

텍스트마이닝을 활용한 남북관계 온라인 신문기사 분석 - 문재인 정부에서의 남북관계 -

이종영
(재)강릉과학산업진흥원

<요약문>

오늘날 빅데이터를 활용하는 기술은 전 산업뿐만 아니라 사회현상도 분석하고 예측하는 기술로 진화하고 있다. 이에 비정형 데이터인 남북관계 관련 온라인 신문 기사를 중심으로 빅데이터 분석방법인 텍스트마이닝을 활용해 신문기사와 연계한 문재인 정부에서의 남북관계에 대해 분석을 하였다. 신문 기사를 수집하여 텍스트 데이터를 분석하는 것은 거시적 관점에서 일반적인 흐름을 살펴보는 데 새로운 접근법이 될 수 있기 때문에, 신문 기사에 나타난 남북관계 기사를 단순 빈도분석, TF-IDF 분석, N-gram 분석을 빅카인즈(BigKinds)에서 제공하는 기사 검색 서비스를 활용하여 분석하였다.

본 연구의 결과 문재인 정부 초기에는 남북관계와 관련된 기사 양이 많았으나 점점 감소하는 등 북한의 특별한 이슈가 있을 경우에만 높은 관심도를 나타내는 것으로 나타났으며, TF-IDF 분석 결과 김정은 위원장 > 북미관계 > 한반도 비핵화 > 평창올림픽 순으로 나타나 남북관계와 밀접한 관련이 있는 명사들이 추출되었음을 알 수 있었다. 그리고 남북관계의 연관성을 살펴보면 ‘북한’, ‘한반도’와 같이 대상이나 관계를 나타내는 단어와 연결되어 있어 문재인 정부의 대북정책과 연관성이 높은 것을 알 수 있다.

본 연구는 텍스트마이닝 기법으로 남북관계에 대한 뉴스 기사를 통해 핵심단어와 중요도를 계량화하여 의미 있는 정보를 찾았다는 점에서 의의가 있다고 하겠으나, 향후에는 남북관계뿐만 아니라 다양한 키워드 및 다양한 매체를 통해 데이터마이닝 기법을 적용해 남북관계에 대해 좀 더 입체적으로 파악할 필요가 있다.

▣ 주제어 : 빅데이터, 텍스트 마이닝, 남북관계, 신문기사, 빅카인즈

I. 서 론

오늘날 기술의 발전으로 방대한 양의 데이터들로 다양한 정형·비정형 데이터를 만들어 내고 있는데, 이러한 데이터를 빅데이터(big data)라고 한다. 빅데이터를 활용하는 기술은 전 산업 뿐만 아니라 사회현상도 분석하고 예측하는 기술로 진화하고 있다.

빅데이터(big data)에서 사회변화를 예측하고 문제 해결에 필요한 가치와 정보를 제공 받고 데이터에서 의미 있는 가치를 추출해 내는 것이 텍스트 마이닝(text mining)으로 관련된 키워드의 사회적 동향과 이슈를 파악하는데 사용하고 있다(전해정, 2019).

우리 주변에는 트위터, 페이스북 같은 SNS(Social Network Service) 등의 사용과 함께 개인정보와 위치정보 등이 포함된 빅데이터(big data) 환경이 도래(김재생, 2014)하고 있고, 소셜미디어를 활용한 빅데이터 분석은 검색 키워드와 연관성이 높은 영향요소를 도출함으로써 미래 변화에 대한 예측이 가능해질 뿐만 아니라 이용자의 관점에서 새로운 가치 및 정보를 창출한다는 의미가 있다(김민형, 2020).

또한, 최근에는 뉴스가 디지털로 축적하여 분석하는 기술이 발전함에 따라 체계적이고 종합적인 분석이 가능하며 주제들에 대한 시각화 또한 가능하게 됨에 따라 뉴스데이터의 의미망 분석은 다각적인 시사점을 도출할 수 있다고 하였다(강주연·이이든·김지수, 2020).

오늘날 북한은 경제발전을 중요한 부분으로 인식하고 있고, 국제사회의 대북제재 틀 속에서도 부분적인 제재의 빈틈을 찾아 일정 정도의 실적을 낼 가능성도 존재하기(탁용달, 2020) 때문에 남북관계에 대해 빅데이터 속에서 의미 있는 상황을 파악할 필요가 있다.

특히, 통일연구원의 통일의식조사(2021)에서 통일이 필요하다는 응답이 상당폭 상승(2020년 52.7% → 2021년 58.7%)하였는데 이는 남북관계의 변화와 밀접하게 연동되어 있다고 볼 수 있고, 2016년 이후 계속 평화공존을 선호하는 비율이 증가하였는데, 이는 남북관계의 미래를 바라보는 관점이 북한을 통일의 대상이 아닌 공존의 대상으로 보는 추세로 변화하고 있다고 할 수 있다.

