 있다고 하였다. 이는 검색 지수가 실증적으로 검증하였기 때문에 검색 지수가 의미 있는 결과를 보여준다는 것이다.

김상배(2017)의 연구에서는 빅데이터는 큰 규모를 활용해서 작은 규모에서는 불가능했던 새로운 통찰과 새로운 형태의 가치를 창출해 내는 일을 가능케 한다고 하면서 빅데이터 분석을

외교안보 정책에 활용하고자 필요한 전제조건과 실현방안을 검토하고 빅데이터 분석의 프레임워크를 제시하였다.

이현진·이승우(2018)의 연구는 최근 주목하고 있는 빅데이터를 활용하여 인터넷을 사용하는 불특정 다수를 대상으로 사회적기업 인식에 관해 네이버 데이터를 활용하였는데 일자리 문제와 사회적기업에 관한 관심의 상관관계는 유의한 정의 관계가 있는 것으로 나타났다는 것을 제시하였다.

그리고, KBS 남북교류협력단(2019)의 조사결과 2019 국민 통일의식 조사 결과 북한의 김정은 정권에 대해 반감을 느낀다는 응답자가 51.6%로 나타났으며, 북한은 경계대상(34.1%)이라는 응답과 협력대상(31.3%)이라는 응답이 비슷하게 <그림 2>와 같이 나타나 최근 북한의 행보에 대한 긍정적인 시각과 부정적인 시각이 공존하는 것으로 나타났다.



<그림 2> 남북한 관계에 대한 인식

출처 : 2019년 국민통일의식 조사 결과(요약), 2019

종합적으로 본 논문과 관련성 있는 선행연구에 대한 검토결과 기존 연구와의 차이점은 다음과 같다. 첫째, 본 연구는 포털 검색 지수를 이용해 북한에 대한 검색량을 파악하여 얼마나 인식하고 있는지와 둘째, 본 논문은 네이버와 구글의 검색 지수를 모두 분석하고 그 결과를 비교해 차이를 점검하고자 하였다.

### Ⅲ. 분석 설계 및 분석 결과

#### 1. 조사대상 및 자료수집 방법

검색 키워드 형태의 빅데이터를 이용한 연구로는 구글 검색지수를 활용한 네이버의 데이터랩과 구글의 트렌드를 활용한 연구들에 따라 2016년 1월 1일부터 2019년 11월 30일까지 북한이라는 핵심 키워드를 이용 키워드 검색 추이를 찾았다.

네이버와 구글은 모두 사용자의 로그(Log)를 수집하고 집적하여 검색서비스를 활용해 새로운 가치로 창출한다는 점에서 유사하나, 네이버는 정보를 가지고 콘텐츠를 구축하고 구글은 정보를 집적해 아카이브를 구축하는 특징이 있다(설진아, 2011, p.160).

네이버의 데이터랩을 통한 핵심키워드 검색 추이 트래픽 값을 수집하고, 20세 이상의 사람

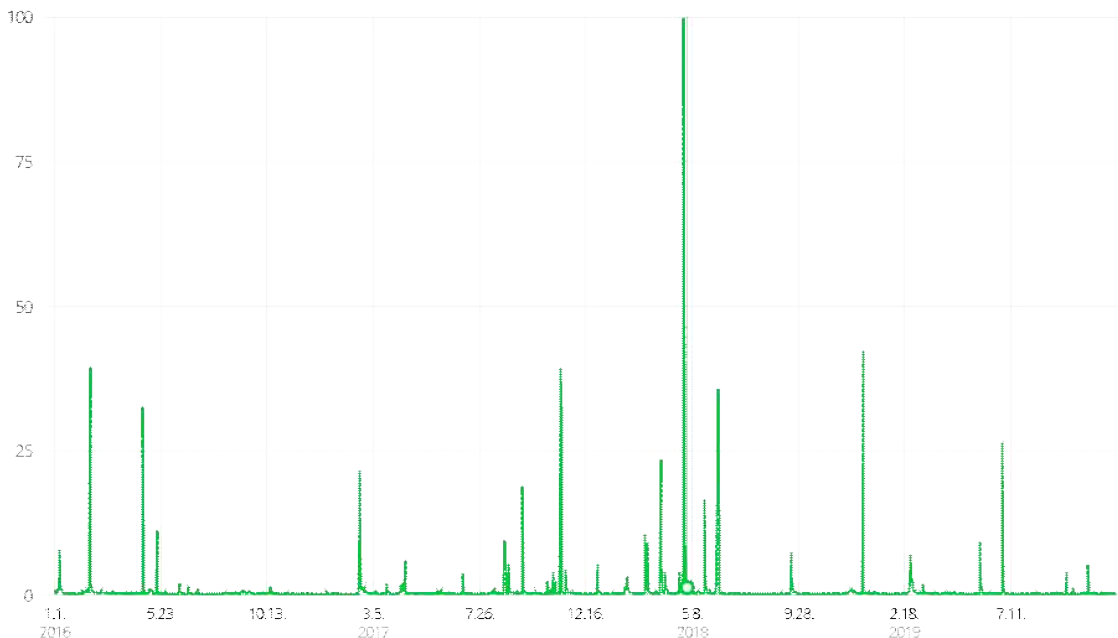
이 검색한 값으로 한정하여 자료를 수집하였다. 네이버 트렌드 데이터는 해당 검색어가 검색된 횟수를 각각 합산하여 조회기간 내 최대 검색량을 100으로 설정하여 상대적인 변화를 나타낸다.

