



www.akht.or.kr

# Journal of Hospitality & Tourism Studies

The Academy of Korea Hospitality &amp; Tourism



## 소셜미디어 빅데이터 분석을 이용한 북한관광 인식 연구

### The study of the perception about unified tourism of Korean peninsula through qualitative research

노희경\* · 채익근\*\*

Noh, Hee-Kyung · Chae, Yi-Geun

\* 공주대학교 관광경영학과 박사 수료, hak06@kognju.ac.kr, 관심분야: 통일관광, 북한관광, 관광빅데이터

\*\* 공주대학교 공과대학 컴퓨터공학부 교수, ygchae@kongju.ac.kr, 관심분야: 통계적데이터처리, 패턴인식, 기계학습

#### 요약

본 연구는 소셜미디어를 대상으로 빅데이터 분석기법을 활용하여 북한관광에 대한 인식의 핵심요소를 파악하고 구성개념 간의 관계를 기술적, 탐색적으로 고찰하고자 한다. 2018년 1월부터 2019년 1월까지의 기간 동안 국내 포털사이트 중 블로그가 가장 활성화 되어있는 네이버 블로그를 대상으로 '북한관광'이 포함된 게시글을 수집해 36,339건의 데이터를 분석에 활용하였다. 분석결과 TF 및 TF-IDF값을 통해 텍스트의 중요도를 측정된 결과 국제적 영향관계, 경제적 효과에 대해 중요하게 인식하고 있는 것으로 분석되었다. 이에 국제 정치적 상황의 안정이 전제되어야 북한관광 재개 및 활성화가 이루어질 수 있으므로 지속적인 국가적 노력과 국민의 관심이 필요하다. 또한, 토픽 모델링 결과 남북연계관광, 대북관광투자, 정치적변화, 통일 환경변화, 관광시장변화, 대북자원투자 총6개의 토픽으로 분류되었다. 향후 남북연계관광 상품 개발 및 활성화를 위해 지자체 및 민간의 준비가 필요하며, 북한의 시장개방에 대비하여 남한자본이 차별받지 않고 시장을 선점하기 위한 철저한 준비가 필요함을 시사한다. 마지막으로 중심성 분석결과 연결정도 중심성이 높게 나타난 텍스트는 북한관광 인식에 직접적인 영향을 주는 요소, 근접중심성이 높게 나타난 텍스트는 북한관광 인식에서 다른 구성개념과 어우러져 시너지 효과를 발휘하는 요소, 매개중심성이 높게 나타난 글텍스트는 북한관광 인식하는데 있어 연결고리로서의 역할이 가장 큰 요소, 위세중심성이 높게 나타난 텍스트는 북한관광 인식에 파급력을 지니는 요소로 해석할 수 있다. 본 연구는 관광분야에서 관광대상지에 대한 인식 연구를 진행함에 있어 소셜미디어 빅데이터를 활용하여 비정형데이터를 객관적 · 계량적인 정형데이터로 변환하여 연구에 활용할 수 있는 방안을 제시하는 선행연구로써 학문적 기여를 한다고 본다.

#### 핵심용어

통일관광,  
북한관광,  
빅데이터분석

#### ABSTRACT

This study, using big data analysis on social media, identifies key elements of perception about North Korea tourism. In Naver blog which is the nation's most active portal site's blog, the posts including 'North Korea tourism' are collected and 36,339 datas are used to analyze from January 2018 to January 2019. According to the analysis result of the importance of text measured through the TF and TF-IDF analysis, it is founded that international relationship and economic effect are important. North Korea tourism resumes and revitalizes only if international political situation is stable, so national effort and public attention is consistently needed. Also, using topic modeling, classified as 6 topics which are liaison tour of South and North Korea, investment of North Korea tourism, political change, change in unification environment, change in tourism market, and resources investment in North Korea. Local government and public are requested to prepare to develop and activate liaison tour of South and North Korea and to thoroughly prepare for South Korea capital to dominate the market in advance indifferently for the opening of North Korea market in the future. Finally, as a result of centrality analysis, text with a high degree centrality are elements having a direct effect on the perception of North Korea tourism, text with a high closeness centrality are elements bringing synergy effect on the perception of North Korea tourism with other components. text with a high betweenness centrality are elements playing an important role in a link on the perception of North Korea tourism, text with a high eigenvector centrality are elements having an impact on the perception of North Korea tourism. This study, in progressing perception study about tourist attraction on tourism, academically contributes as a preceding study which suggests ways to utilize by converting unstructured data to objective and quantitative structured data using social media big data.