이에 문재인 정부가 출범하고 북한과의 관계가 호전되기도 하였으나 국제정세의 흐름에 따라 시시각각으로 변화하고 있는 것이 최근의 실정으로 북한 문제는 국내의 단순 보수와 진보의 논리, 북한 핵 위기 등 단순 논리에서 보다는 대내와 대외의 관점이 연결된 상태에서 이루어져야 함(황지환, 2011)에 따라 우선 국내의 신문기사들 속에서 남북관계에 대한 현상을 찾아내는 것이 필요한 상황이다.

본 연구에서는 비정형데이터인 남북관계 관련 온라인 신문기사를 중심으로 빅데이터 분석방법인 텍스트 마이닝을 활용해 신문기사와 연계한 남북관계에 대해 분석을 하고자 하였다.

신문기사를 수집하여 텍스트 데이터를 분석하는 것은 거시적 관점에서 일반적인 흐름을 살펴보는 데 새로운 접근법이 될 수 있을 것이라 보여 진다.

시간적 범위는 문재인 정부가 출범한 2017년 5월 10일부터 2021년 11월 25일까지로 빅카인즈(bigkinds.or.kr)를 활용하여 신문기사에서 나타난 남북관계의 현상을 데이터 마이닝으로 분석하였다.

본 연구의 구성은 다음과 같다. 2장에서는 선행연구로 텍스트 마이닝과 남북관계에 대한 선행연구를 분석한다. 3장에서는 분석결과를 도출하고 4장에서는 결론과 한계점을 제시하였다.

본 연구는 북한과의 관계가 급변하고 있는 시점에 남북관계의 관심에 대한 변화를 선제적으

로 파악하고 향후 남북관계에 대한 정책방안을 마련하는데 필요한 기초자료에 도움이 될 수 있을 것으로 판단한다.

II. 선행연구

1. 텍스트마이닝

빅데이터 시장은 연평균 성장률 10.6%(2020년 1,388억 8,600만 달러 → 2025년 2,294억 2,300만 달러) 증가하고 있으며, 일상적인 비즈니스에 유입되고 더 나은 의사결정과 전략적 통찰력을 얻기 위해 빅데이터를 분석 한다(연구개발특구진흥재단, 2021).

광범위하게 수집된 텍스트 데이터는 빅데이터 분석 방법 중 하나인 텍스트 마이닝(Text Mining)으로 분석가능하며, 이상아·최범성·최진혁(2020)은 텍스트 마이닝을 활용한 신문기사의 발달장애 조기선별과 조기중재에 대해 빅데이터 분석방법을 활용하여 우리사회의 인식특성을 분석하였다.

텍스트 마이닝을 활용한 연구로는 오신휘·김혜진(2020)의 저출산 관련 변화과정을 비교분석하기 위해 텍스트 마이닝과 동시출현단어 분석기법을 활용하여 실증 분석을 수행하였다.

심재권(2020)은 연구 동향을 분석하기 위해 텍스트 마이닝 방법을 활용하였는데, 신뢰도와 타당성에 대한 문제로 많은 양의 데이터를 분석하기 위한 새로운 방법으로 LDA 기반 토픽 모델링 연구방법론이 기존의 분석 방법에서 확인하지 못했던 주제까지 발견할 수 있어 적합하다고 인용하였고 연구 동향 파악에 사용하였다.

우경숙·서주환(2021)은 네트워크가 관계를 형성하는 디지털 기술의 발전에 따른 정보와 지능화를 이용한 스마트 가든의 개념 및 인식경향을 파악하고자 텍스트 마이닝 기법의 빈도분석, 키워드 연결망분석, 이미지 분석을 활용하였다.

뉴스 데이터를 이용한 강주연·이이든·김지수(2020)는 포털 사이트에서 뉴스데이터를 추출하고 분석하여 Z 세대와 관련된 의미구조를 확인하였는데, Z 세대의 특성을 설명하는 단어와 소비와 관련된 단어를 추출하는 등 뉴스데이터 분석을 통해 Z 세대에 대한 다양한 의미연결 구조망을 확인하였다.

그리고 텍스트 마이닝을 활용한 북한연구로 김민형(2020)은 텍스트 마이닝 분석을 실시하여 기존연구의 조사방법과 표본추출의 한계를 극복하였으며, 최근 사회적 분위기를 반영하여 정상회담, 통일, 개방, 협력, 남북 등의 용어가 빈번하게 나타났으므로 법적 또는 제도적 시스템을 구축하여 관광분야의 남북 교류를 지속화 하여야 한다고 하였다.

이에 본 연구에서도 텍스트 마이닝 기법의 빈도분석, 키워드 연결망 분석, 이미지 분석으로 남북관계에 대해 분석하였다.