검색량을 총량으로 나누어 0에서 100까지의 상대적인 수치로 나타내기 때문에 환경적인 영향을 최소화하고 트렌드 데이터가 가진 본질적인 특성을 분석에 활용할 수 있으며 데이터의 변화와 비교에 쉽다는 장점이 있다(이현진외 1, 2018).

구글 트렌드에서는 사용자가 입력한 검색어 빅데이터를 분석하여 사용자가 특정시간에 특정 지역에서 검색어를 입력한 검색량을 표준화하여 제공하는 것으로 네이버 데이터랩과 같이 100을 기준으로 상대적인 수치로 나타내고 1주일 단위로 나타낸다.

## 2. 네이버 검색 빈도 분석을 통한 북한 검색 현황

네이버 데이터랩에서 2016년 1월부터 2019년 11월 30일까지 북한이라는 키워드로 모바일 및 PC와 19세 이상 남녀를 포함하여 검색한 결과 아래의 <그림 3>과 같이 나타났으며, 최대 검색량을 100이라고 했을 때 2018년 4월 27일 검색량이 100으로 가장 많았으며 2018년 12월 24일 42.4, 2016년 2월 18일 39.4, 2017년 11월 12일 39.3의 순으로 나타났다.



<그림 3> 북한 키워드 검색 빈도 추이

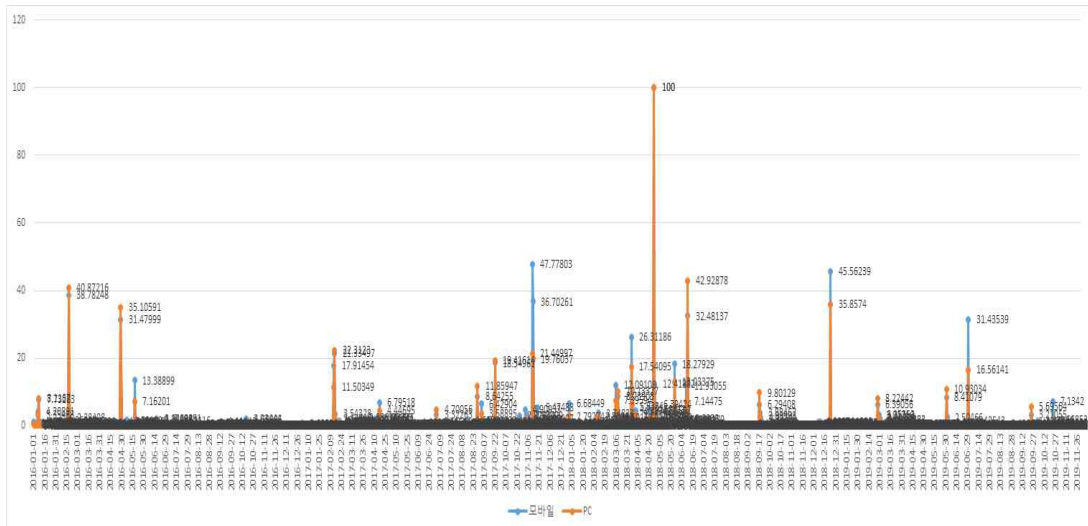
출처 : 네이버 데이터랩

포털인 네이버에서 2018년 4월 한 달 동안 북한이라는 키워드로 검색된 뉴스 기사가 58,016건으로 북한과의 대화 및 남북경협 기대감에 대한 기사가 많았으며, 4월 27일 1일간 북한 관련 기사 수가 13,538건으로 높았다. 이는 남북정상회담이 군사분계선에서 역사적인 악수를 TV와 언론 보도가 나온 날로 관련 검색 빈도가 급증한 것으로 보인다. 12월 24일의 경우 기사건수가 869건으로 대북제재 여파와 하노이의 베트남-북한 축구 평가전 관련인 것으로 보인다.