#### KEYWORDS

unified tourism,  
north korea tourism,  
big data analysis

#### ACCEPTANCE INFO.

2019. 04. 20 원고 접수  
2019. 05. 21 1차 수정본 접수  
2019. 06. 04 2차 수정본 접수  
2019. 06. 14 3차 수정본 접수  
2019. 06. 30 최종 수정본 제출  
2019. 06. 30 최종 게재 확정  
3인 익명심사필

## I. 서론

남북의 문화 관광교류 협력은 여타 남북 협력 사업에 비해 상대적으로 사회적 갈등을 유발할 가능성이 낮을 뿐만 아니라, 평화적 남북 관계 형성과 남북 사회통합의 기초를 제공할 수 있으며, 장기적으로는 통일비용도 절감할 수 있어 그 중요성이 매우 크다(이해정·이용화, 2018). 남북 간의 교류·협력은 두 가지 측면에서 의미를 가지고 있다. 하나는 분단의 고통을 근원적으로 해결하기 위하여 평화적 통일을 이루어야 한다는 당위성을 확신시키는 측면이다. 또 하나는 통일 과정에서 남북 간 인적 교류 활성화는 관계 개선을 촉진시키고, 이를 통해 민족의 동질성을 회복하며 나아가 북한 사회의 변화를 유도할 수 있는 중요한 매개체로 작용하는 측면이다(최의철, 2000). 남북 교류 협력의 재개 및 활성화는 한반도의 상호 이익, 민족경제 공동체 형성 촉진, 균형적 발전, 그리고 남북 긴장완화 등에서 매우 중요한 의미를 가진다(홍순직, 2014). 나아가 국내외 관광객의 관광 목적지 다양화뿐만 아니라 우리나라 관광의 매력성, 안전성, 안정성 등 국가 이미지와 브랜드 또한 상승할 것으로 사료된다(김상태, 2018).

1998년 11월 18일을 시작으로 금강산, 개성 등의 관광교류가 11년간 지속되었으나 2008년 잠정적으로 중단된 지 10년이라는 시간이 흐르고 있다. 그간 금강산 관광을 재개해야 한다는 목소리가 지자체는 물론 시민단체에서 지속적으로 제기되었으나, 5.24 조치, 핵무기 및 대륙간탄도미사일 실험에 따른 대북제재 강화 등으로 남북 관계가 더욱 경색되면서 크게 힘을 얻지 못했다(황수환, 2018). 하지만, 2017년 6월 북한이 평창동계올림픽 참가 첫 제안을 시작으로 1차, 2차 남북 정상회담, 판문점 선언 등으로 남북 교류 협력 사업의 재개에 대한 기대감이 커지고 있다(김홍태, 2018). 정동준 등(2019)의 조사에 따르면 금강산 관광 재개 찬성에 대한 국민적 의견은 2010부터 2017년까지 다소 줄어드는 추세였지만, 여전히 찬성 여론이 높다(정동준 등, 2019). 이용화·이해정(2017)의 2017년 10월 남북 관계 전문가를 대상으로 설문조사를 한 결과에 따르면, 금강산 관광 재개 시 남북 관계에 긍정적 영향을 미칠 것이라는 응답이 정치적 성향과는 별도로 90.8%에 달했다. 금강산 관광 재개 시 방문할 의향이 있냐는 답변에서도 88.7%가 방문 의사가 있다고 응답했다(이부형 등, 2018). 국민들의 북한관광에 대한 찬성, 중립 또는 반대의 입장을 수치로서 파악하는 것도 중요하지만 북한관광에 대한 어떤 감정, 인식 등을 가지고 있는지 확인하는 작업(노희경, 2019)은 현시점에서 상당한 의미가 있는 필요한 연구라고 할 것이다. 또한, 문재인 정부는 “남북 관계 복원을 위해 국제사회의 대북제재를 훼손하지 않는 범위 내에서 민간 차원의 교류를 유연하게 검토한다.”는 정책 방향을 표명하였다. 하지만 북한의 도발 등을 감안하여 방북 등의 구체적인 조치는 북한의 태도 및 국제사회와의 공조, 국민 여론 등을 고려하여 신중하게 검토하겠다고 밝히고 있다(통일부, 2018). 따라서 남북은 한민족이기는 하지만 반세기가 넘는 장기간의 분단으로 인하여 현재는 매우 이질적인 문화와 사회체제를 가지고 있는 실정이므로 이런 상황에서 남북 관광을 통한 교류의 당사자가 될 민간 차원의 인식에 대한 파악(신용석 등, 2005)은 무엇보다 중요하다. 이에 본 연구에서는 국민들의 생각과 의견, 경험, 정보 등을 생산하고 확산시키는 개인의 의견이 담긴 소셜미디어 빅데이터를 대상으로 텍스트 마이닝 기법과 텍스트 네트워크 분석을 통해 남북 관광교류 재개 및 활성화에 대비한 방향성 설정의 시사점을 제시하고자 한다.