2. 남북관계

김정은 체제하의 북한의 안보정책은 핵무력 건설에 집중했던 병진노선의 전략노선으로부터 ‘국가 핵무력 건설 완성’선언 이후의 대외관계 개선에 나서게 되는 경제건설 총노선의 시기로 변화되어 왔고(김연수, 2020) 이와 연결되어 북한은 김정일 시기 자본주의 세계 경제를 생존을 위한 공간으로 인식하고 이 공간 속에서 대외경제 관계를 어떻게 만들어 나갈 것인가에 관

심의 초점을 두었으며, 김정은 집권 이후에는 세계화를 세계경제 변화의 하나로 인정하고 이전과는 확연하게 다른 인식의 전환을 가져오고 있다. 이에 우리는 북한의 변화를 읽어내고 이에 맞는 적절한 정책을 수립하고 실천해야 한다고 하였다(정영철, 2018).

또한, 남북관계의 사건이 북한인식과 대북정책 등 전반적인 통일 의식에 어떤 영향을 미치는지, 그 효과와 강도를 살펴보면 남북관계 사건이 안보불안 의식에 미치는 영향과는 달리 통일자체의 인식을 변화시키는 효과는 제한적으로 나타남(김병로, 2017)에 따라 남북관계에서의 사건은 국민의 남북관계에 대한 변화가 있을 수 있으므로 이에 대해 살펴볼 필요가 있다.

황지환(2011)은 북한문제 인식의 문제점과 새로운 접근의 필요성에서 대북포용정책이나 대북압박정책 모두 북한정권의 속성이나 세계 질서에 대한 보다 현실적인 사고를 통해 재조정될 필요가 있다고 하였다.

특히, 남북관계에서 북한이탈주민에 대한 관계에서도 김종원(2019)은 남한주민들이 북한에 대해 더 많이 접촉하고 북한 사회를 알아갈수록 북한이탈주민에 대한 친근감이 높고, 남한사회 수용여부에도 긍정적이라는 것을 확인할 수 있었다. 향후 남북관계의 변화 및 개선을 위해서는 서로간의 관계를 확인할 필요가 있겠다.

김재선(2014)은 국내 각 언론매체의 북한관련 뉴스보도에 대한 증가 양상에 대해 언론 현실에서 국민들의 북한 관련 의식에 매우 커다란 영향을 미치고 있는 북한 현장 취재기자들의 취재원에 대한 인식을 주관성 연구방법인 Q 방법을 이용해 분석하였고, 4개의 유형을 도출하기도 하였다. 이는 국내 언론들이 동일한 북한 관련 이슈에 대해서도 자신의 이념적 성향에 따라 보도의 내용과 방식을 다르게 구성해서 보도함으로써 북한 이슈에 대한 의견과 태도 형성에 큰 영향을 미친다는 전제에서 출발하였다.

박시영(2019)은 김정은 정권은 위협회피전략으로 전환 이후에 남북관계의 전략적 가치를 높게 평가했다는 것으로 남북관계가 북한에게 있어 완충제, 중재자로서의 충분한 전략적 가치를 인식하기 시작했다는 것이다.

이에 남북관계는 다양한 시각으로 변화되고 있으며, 인식의 변화에 따라 달라짐에 따라 국내 각 언론 매체에서 남북관계에 대한 기사를 살펴보고 그 기사들의 의미 있는 현상을 살펴보고 남북관계의 변화 현상을 확인하였다.

Ⅲ. 분석 설계 및 분석 결과

1. 자료수집

본 연구는 남북관계에 대한 분석을 통해 우리 사회에서 남북관계를 어떻게 바라보고 있는지를 알아보고자 한다. 이를 위한 자료 수집은 한국언론진흥재단에서 운영하는 빅카인즈(BigKinds)에서 기사검색 서비스를 활용하였다. 빅카인즈에는 중앙지, 경제지, 지역종합지, 방송사, 전문지 등에서 보도한 내용을 분석대상으로 하였다.

이에 문재인 정부 출범 이후 남북관계에 대한 신문기사를 분석하여 남북관계와 관련된 주요 단어를 비교하여 우리사회의 남북관계에 대한 흐름을 알고자 하였다. 따라서 자료 수집기간은 문재인 정부가 출범한 2017년 5월 10일부터 2021년 11월 25일까지 설정하였고 신문기사 검색에는 '남북관계'로 검색하였다.