4월 27일이 역사적인 날로 국민들은 북한에 대한 검색어를 PC로 검색한 것으로 나타났다

대부분의 상위를 차지한 빈도는 PC로 검색한 것으로 나타났다.

이는 젊은 세대보다는 PC가 자유로운 세대가 일반적으로 검색한 것으로 보이며, 모바일 검색이 가장 높은 날은 2017년 11월 12일로 651건의 기사가 검색되었으며, 북한이 탄도미사일에 사용되는 고체 연료식 엔진 연소실험이 있었던 날이었다.

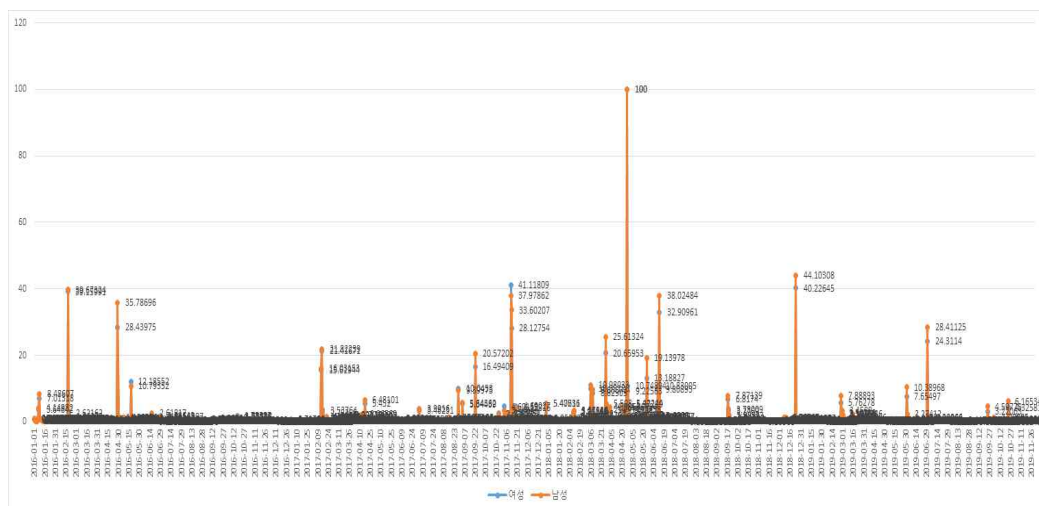


<그림 4> 북한 키워드 중 모바일 및 PC 사용 빈도 추이

주 : 네이버 데이터랩 자료를 Excel로 변경하여 표현

출처 : 네이버 데이터랩

네이버 데이터랩을 통해 검색한 주체가 남자인지 여자인지 구분한 결과 남성이 압도적으로 많았으나 2017년 11월 12일은 여성이 더 많이 검색한 것으로 나타났는데 이는 북한의 위협에 더 민감하게 여성이 작용하지 않았나 하는 추측이다.