## II. 이론적 배경

### 1. 북한관광

금강산 관광은 1998년 4월 30일 김대중 정부의 ‘남북경협 활성화 조치’에 따라 기업인의 방북이 허용된 이후 시작되었다. 1998년 6월 정주영 현대그룹 명예회장이 조선아태평화위와 ‘금강산 관광 및 개발사업’에 합의하고, 그 해 11월 금강호가 동해항에서 장전항을 향해 첫 출항하였지만, 2003년 9월 육로관광이 허용됨으로써 더욱 활성화되기 시작했다. 2002년 제2차 북핵 위기가 발발하여 어려움을 겪었으나, 2007년 6월 내금강 관광이 개시되어 본격적 궤도에 올랐다. 그러나 2008년 7월 금강산 관광객이 피격 사망함으로써 중단되었다. 금강산 관광 이외에 개성관광은 2007년 12월 본격 시작됐으나 2008년 11월 중단되었으며, 백두산 관광은 2008년 상반기 시행 예정이었으나 실행되지 못했다(통일부 홈페이지).

금강산 관광은 정부 당국 간 교류와 민간의 인적 교류를 개방하여 민간인들이 북한 땅을 자유롭게 왕래하고 북한 주민들과 자연스럽게 접촉하여 민족 동질성 회복의 기회를 제공한 최초의 민간사업이다. 당국 간 대화 통로가 없던 상황에서 민간의 관광 경험을 통한 화해 분위기 조성은 극단적인 대립 관계 형성 예방과 비공식적인 외교 채널 역할 수행, 남북정상회담 성사의 기반 조성 등에 기여하였다. 그뿐만 아니라 북한관광 사업을 계기로 개성공단 사업 등 남북 간 경험이 본격화될 수 있었으며, 통일 한국에 대비한 국토의 균형 있는 발전에 기여하였다.

이와 같이 남북 간 동질성 회복 및 점진적 통일의 기반을 마련하기 위해 민간 차원의 관광교류의 재개는 상당히 중요한 의미를 가진다고 본다. 국내 남북 관광 교류와 관련된 연구는 다음과 같다.

박창수 등(2006)는 북한 개성 방문 관광객을 대상으로 통일의식과 통일 당위성, 남북 관광 교류 정책에 따른 북한관광 만족도, 방문 가치 및 재방문 의사에 관한 연구를 진행하였다. 요인분석을 통해 '통일에 대한 당면 문제 및 대비 의식', '남북 관광 교류 정책', '통일의 당위성' 3개의 요인이 도출하였으며, 도출된 요인이 관광 만족도, 재방문의사 및 추천의사에 미치는 영향을 분석하였다. 남북 관광 교류 정책에 대한 태도가 긍정적일수록 북한관광 만족도에 대하여 긍정적인 영향을 미치는 것으로 나타났으며, 재방문 및 추천의사에 영향을 미치는 요인은 남북 관광 교류 정책과 통일 당위성으로 나타났다.

강인원(2006)는 금강산 방문객을 대상으로 한 연구를 통해 북한관광 속성을 도출하였으며 이를 토대로 북한관광 속성과 만족도, 충성도와와의 인과관계를 검증하였다. 안전·신뢰성, 접근·비용성, 전통·상품성, 위락·휴양성, 시설·서비스성, 정보·체제성의 6개 북한관광 속성 요인이 의미 있는 속성으로 확인되었으며, 종속변수인 만족도에 위락·휴양성을 제외한 5개 관광 속성 요인이 영향을 미치는 것을 확인하였다.