2. TF-IDF 분석

신문기사 데이터에서 명사화 처리된 키워드들의 특정 문서 내 중요도 분석을 위해 단어 빈도-역문서 빈도(Term Frequency-Inverse Document Frequency, TF-IDF)를 산출하였다. 이는 문서에서 특정 키워드가 출현하는 빈도와 이와 연관된 많은 문서에서 공통된 출현 빈도의 양쪽 특징을 종합한 분석 방법으로 어떤 문서에서 나타난 특정 키워드가 해당 문서의 특징을 어느 정도로 표현하는지 가중치 값으로 계산한 것이며, TF-IDF의 계산방식은 특정 단어의 상대적 출현빈도와 전체 문서에서 특정 단어가 출현하는 문서 비율의 역수에 관계된 값을 서로 곱하여 계산한다(정민경.최규완, 2021, 전해정, 2019). 본 연구에서도 이 방법을 채택하였다. TF-IDF 공식은 다음과 같다.

$$TFIDF(w, d) = TF(w, d) \times \log\left(\frac{N}{DF(w)}\right)$$

$TF(w, d)$: 문서 d 에 단어 w 가 나타나는 횟수
 $DF(w)$: 단어 w 가 들어가는 문서의 총수
 N : 전체 문서의 총수

3. N-gram 순열

빈도분석을 통해 나온 단어들은 원래의 자기 자리에 있던 순서를 완전히 무시하고 처리되므로 단어 자체가 가지는 의미를 제대로 표현하지 못할 수 있다. 이를 보완하기 위해 문서에서 인접한 단어들을 쌍으로 묶어 표현하는 N-gram 순열기법을 적용한다(박종영.서충원, 2015).

N-gram이란 연속한 n개의 단어를 하나의 분석 단위로 두는 언어분석방법으로 n개 단어의 연쇄를 확률적으로 나타내어, 실제로 표현된 문장의 기록을 확인하여 분석하는 개념(이상아.최범성.최진혁, 2020)으로 빈도분석을 통해 나온 단어들은 원래의 자기 자리에 있던 순서를 완전히 무시하고 처리되므로 단어 자체가 가지는 의미를 제대로 표현하지 못할 수 있는데 이를 보완하기 위한 기법이다(전해정, 2019).

남북관계에 대한 신문기사에서 남북관계와 연관되는 상황을 빅카인즈를 통해 살펴보았다.

4. 분석 결과

본 연구에서 사용된 자료는 빅카인즈를 활용하여 2017년 5월 10일부터 2021년 11월 25일 까지의 문재인 정부하의 남북관계에 대한 신문 기사를 수집하여 분석하였다. 분석기간의 남북관계 뉴스검색 결과 2017년 5,575건, 2018년 26,574건, 2019년 10,811건, 2020년 12,707건, 2021년 6,452건으로 총 62,119건이 검색되었다. <표 1>은 각 언론사별 남북관계 기사의 월별 기사빈도이다.

<표 1> 남북관계 월별 건수

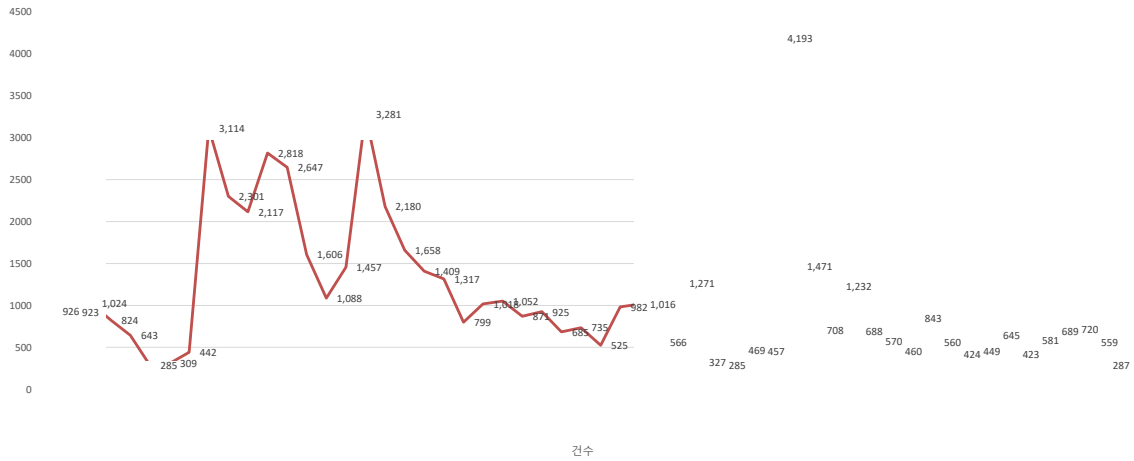
기 간	건 수	기 간	건 수
2017.5	926	2019.9	525
2017.6	923	2019.10	982
2017.7	1,024	2019.11	1,016
2017.8	824	2019.12	566
2017.9	643	2020.1	1,271
2017.10	285	2020.2	327
2017.11	309	2020.3	285
2017.12	442	2020.4	469
2018.1	3,114	2020.5	457
2018.2	2,301	2020.6	4,193
2018.3	2,117	2020.7	1,471
2018.4	2,818	2020.8	708
2018.5	2,647	2020.9	1,232
2018.6	1,606	2020.10	688
2018.7	1,088	2020.11	570
2018.8	1,457	2020.12	460
2018.9	3,281	2021.1	843
2018.10	2,180	2021.2	560
2018.11	1,658	2021.3	424
2018.12	1,409	2021.4	449
2019.1	1,317	2021.5	645
2019.2	799	2021.6	423
2019.3	1,018	2021.7	581
2019.4	1,052	2021.8	689
2019.5	871	2021.9	720
2019.6	925	2021.10	559
2019.7	685	2021.11	287
2019.8	735		