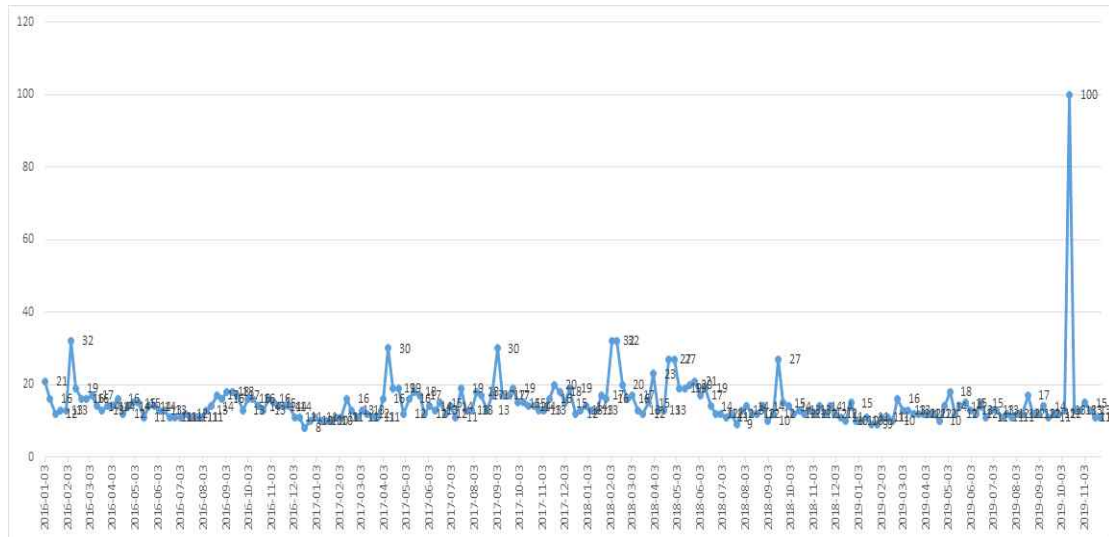
<그림 5> 북한 키워드 중 남성 및 여성 사용 빈도 추이

주 : 네이버 데이터랩 자료를 Excel로 변경하여 표현

출처 : 네이버 데이터랩

## 3. 구글 트렌드 빈도 분석을 통한 북한 검색 현황

구글 트렌드에서 2016년 1월부터 2019년 11월 30일까지 북한이라는 키워드로 검색한 결과 아래의 <그림 6>과 같이 나타났으며 이 검색빈도는 1주일 단위로 나타난 수치이다. 2019년 10월 13일의 검색량이 100으로 가장 높았다. 이날의 북한관련 기사는 650건으로 한미연합훈련 비방에 따른 비방이 있었고 한국축구대표팀(파울루 벤투 감독)의 북한 원정이 있었다.



<그림 6> 구글 트렌드 북한 키워드 검색 빈도 추이

주 : 구글 트렌드 자료를 Excel로 변경하여 표현

출처 : 구글 트렌드

구글 트렌드에서 북한 검색 키워드를 지역별로 살펴보면 강원도가 검색 빈도량이 가장 높은 100으로 나타났고, 서울특별시, 경상북도 순으로 나타났다. 이는 접경지역인 강원도와 북한과 인접한 서울에서 가장 많은 관심을 갖고 있는 것으로 판단된다.



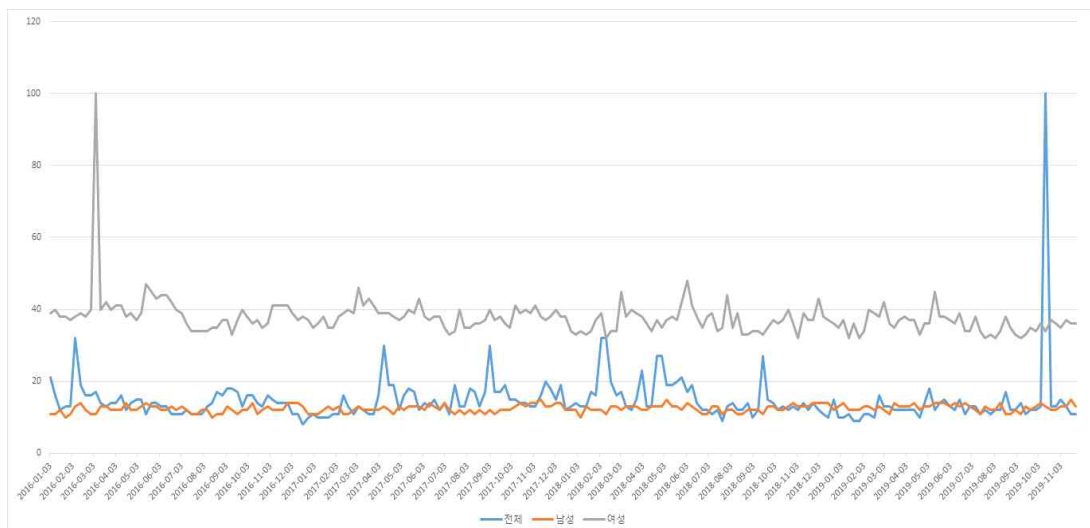
<그림 7> 구글 트렌드 북한 키워드 지역별 빈도 추이

주 : 구글 트렌드 자료를 Excel로 변경하여 표현

출처 : 구글 트렌드

또한, 구글 트렌드에서 북한에 대한 키워드 검색으로 남성과 여성의 검색 빈도를 보면 평균이 15인데, 남성은 12, 여성은 38의 검색빈도가 나타나 여성이 북한에 대해 더 민감하게 작용하는 것으로 보여 진다.

특히, 2016년 3월 6일 검색빈도수를 보면 여성이 100으로 나타났으며, 이날 북한관련 기사 건수는 552건으로 북한의 미본토 목사발 발언으로 위협이 있던 날이었다.