황인학(2009)은 북한관광의 서비스 품질이 관광객의 만족 및 행동 의도에 미치는 영향 관계를 분석하였으며, 장동석·고계성(2014)은 북한관광 재개 및 활성화를 위한 전문가들의 남북 관광에 대한 인식 변화를 연구하였다. 그 결과 전문가들은 남북 교류 협력에 향후 관광 부분이 차지하는 비중이 더욱 증가하고, 남북 관광교류가 남북 관계에 긍정적인 영향을 미칠 것이며, 남한 관광객에 대한 북한관광 개방 지역이 점차 확대될 것이라고 예상했지만 긍정적인 전망의 강도는 2006년의 전망에 비해 감소한 것으로 분석되었다.

장동석(2018)은 통일의식을 구성하는 세부 요인을 확정적으로 도출하고, 세부 통일 의식 요인과 북한관광 인식의 상관관계를 분석하였다. 요인분석 결과 통일의식은 통일 필요성 요인, 통일 이익 요인, 남북 유사 인식 요인, 교류 및 지원 요인의 4개 요인이 세부 구성요인으로 확인되었다. 통일 필요성 요인과 교류 및 지원 요인이 북한관광에 유의한 영향을 미치는 것을 확인하였으며, 통일의식과 북한관광 인식을 동시에 고려하여 공동 특성을 가지는 독립 집단을 규명하였다. 두 개의 독립된 군집이 확인되었는데 보다 큰 집단이 통일의식과 북한관광에 대해 상대적으로 우호적이었고 작은 집단은 상대적으로 모두 비우호적이었다.

노희경(2019)은 질적 접근을 통하여 한반도 통일 관광에 대한 인식을 파악하고 통일 관광 정책의 방향성을 제시하였다. 참가자들이 인식하는 한반도 통일관광 요인으로는 자원요인, 인프라요인, 문화요인, 여가요인, 안전성요인, 경제요인으로 분류되었으며, 경제적, 문화적 측면에서 매우 긍정적으로 인식하고 있었으며, 강한 기대감을 가지고 있었다. 또한, 북한의 인프라 구축에 있어 남한의 참여에 대한 경제적 기대감을 가지고 있는 것을 확인하였지만 경제적 격차로 인한 갈등, 문화갈등 등에 대해 부정적으로 인식하는 측면도 확인되었다.

이상과 같이 북한관광에 대한 인식 연구는 진행되었다. 현재 북한관광 교류의 재개를 위한 노력이 지속되고 있는 가운데 국민들의 북한관광에 대한 인식을 파악 할 필요(배준호·이재형, 2018; 노희경, 2019)가 있는 시점이라고 판단된다.

## 2. 소셜미디어 빅데이터 분석

소셜미디어는 이용자의 생각과 의견, 경험, 정보 등을 생산하고 확산시키기 위해 사용하는 개방된 플랫폼인 블로그, 유튜브와 소셜 네트워크 서비스인 페이스북, 트위터 등을 의미한다(이동훈 등, 2010). 소셜미디어의 발달과 스마트 기기의 다양화와 보급증가로 많은 형태의 빅데이터가 실시간으로 생산 및 유통되고 있다(정순여 등, 2018). 이와 같은 빅데이터의 분석 기법으로는 데이터마이닝, 텍스트 마이닝, 평판분석, 소셜분석, 클러스터분석, 현실마이닝 등 크게 6