자료 : 빅카인즈

<그림 1>은 남북관계에 대한 기사 빈도의 추이를 나타낸 것으로 문재인 정부 초기에는 남북관계와 관련된 기사 양이 많았으나 점점 감소하다가 2020년 6월 가장 많은 양의 기사가 나타나고 있다. 2020년 6월 4일 남북공동연락사무소 폐쇄, 남북군사합의 파기, 개성공단 철거 등을 거론했으며 6월 16일 개성 남북공동연락사무소 청사 폭파 등 다양한 이슈가 있어 남북관계에 대한 기사의 양이 많아졌다가 급격히 줄어들었다. 따라서 남북 및 북미관계 개선의 출발점은 남북연락.협의 채널 복구에서 시작되는 것이 바람직하고, 신뢰는 관계의 반복 속에서 형성되는 것이기 때문에 소통이 필요하다(최용환, 2021).

그러나 김병로(2017)는 북한의 도발사건은 안보불안을 15~20%를 상승시키는 효과를 지니며, 그 효과는 시간의 흐름에 따라 감소하며 유사한 사건이 반복되면 그 효과는 거의 발휘되지 않는 것으로 분석했고, 남북관계 사건이 안보불안 의식에 미치는 영향과는 달리 통일자체의 인식을 변화시키는 효과는 제한적으로 보았다.

이제는 경험과 인식변화에 대해 이해하고 남북관계에 대해 종합적으로 판단할 시기가 되었다고 본다.

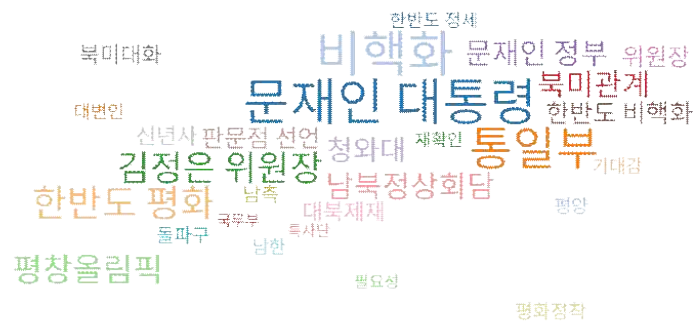


자료 : 빅카인즈

<그림 1> 남북관계 기사 빈도수

다음으로 빅카인즈에서 남북관계에 대해 어떤 단어가 많이 노출되었고 서로 관련성이 있는지에 대해 시각화하여 살펴보았다. 이는 검색결과 중 분석 뉴스와 연관성(가중치, 키워드 빈도수)이 높은 키워드를 시각화 한 것으로 이는 남북관계와 연관성이 가장 높은 건수를 1,000건으로 하여 워드 클라우드(Word Cloud)기법을 사용해 시각화하였다.

시각화 결과 문재인 대통령, 비핵화, 통일부, 한반도평화 등의 단어가 많이 언급된 것으로 나타났다. 워드 클라우드는 단어의 출현빈도가 많은 수록 큰 글자로 나타나며, 단어의 출현빈도가 비슷하면 같은색, 같은 크기로 나타난다.



자료 : 빅카인즈

<그림 2> 남북관계에 대한 연관어 분석

빅카인즈 메뉴얼에서 이 데이터의 유형은 가중치를 선택한 것으로 가중치 정보는 토픽랭크 알고리즘에 기반 해 출력된 결과이며, 토픽랭크는 사용자가 입력한 질의어에 기반하여 검색한 결과에 대해 공식적 분석과 워드 클러스터링을 하여 동적으로 상호 작용하는 시맨틱 네트워크를 생성하는 알고리즘이다.

사용자 질의어에 의해 제시된 키워드의 배열은 의미적 유사도 수식에 따라 결정되고 같은

기본 질의라 하더라도 사용자가 통찰을 얻어내기 위해 입력한 다양한 조건(날짜, 카테고리)에 따라 시맨틱 네트워크를 구성할 수 있게 편리한 인터페이스이다. 가중치로 보면 문재인 대통령 90.47, 비핵화 89.75, 통일부 82.41, 한반도평화 62.29로 나타났다.