<그림 8> 구글 트렌드 북한 키워드 성별 빈도 추이

주 1) 구글 트렌드 자료를 Excel로 변경하여 표현

2) 빈도 100인 날짜의 기사 건수는 네이버 뉴스 카테고리에서 검색한 수치임

출처 : 구글 트렌드

<표 2>에서와 같이 지역별로 남성과 여성의 관심도에 대해 살펴보면 강원도에서 북한 검색어가 24%로 가장 많고 여성이 58%, 남성이 18%로 여성의 빈도가 높게 나타난다. 이는 전 지역에 공통적으로 나타나고 있다.

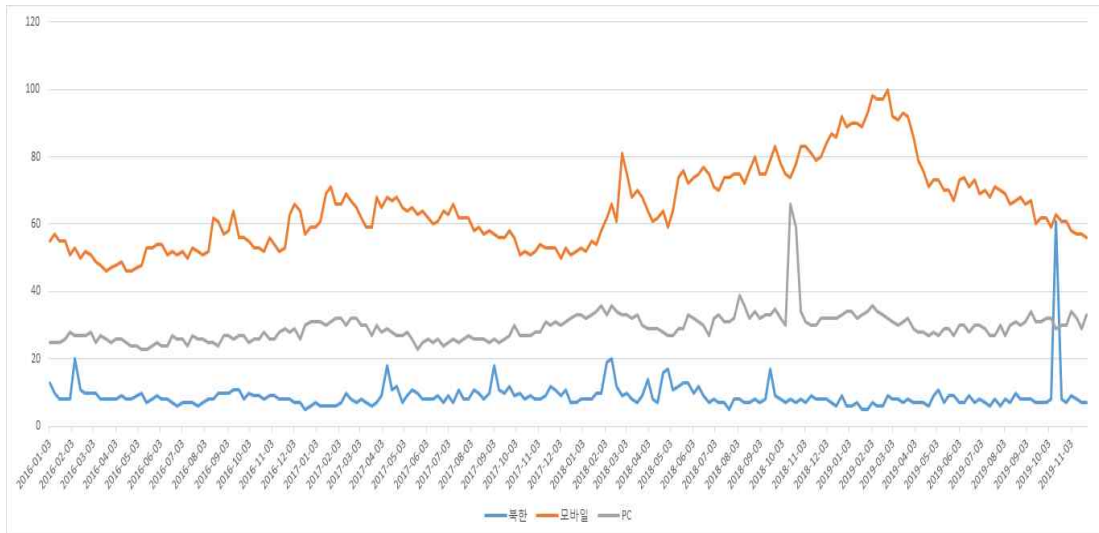
<표 2> 지역별 북한에 대한 관심도

단위 : %

지역	북한	남성	여성
강원도	24	18	58
서울특별시	23	19	58
경상북도	23	19	58
충청북도	23	19	58
충청남도	24	19	57
경상남도	22	19	59
부산광역시	21	21	58
대전광역시	22	20	58
대구광역시	22	20	58
인천광역시	21	19	60
경기도	22	19	59
전라남도	21	20	59
전라북도	21	22	57
울산광역시	21	20	59
제주특별자치도	21	19	60
광주광역시	18	20	62

출처 : 구글 트렌드

그리고, 구글 트렌드에서 북한에 대한 키워드 검색은 주로 모바일이 65%로 많고, PC가 29%로 나타나 구글에서는 모바일로 검색하는 빈도수가 높은 것으로 나타났다. 이중 모바일로 검색빈도가 가장 높은 날은 2019년 2월 24일로 이날 북한관련 기사 건수는 1,244건으로 세계적 투자가인 짐로저스가 북한은 매력시장이라는 발언과 김정은 위원장의 하노이행 출발을 이례적으로 신속 보도 한 날이었다.