가지가 있다(이동준 등, 2016). 본 연구는 텍스트 마이닝 기법을 활용하고자 한다. 텍스트 마이닝 기법은 최근 질적 데이터의 분석이 가능해짐에 따라 실시간으로 생성되는 대용량의 데이터를 분석할 수 있는 기술이 현실화 되면서 지식 추출 및 트렌드 탐지 그리고 미래를 예측하는 등 다양한 연구가 활발히 진행되고 있다(김정태 등, 2013). 본 연구에서는 텍스트 마이닝의 대표적 기법인 토픽모델링(Topic Modeling)에 속하는 LDA(Latent Dirichlet Allocation)기법을 활용하였다. 토픽모델링은 문서(document)내 말뭉치(corpus) 분석을 통해 각 문서에 잠재된 주제를 통계적으로 파악하는 기법이다. 즉, 문서 간에 나타난 단어 사이의 토픽이 잠재적으로 내재한다는 가정 하에 토픽모델링이 실시된다(김용진·손용훈, 2018; Blei, 2012; Blei et al., 2003; Kim & Baek, 2016; Steyvers & Griffiths, 2007; Wiedemann, 2013). 토픽모델링은 텍스트 분석기법 중 최근 주목받고 있는 방법으로(김용진·손용훈, 2018; Jockers, 2014), 텍스트 데이터에 사용된 단어의 동시사용 패턴을 기반으로 특정 주제, 이슈, 그룹을 자동으로 추출한다(김용진·손용훈, 2018; Battisti et al., 2015; Blei, 2012; Blei et al., 2003; Steyvers & Griffiths, 2007). LDA(Latent Dirichlet allocation) 토픽모델링은 3단계에 걸쳐 진행되며, 하나의 문서에 복수의 토픽이 내재되어 있다는 것을 전제로 문서 내의 여러 단어들의 조합을 토대로 문서의 토픽을 확률적으로 파악하는 통계적 분석 방법이다(Blei et al., 2003). LDA 기법은 다른 토픽모델링에 비해 결과 해석이 용이하고(Blei, 2012) 방대한 비정형 데이터를 축소해 의미 있는 토픽을 도출하는데 장점(Griffiths & Steyvers, 2004)이 있다. 또한, 사회과학자들에게 네트워크 이론은 다양한 학문분야의 사회현상을 설명해주는 것으로 인식되고 있다(이미경, 2018; Borgatti et al., 2009). 사회 네트워크 분석의 장점은 비정형화된 텍스트(text)데이터로부터 구조화된 패턴의 정보를 추출하여 커뮤니케이션 과정의 양상을 체계화할 수 있으며(남인용·박한우, 2007), 연계 전체의 특성으로 연계에 포함된 사회적 행위를 설명하려는 시도라는 점에 의미가 크다(안재섭, 2009). 소셜 네트워크분석은 수 많은 노드(node)들과 그 노드들 사이의 무수히 다양한 관계들로 인해 계산론적으로 접근하는 복잡한 분야로(윤영일, 2016) 소셜네트워크 분석은 기술적 분석(descriptive analysis)과 추론적 분석(inferential analysis)으로 구분할 수 있다. 기술적 분석은 네트워크 상의 지표를 도출하여 해석하는 기능을 하며, 문제해결을 위한 해결책을 제안한다. 추론적 분석은 키워드간의 연관성을 규명하여 관계를 예측하는 기능을 한다(Borgatti et al. 2013; 백승현 등, 2018). 대표적인 네트워크 분석지표로 연결된 구조적 특성을 파악하기 위해서는 밀도(density), 중심성(centrality), 중심도(centralization) 등의 지표가 사용된다(윤영일, 2016; Hansen et al., 2010). 사회과학에서 주로 사용되는 연결망 분석기법인 '텍스트 네트워크 분석'은 '사회연결망 분석'을 커뮤니케이션 메시지에 적용한 것으로서, 텍스트 내에서 메시지를 형성하는 단어들의 구조적 관계를 분석하여 의미를 도출하는 분석방법이다(이수상, 2014). 빅데이터를 활용한 소셜네트워크 분석은 다양한 분야에서 학문적, 실무적으로 활용되고 있으며, 관광 분야에서도 관심이 증가하고 있다.

장미화·윤영일(2016)는 소셜미디어 빅데이터 분석을 통해 캠핑에 대한 정부 정책의 문제점을 확인하고 향후 수립 방향성을 제시하였다. 김경희 등(2016)는 재방문한 중국과 일본관광객을 대상으로 사회연결망을 이용한 서울지역 관광상품의 만족요인과 불만족 요인을 파악하여 실무적 시사점을 제시하였다. 윤영일·하동현(2017)는 소셜미디어 빅데이터를 통한 비즈니스호텔의 인식에 대한 연구를 진행하였으며, 서정아·오익근(2017)는 온라인 빅데이터를 통한 사회연결망 분석을 실시하여 대구의 관광지 이미지를 분석하였다. 김정현 등(2017)는 사회연결망 분석을 통한 축제 연계 관광 프로그램 개발에 관한 연구를 진행했으며, 이대은 등(2017)는 사회연결망 분석을 통한 중국인 개별관광객의 다목적 지 방문 패턴을 분석하였다.