TF-IDF 기법은 주관적이거나 혹은 객관적이라고 하더라도 그 문서가 담고 있는 감정을 핵심 단어를 통해 속성을 찾는 과정(박종영, 서충원, 2015)으로 빅카인즈의 데이터를 활용하여 <표 2>와 같이 분석하였다. TF-IDF 분석 결과로 김정은 위원장이 가장 높은 빈도가 나타났으며 TF-IDF 지수는 463.13이었으며, 북미관계 TF-IDF 402.66, 한반도 비핵화 TF-IDF 348.45, 평창올림픽 TF-IDF가 341.36 순으로 나타나 남북관계와 밀접한 관련이 있는 명사들이 추출되었음을 알 수 있었다.

남북관계에서 북한정권과 북미관계에 따라 남북관계의 상황이 달라짐을 알 수 있으며, 한반도 비핵화 문제도 남북관계에 있어 중요한 의미를 갖는 키워드라고 할 수 있다. 이와 같은 키워드로 보았듯이 남북관계는 내부의 문제보다 외부의 문제로 인해 달라진다는 것을 알 수 있다. 특히, 평창동계올림픽은 내외부와 모두 공존한 행사로 남북관계에서 아주 의미 있는 키워드로 볼 수 있다.

<표 2> 남북관계 TF-IDF 분석 결과

순위	단 어	빈도수	TF-IDF	순위	단 어	빈도수	TF-IDF
1	문재인대통령	904	335.01	15	북미대화	280	294.17
2	비핵화	897	92.82	16	신년사	274	208.78
3	통일부	824	51.49	17	남측	225	135.86
4	한반도평화	622	264.95	18	평화정착	217	225.89
5	김정은위원장	572	463.13	19	돌파구	208	222.68
6	평창올림픽	516	341.36	20	남한	205	164.28
7	북미관계	475	402.66	21	한반도정세	205	221.59
8	남북정상회담	470	247.80	22	평양	201	106.86
9	문재인정부	426	307.25	23	대변인	193	118.92
10	청와대	416	186.60	24	기대감	188	221.93
11	한반도비핵화	359	348.45	25	재확인	180	221.25
12	판문점선언	328	277.05	26	필요성	148	170.01
13	위원장	320	25.39	27	국무부	119	124.45
14	대북제재	281	219.15	28	특사단	115	134.26

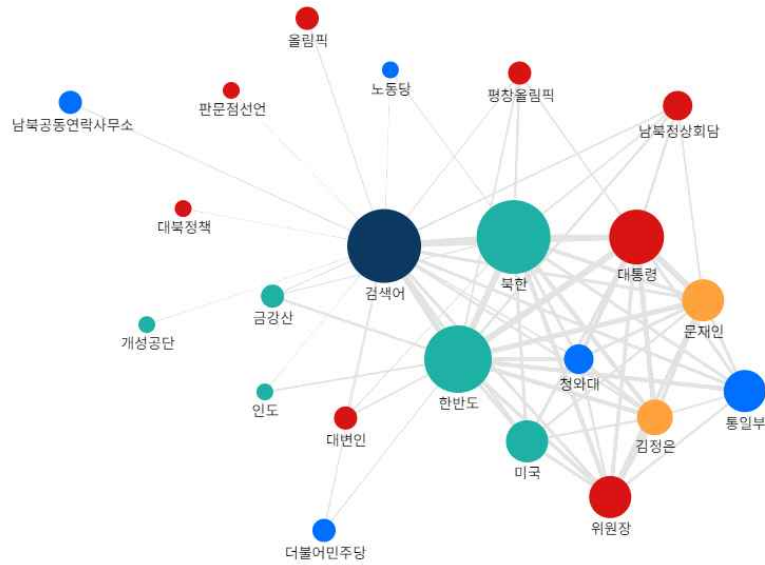
빅카인즈에서 남북관계 뉴스에서 추출된 개체명 사이의 연결 관계를 네트워크 형태로 시각화하여 보았다. 관계도 분석의 추출방식은 검색결과 중 정확도 상위 100건의 뉴스 본문을 형태소로 분석하여 명사상당 어구를 추출 한 것으로 이후 추출된 명사상당어구에 개체명 분석 알고리즘(Structured SVM)을 적용하여 개체명의 관련기사 건수를 고려해 가중치를 부여한 것이다.

검색결과 중 정확도 상위 100건의 분석 뉴스에서 추출된 인물 2건, 장소 6건, 기관 5곳, 키워드 8개의 연결 관계를 네트워크 형태로 시각화하였다.

이와 같은 분석 결과에 의하면 <그림 3>에서 남북관계는 네트워크의 중심부에 위치하여 ‘북한’, ‘한반도’와 같이 대상이나 관계를 나타내는 단어와 연결되어 있다.

남북관계는 북한과 우리정부와 연관이 높고, 우리정부는 대통령, 통일부와 연결되어 있는데 문재인 정부의 대북정책과 연관성이 높은 것을 알 수 있다.

그리고 한반도는 미국과 연결되어 있고, 미국은 김정은 위원장과 연결되어 있음에 따라 남북관계는 미국과 북한과의 연관성에 따라 남북관계 상황이 달라질 수 있음을 알 수 있다.