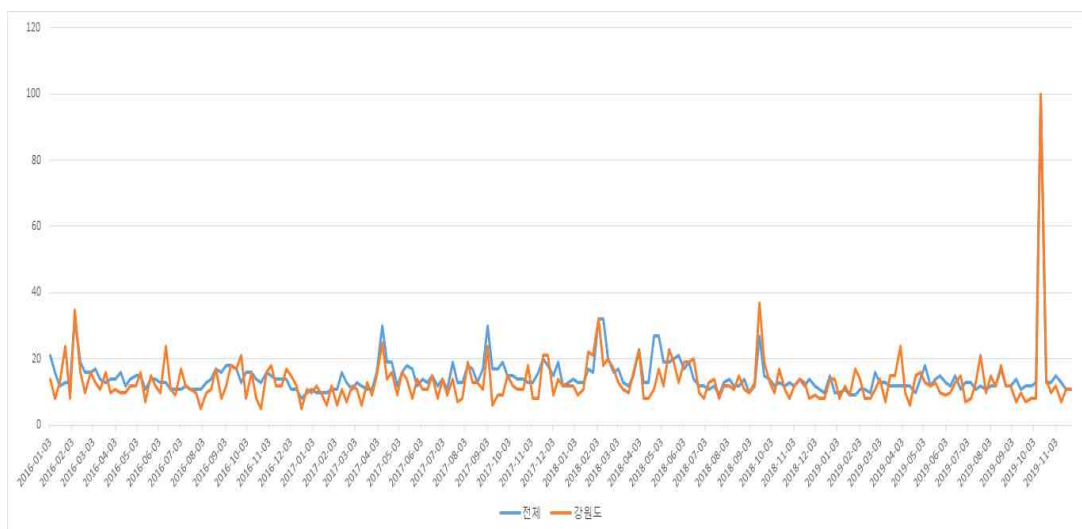
<그림 9> 구글 트렌드 북한 키워드 기기별 빈도 추이

주 1) 구글 트렌드 자료를 Excel로 변경하여 표현

2) 빈도 100인 날짜의 기사 건수는 네이버 뉴스 카테고리에서 검색한 수치임

출처 : 구글 트렌드

특히, 접경 지역으로 북한과 밀접한 관계가 있는 지역인 강원도에서 북한에 대한 검색 결과를 보면 2019년 10월 13일이 가장 높은 검색빈도인 100을 나타냈는데, 이날의 북한 관련 기사는 650건으로 한미연합훈련에 따른 비방이 있었고 한국축구대표팀(파울루 벤투 감독)의 북한원정이 있었던 날로 구글 북한 검색의 전체적인 검색 빈도와 같이 나타났다.



<그림 10> 구글 트렌드 북한 키워드 강원도 빈도 추이

주 1) 구글 트렌드 자료를 Excel로 변경하여 표현

2) 빈도 100인 날짜의 기사 건수는 네이버 뉴스 카테고리에서 검색한 수치임

출처 : 구글 트렌드

그리고, 강원도 내에서 지역별로 보면 춘천시가 100으로 나타났고, 원주시가 77, 강릉시가

74로 나타났다. 이는 인구적인 측면과 정치적인 측면, 접경지역인 측면 등 다양한 요소가 있을 것으로 보여 진다.

## V. 결론 및 한계점

### 1. 결론

빅데이터 분석은 사회현상에 대해 새로운 분석 방법의 가능성을 제시하고 있는데, 분석 대상에 대한 구체적인 정보를 알지 못하더라도 인터넷 검색 빈도 빅데이터 분석만으로도 관심 있는 현상에 대해 개략적인 파악 및 분석이 가능하고 행태적 특성에 대해 보다 정확한 분석 및 예측이 가능(김석영외 1, 2018, p.9)하다는 점으로 북한이라는 키워드를 활용해 국민들이 북한에 대해 얼마나 검색하고 인지하고 있는지를 확인한 연구라고 하겠다.

네이버의 데이터랩 검색지수는 최대값을 기준으로 작성되기 때문에 별도로 검색된 검색 지수간 값들을 서로 비교할 수 없고, 구글 트렌드 검색통계도 역시 주 단위로 작성 되는데, 네이버와 유사하게 최대 검색량을 100으로 하여 상대적 값으로 지수 형태로 표시된다(이종민외 2, 2017, p.140).

본 연구에서는 빅데이터인 네이버의 데이터랩과 구글의 구글트렌드를 활용해 ‘북한’이라는 키워드의 검색빈도를 알아보려고 하였다. 네이버와 구글의 검색빈도에서는 차이가 있었지만 대체적으로 북한과 연관된 이슈가 있었을 때 검색빈도가 높게 나타나는 것을 알 수 있었다. 그러나 네이버와 구글을 사용하는 계층이 다르고, 사용되는 검색빈도를 추출하는 것에서 차이가 있는 것으로 보인다.

특히, 북한이라는 특수성에 따라 접경지역인 강원도와 서울에서 북한을 바라보는 시각이 높기 때문에 검색빈도가 높게 나타났고, 남성과 여성의 빈도를 보면 상대적으로 여성의 검색빈도가 높았으며, 네이버 데이터랩은 PC에서 구글 트렌드는 모바일에서 높은 검색 빈도가 나타났다. 이는 네이버의 경우 전연령층에서 고루 사용한 반면, 구글은 젊은층의 높은 사용이 있다고 추측되어 진다.