이상과 같이 관광분야에서 소셜미디어 빅데이터를 활용한 연구가 점차적으로 증가하고 있다. 하지만 북한과 관련된 연구에서 소셜 빅데이터를 활용해서 진행한 연구는 송태민(2015)의 연구가 유일하다. 그러나 송태민(2015)은 소셜 빅데이터를 활용한 북한관광에 대한 연구가 아닌 통일인식에 대한 연구를 진행하였으며, 그 결과 핵무기와 천안함은 미국, 중국, 일본, 러시아 순으로 밀접하게 연결된 것으로 나타났다. 또한, 통일대박, 남북공동선언, 정상회담, 휴전의 버즈(댓글)가 많을 때 통일 인식은 찬성확률이 높은 반면 간첩, 핵무기, 천안함의 버즈가 많을 때 통일인식의 반대 확률이 높은 것으로 분석하였다.

북한관광에 대한 국민들의 인식이 반영된 소셜미디어 데이터분석이 연구 가치가 있음에도 불구하고 아직 선행연구가 존재하지 않는다. 본 연구는 소셜미디어 빅데이터를 대상으로 국민들의 북한관광에 대한 인식을 파악하고자 한다.

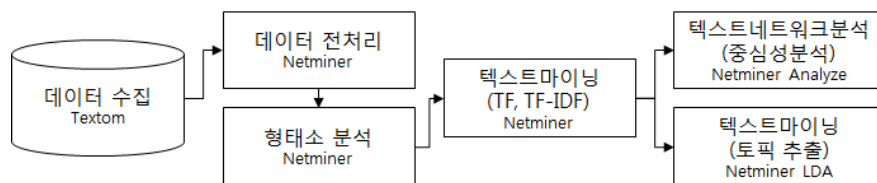
### III. 연구 설계

#### 1. 분석대상 및 자료수집

본 연구는 소셜미디어 빅데이터를 활용하여 북한관광에 대한 국민의 인식을 파악하기 위하여 텍스트 마이닝 빅데이터 분석 솔루션인 텍스트롬(TEXTOM)을 활용하여 데이터를 수집하였다. 데이터 수집을 위한 키워드는 ‘북한관광’으로 국내 최고의 점유율을 가지는 포털사이트에서 제공하는 대표적인 소셜미디어인 네이버 블로그(안효선·이인성, 2016)의 원문 데이터를 수집하였으며, 데이터 수집기간은 김정은 국무위원장의 신년사에서 ‘평창동계올림픽 대표단 파견 및 대화 용의 표명’의 시점(2018남북정상회담 홈페이지)인 2018년 1월 1일부터 IOC, 북한 평창동계올림픽 참가 최종 승인일(2018남북정상회담 홈페이지)인 2019년 1월 31일까지의 기간으로 약 1년을 설정하였으며, 데이터의 중복을 피하기 위하여 엑셀을 활용하여 중복되는 데이터는 삭제처리 후 분석을 진행하고자 한다.

#### 2. 분석방법 및 절차

소셜빅데이터를 활용하여 북한관광에 대한 인식을 파악하기 위해 <그림 1>과 같이 수집된 데이터에 대해 전처리와 형태소 분석(morphological analysis)을 실시한 후 빅데이터 기반의 TF와 TF-IDF 분석, 토픽 모델링의 LDA 기법을 적용하여 분석을 실시하였다. 또한, 텍스트 중 핵심적 역할을 하는 텍스트를 알아보기 위하여 중심성 분석을 수행하였다.



〈그림 1〉 분석방법 및 절차

본 연구에서는 자료의 수집을 위해서 소셜 매트릭스 프로그램인 ‘텍스롬(Textom)’을 활용하였다. 텍스롬은 포털 검색 사이트 자료를 검색하여 데이터 및 연관 키워드 순위를 제공하고 검색 키워드의 공출현 빈도에 따른 매트릭스 정보를 제공하는 네트워크분석 시 활용되는 소프트웨어이다(김보경·김미경, 2015).