자료 : 빅카인즈

<그림 3> 남북관계에 대한 N-gram 분석

IV. 결론 및 한계점

본 연구는 비정형데이터인 ‘남북관계’ 키워드를 온라인 신문기사를 중심으로 빅데이터 분석 방법인 텍스트 마이닝을 활용해 신문기사와 연계한 남북관계에 대해 분석을 하였다.

신문기사를 수집하여 텍스트 데이터를 분석하는 것은 거시적 관점에서 일반적인 흐름을 살펴보는 데 새로운 접근법이 될 수 있을 것이라 보여지며, 신문기사에 나타난 남북관계 기사를 단순 빈도분석, TF-IDF분석, N-gram 분석을 한국언론진흥재단에서 운영하는 빅카인즈(BigKinds)에서 제공하는 기사검색 서비스를 활용하였다. 문재인 정부하에서 남북관계 기사를 살펴보기 위해 2017년 5월 10일부터 2021년 11월 25일까지의 신문기사를 수집하여 분석하였다.

본 연구의 결과는 다음과 같다. 첫째, 문재인 정부 초기에는 남북관계와 관련된 기사 양이 많았으나 점점 감소하다가 2020년 6월 가장 많은 양의 기사가 나타나고 있다. 이때 남북공동연락사무소 폐쇄, 남북군사합의 파기, 개성공단 철거 등을 거론했으며, 개성 남북공동연락사무소 청사 폭파 등 다양한 이슈가 있어 남북관계에 대한 기사의 양이 많아졌다가 급격히 줄어들었다. 남북관계 사건이 안보불안 의식에 영향을 미치지만 통일자체의 인식변화시키는 효과에는 제한적(김병로, 2017)이라는 연구가 있는데, 이와 같이 남북관계에 대해서는 특별한 이슈가 있을 경우에만 높은 관심도를 나타내는 것으로 지속적인 남북관계에 대한 현상을 유지하는 방안

이 필요할 것으로 보인다.

둘째, 신문기사 데이터에서 명사화 처리된 키워드들의 특정 문서 내 중요도 분석을 위한 TF-IDF 분석 결과 김정은 위원장 > 북미관계 > 한반도 비핵화 > 평창올림픽 순으로 나타나 남북관계와 밀접한 관련이 있는 명사들이 추출되었음을 알 수 있었다.

남북관계에서 북한정권과 북미관계에 따라 남북관계의 상황이 달라짐을 알 수 있으며, 한반도 비핵화 문제도 남북관계에 있어 중요한 의미를 갖는 키워드라고 할 수 있다. 이는 북한문제에 대한 대응이 대내-대외의 관점에서 연결된 상태에서 이루어져야 하며 대북포용정책이나 대북압박정책 모두 북한 정권의 속성이나 세계질서에 대한 보다 현실적인 사고를 통해 재조정될 필요가 있다(황지환, 2011)는 연구와 일맥상통하며, 이와 같은 키워드로 보았듯이 남북관계는 내부의 문제보다 외부의 문제로 인해 달라진다는 것을 알 수 있다. 특히, 평창동계올림픽은 내.외부와 공존한 행사로 남북관계에서 아주 의미 있는 키워드로 볼 수 있다.

셋째, 남북관계에 대한 신문기사에서 남북관계와 연관성을 살펴보면 네트워크의 중심부에 위치하여 ‘북한’, ‘한반도’와 같이 대상이나 관계를 나타내는 단어와 연결되어 있다. 이는 남북관계는 북한과 우리정부와 연관이 높고, 우리정부는 대통령, 통일부와 연결되어 있는데 문재인 정부의 대북정책과 연관성이 높은 것을 알 수 있다. 따라서 통일문제와 북한문제에 대한 인식은 한반도 정세 변화에서 정권별 시기별로 달라지는 대북정책 이슈가 당파성 변수의 영향이 더 크기 때문에 북한 인식의 경향을 올바르게 수렴할 수 있는 정책을 고민할 필요가 있다고 하였다(이성우, 2020).

그리고 한반도는 미국과 연결되어 있고, 미국은 김정은 위원장과 연결되어 있음에 따라 남북관계는 미국과 북한과의 연관성에 따라 남북관계 상황이 달라질 수 있음을 알 수 있다. 앞으로 미.북간에 진행되는 비핵화 협상 속에서 위협인식, 지속적인 대북제재가 체제의 안정성에 미치는 영향, 이로 인한 김정은의 정치적 통제력 변화 등이 중요한 변수가 될 수 있고(박시영, 2019), 북한의 핵문제 해결의 실마리는 결국 남한이 마련해야 한다는 인식(박순성, 2013)이 필요하므로 남북관계에 대한 전략적 접근이 필요할 것이다.