2018 정보화통계집(한국정보화진흥원, 2018)에 따르면 2017년 12월 말 기준 전국 사업체 76.7%가 데스크탑 컴퓨터를 보유하고 있고, 인터넷 접속이 가능한 업체는 80.6%이며 인터넷 접속방법으로는 광랜이 59.1%, 무선랜 33.7%, 모바일 인터넷 26.5% 순으로 나타났고, 모바일 기기 이용률은 48.6% 수준으로 추정되었다. 이는 일반적으로 PC를 통한 검색이 많고, 모바일도 높은 수준으로 사용하는 것으로 보여진다.

이를 통해 국민들이 북한에 대해 언제 검색빈도가 높아지고, 어떤 기기에서 어떤 성별에서 관심도가 있는지 알게 되었으며, 앞으로 북한과의 인식을 개선하기 위해서는 국민이 어떻게 바라보느냐도 중요하지만, 다양한 계층의 빈도를 보고 희망을 갖고 현실적으로 접근해야 하며 대북지원이나 경험의 경우 조급한 마음에 한걸음 앞서려고 하지 말고 국민과의 충분한 인식이 된 다음 논의해도 될 것이다(이상현, 2018).

### 2. 시사점 및 한계점

본 연구의 한계는 동일한 이슈라도 키워드를 어떻게 선정하느냐가 중요한데, 현 북한과 관

련하여 단절된 형태로 이어오다 보니 북한과 관련하여 너무 단순한 단어인 ‘북한’이라는 키워드를 사용한 점이라는 것이다. 단어가 너무 일반적인 단어이거나 혹은 분석에 방해되는 단어일 경우 이의 전처리가 필요한데 현 시점에 특정한 단어를 활용할 수 없다는 문제점이 있기 때문에 향후 이러한 부분에서 구체화가 필요할 것이다.

특히, 인터넷 검색 행위가 다양한 의도를 모두 포괄하고 있기 때문에 생기는 한계를 인지하고 트렌드 자료는 상대적인 검색량을 제공하고 있지만 데이터의 왜곡이 전혀 발생하지 않았다고 보기는 어렵기 때문에(이현진외 1, 2018) 북한이라는 키워드가 명확했다고는 할 수가 없다. 따라서 향후에는 특정한 키워드를 가지고 장기적인 측면에서 검증하는 것이 필요하다고 하겠다.

또한, 네이버의 데이터랩과 구글의 트렌드는 검색 키워드에서 검색빈도가 차이가 있는 것으로 보여짐에 따라 향후 두 검색빈도의 차이를 규명하고 공통된 이슈가 될 수 있도록 조정이 필요하고 검색량이 절대적인 것은 아니라는 것을 이해하고 해석해야 할 것이며, 연령대에 따라 북한에 대한 인식이 달라질 수 있음에 따라 향후 구체화된 연령과 키워드를 활용하여 검색빈도를 찾아낸다면 국민의 북한과 관련한 이슈에 대해 더 명확히 알 수 있을 것이라 보여진다.

## 참고문헌

- KBS남북교류협력단.공영미디어연구소. (2019). “2019년 국민 통일의식 조사 결과(요약).”, 2019
- 김상배. (2017). “외교안보 분야 빅데이터 분석 : 어떻게 수행할 것인가?”, 『서울대학교 국제문제연구소 위킹페이퍼』. No.25, 2017.12.11
- 김석영.정원석. (2018). “빅데이터에 나타난 보험산업.”, 『KIRI리포트』, 2018.12.10
- 김수연.도지훈.김보라. (2018). “빅데이터.”, 『KISTEP 기술동향브리프』, 2018-11호, 2018
- 김형중.김진화. (2017). “소셜 빅데이터 분석을 통한 소비자 가치 인식 연구 : 신규 스마트폰을 중심으로.”, 『한국전자거래학회지』, 제22권 제1호, 2017, 123-146.
- 박선우. (2018). “빅데이터 시대와 데이터 융합.”, 『초점』, 제30권 1호, 통권 661호, 2018.1.16
- 박수지.신지영.정철. (2017). “네이버 데이터랩을 활용한 수요예측과 관광 활성화 방안 : 2017 울산 방문의 해를 중심으로.”, 『한국관광학회 국제학술발표대회집』 82권 0호, 2017. 7, 382-391.
- 배영. (2019). “데이터로 본 한국인 포털댓글 급증한 9월-그 중심엔 ‘조국’.” 『한국일보 오피니언 칼럼』, (12월 21일)
- 설진아. (2011). “대학생집단의 인터넷 포털 사이트에 대한 사용자 평가 연구 : 네이버와 구글을 중심으로.”, 『언론과학연구』, 제11권 1호, 2011.3., 157~186.
- 송샘.이재목. (2019). “한반도 주변국에 대한 세대별 인식 차이 분석 : 남남 갈등과 세대갈등의 중첩 가능성 연구.”, 『지역과 세계』, 제43집 제1호, 2019., 117-141
- 송태민.(2012). “빅데이터를 활용한 자살요인 다변량 분석 : Google 검색 트렌드 적용.”, 『보건.복지 ISSUE&FOCUS』, 제 168호(2012-49), 2012.12.14
- (2015). “소셜 빅데이터를 활용한 국민 통일인식 동향분석.”, 『보건.복지

- ISSUE&FOCUS』, 제269호(2015-01), 2015.1.5
- 송해엽.양재훈. (2017). “포털 뉴스 서비스와 뉴스 유통 변화 2000-2017 네이버 뉴스 빅데이터 분석.”, 『한국언론학보』, 61권 4호, 2017, 74-109.
- 양승준. (2019). “서울시면 10명중 8명 ”남북관계 인식차로 우리사회 내부갈등 심각“.” 『한국일보』, (12월 25일)
- 왕린.이주경.황지현. (2019), “노인의 건강 및 안전문제에 대한 빅데이터 분석.”, 『한국콘텐츠학회논문지』, Vol.19.No4, 2019, 336-344.
- 유경원. (2019). 빅데이터를 활용한 경제데이터분석 사례 및 방법론 연구, 국회예산정책처 연구용역보고서, 2019.10.3
- 이관섭.우종필. (2018). “네이버 빅데이터를 이용한 자동차 관심도 분석.”, 『한국유통학회학술대회 발표논문집』, 2018.4, 98-102.
- 이궁희. (2017). “뉴스 빅데이터를 이용한 경기판단 : 빅카인즈 뉴스 경기지수의 개발.”, 『통계연구』 제22권 제2호, 2017, 67-94.
- 이기항. (2016). “언어 자료의 보고, 빅데이터.”, 『새국어생활』, 제26권 제2호, 2016.6.30
- 이상현. (2018). “남북정상회담 평가와 북미정상회담 전망.”, 『KDI 북한경제리뷰』, 5월호, 2018, 37-41.
- 이성우.우준모. (2019). “권력변동에 따른 대북정책 변화와 여론 추이의 상관성:박근혜정보(2015)와 문재인정부(2017) 사례비교.”, 『정치.정보연구』, 제22권 1호, 2019.2.28, 163-198.
- 이종민.이종아.정준호. (2017). “포털 검색 지수를 활용한 전세 가격 예측-네이버.구글을 중심으로.”, 『부동산학보』, Vol.68, 2017, 134-148.
- 이준희. (2019). “북한의 김대중.박정희 정권에 대한 이미지와 남북관계.”, 『국방연구』, 제62권 제1호, 2019.3, 35-62.
- 이현진.이승우. (2018). “사회적기업 인식에 관한 연구 : 네이버 트렌드 데이터를 활용하여.”, 『사회적기업연구』, 11권 1호, 2018, 51-74.
- 최동준. (2019). “홍남기 ”신뢰받지 못하는 통계 의미없어, 빅데이터 기반조성 ” 『한국일보』, (12월 27일)
- 채영길.유용민. (2017). “네이버.다음 모바일 포털 뉴스 플랫폼의 19대 대통령 선거기사 분석.”, 『사이버커뮤니케이션학보』, 통권 제34권 4호(2017년 12월), 195-242.
- 최재경. (2016). “빅데이터 분석의 국내외 활용 현황과 시사점.”, 『KISTEP InI』 14호, 2016.7.4
- 한국정보화진흥원. (2018). “2018 정보화 통계집”, 2018
- 한선. (2009). “인터넷 포털 네이버의 정치경제학:시장구조와 지지매커니즘을 중심으로.”, 『한국언론학회학술대회발표논문집』, 2009.9, 41-47.
- 한신갑. (2015). “빅데이터와 사회과학하기 : 자료기반의 변화와 분석전략의 재구상.”, 『한국사회학』, 제49집 제2호(2015년), 161~192.
- Lazer, David, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, Tony Jebara, Gary King, Michael Macy, Deb Roy, and Marshall Van Alstyne. (2009). “computational Social science.” 『Science』, 323(5915), 721-723.

국립중앙과학관, 빅데이터, <http://www.science.go.kr/>  
네이버 지식백과 “빅데이터 정의”

네이버 데이터랩 <http://datalab.naver.com/>  
구글 트렌드 <http://trends.google.co.kr/trends/?geo=kr/>