본 연구는 텍스트 마이닝 기법으로 남북관계에 대한 뉴스 기사를 통해 핵심단어와 중요도를 계량화하여 의미 있는 정보를 찾았다는 점에서 의의가 있다고 하겠다.

남북관계는 일시적으로 나타나는 것이 아니라 정부정책이나 대내외 환경의 변화에 따라 다양하고 급변하는 것이 현실이다. 따라서 향후에는 빅데이터를 활용하여 남북관계에 대해 국민의 인식에 대한 이해와 관심이 필요하며, 이를 위한 다양한 정책을 수립할 필요가 있겠다.

특히, 남북관계뿐만 아니라 다양한 키워드 및 다양한 매체를 통해 데이터 마이닝 기법을 적용해 남북관계에 대해 좀 더 입체적으로 파악할 필요가 있겠다.

참고문헌

- 강주연.이이든.김지수. (2020). “텍스트마이닝을 활용한 ‘Z세대’관련 뉴스데이터 의미연결망 분석.”, 『미래청소년학회지』, 제17권 2호, 25-48.
- 김민형. (2020). “빅데이터를 활용한 북한관광 인식에 관한 연구.”, 『MICE관광연구』, 제20권 제3호, 95-107.
- 김병로. (2017). “남북관계의 사건이 통일의식에 미치는 효과.”, 『통일과평화』, 9집 2호, 157-184.

- 김연수. (2020). “김정은체제 하 북한의 안보정책 진단: ‘위협인식’과 대응을 중심으로.”, 『북한 연구학회보』, 제24권 제2호, 103-138.
- 김재생. (2014). “빅데이터 분석 기술과 활용사례.”, 『한국콘텐츠학회지』, 제12권 제1호, 14-20.
- 김재선. (2014). “Q방법을 통한 한국 신문기자들의 북한에 대한 인식유형 연구.”, 『언론과학연구』, 제14권 4호, 5-52.
- 김종원. (2019). “북한이탈주민의 재이주: 남한주민의 정치 인식을 중심으로.”, 『글로벌정치연구』, 12-1호, 41-65.
- 박순성. (2013). “북한 핵문제 해결과 한반도-동북아 지역 평화를 위한 인식의 전환.”, 『성균관차이나브리프』, 제1권 제2호, 106-109.
- 박시영. (2019). “핵무력 완성 선언 전후 북한의 위협인식과 ‘선호의 역전’: 위협감수에서 위협 회피로.”, 『국방연구』, 제62권 제1호, 1-34.
- 박종영. (2015). “TF-IDF 가중치 모델을 이용한 주택시장의 변화특성 분석.”, 『부동산학보』, 제63권, 46-58.
- 심재권. (2020). “텍스트마이닝을 활용한 연구 동향 분석.”, 『창의정보문화연구』, 제6권 제1호, 23-30.
- 연구개발특구진흥재단. (2021). “빅데이터 시장.”, 『글로벌시장동향보고서』, 2021.06, p.5.
- 오신휘. 김혜진. (2020). “텍스트마이닝 기법을 활용한 ‘저출산’ 관련 연구동향 분석: 정부의 저출산 정책 추진 과정과의 비교를 중심으로.”, 『보건사회연구』, 제40권 제3호, 492-533.
- 우경숙. 서주환. (2020). “스마트가든의 인식경향에 관한 연구.”, 『한국조경학회지』, 제49권 제2호, 51-60.
- 이상민. 민태은. 윤광일. 구분상. (2021). “KINU 통일외식조사 2021.”, 『통일연구원』, 2021.7
- 이상아. 최범성. 최진혁. (2020). “텍스트마이닝을 활용한 발달장애 신문기사 분석: 조기선별과 조기중재를 중심으로.”, 『특수아동교육연구』, 제22권 제1호, 1-27.
- 이성우. (2020). “한반도 정세 변화에 따른 북한 관련 인식태도: 2018년과 2019년 비교를 중심으로.”, 『한국과 국제정치』, 제36권 제4호, 109-135.
- 전해정. (2019). “빅데이터와 텍스트마이닝을 이용한 부동산시장 동향분석.”, 『디지털융합연구』, 제17권 제4호, 49-55.
- 정영철. (2018). “김정은 시대 북한의 세계경제 인식의 변화: 침략과 약탈의 공간에서 생존과 경쟁의 공간으로.”, 『통일과평화』, 제10권 1호, 187-230.
- 최용환. (2021). “남북공동연락사무소 폭파 1년, 남북 연락.협의 채널 복구 필요성.”, 『국가안보전략연구원』, 이슈브리프 269호, 1-4.
- 탁용달. (2020). “김정은 시대 북한의 대외경제: 인식, 정책 그리고 전망.”, 『북한연구학회보』, 제24권 제1호, 121-152.
- 황지환. (2011). “북한문제 인식의 문제점과 새로운 접근의 필요성.”, 『통일과 평화』, 3집 2호, 3-32.