다음으로 텍스트 형태로 이루어진 비정형데이터를 자연어처리 알고리즘을 이용하여 분석 가능한 정형 자료 형태로 변환하기 위하여 자동화된 텍스트 분석(automated text analysis) 프로그램인 넷마이너(NetMiner)를 이용하여 자연어처리(Natural Language Processing: NLP) 기능을 활용하여 네트워크 분석이 가능한 형태인 정형화된 자료로 변환하였다. 자료의 정형화 과정에서 형태소 분석을 수행하여 명사를 추출(tagging)하는 과정을 거쳤다. 또한, 유의어(thesaurus), 유사어(defined words), 제외어(exception list)를 지정하는 과정을 반복하여 수집된 데이터 문서를 정제하였다. 정제된 데이터를 활용하여 ‘단어 빈도(Term Frequency: TF)’와 ‘단어 빈도-역문서 빈도(Term Frequency-Inverse Document Frequency: TF-IDF)’를 도출하였다. TF-IDF는 하나의 텍스트 그룹에서 특정 단어의 중요도는 해당 텍스트 내의 출현 빈도에 대해 비례하고, 전체 데이터에 대한 출현 빈도와 반비례하도록 가중치를 산출하는 방법으로(Han & Kamber, 2011) TF-IDF 값을 통해 일반적 단어를 제거하고 유의미한 단어를 추출할 수 있다. 토픽 모델링 기법은 대량의 문서에 존재하는 단어들 간의 연관 관계를 분석하여 문서가 담고 있는 잠재적 의미구조를 가늠해볼 수 있게 해주는 기법이다(박재현 등, 2017). 이때, 토픽은 서로 같이 등장할 확률이 높고 유사한 의미를 가지는 단어들의 집합이라고 할 수 있는데, 본 연구에서는 다양한 토픽 모델링 방법 중에서도 잠재 디리클레 할당(Latent Dirichlet Allocation: LDA) 기법을 활용하였다. LDA 기법은 다른 토픽 모델링 기법에 비해 결과 해석이 용이하고(Blei, 2012), 과적합(overfitting) 문제를 해결하기 때문에 방대한 비정형 데이터로부터 여러 가지 토픽을 도출

하는 데 유리하다(Blei et al., 2003). 특히 토픽모델링은 대량의 텍스트에서 효과적으로 자료를 분석하는 방식으로 동시에 자주 등장하는 단어의 집합을 통해 토픽을 구성하고 제시한다(Steyvers et al., 2007). 마지막으로 텍스트 네트워크분석(Text Network Analysis)은 특정 주제에 대해 어떤 방식으로 어떤 단어를 사용하고 어떤 방식으로 배치했는지 구조적으로 분석할 수 있게 한다(차민경 · 권상희, 2015). 텍스트 네트워크분석을 통한 소셜네트워크 분석은 정형화된 정보를 비정형데이터에서 추출하여(김해원 · 전채남, 2014) 단어들 간 연결 구조 파악 및 연결정도 중심성을 분석하고 관계의 정도를 계량화 한다. 이를 분석하기 위하여 넷마이너(NetMiner)4 를 활용하였다.

## IV. 분석결과

### 1. TF 및 TF-IDF 결과

본 연구에서 사용된 데이터는 2018년 1월 1일부터 2019년 1월 31일까지 네이버 블로그를 대상으로 '북한관광'이 포함된 게시 글을 수집해 36,399건의 문서 데이터를 분석에 활용하였다. 수집된 데이터는 정제과정 및 형태소 분석을 통해 3,306개의 텍스트로 구성되었으며, '단어 빈도(Term Frequency: TF)'와 '단어 빈도-역문서 빈도(Term Frequency-Inverse Document Frequency: TF-IDF)'를 도출한 결과 <표 1>과 같이 분석되었다. 단어 빈도 분석 결과는 금강산, 남한, 중국, 관련주, 여행, 남북, 관광객, 김정은, 개발, 미국 등의 순으로 나타났지만, 단어 빈도-역문서 빈도 분석 결과는 금강산, 남한, 중국, 남북, 관련주, 여행, 관광객, 개발, 미국, 평양 등의 순으로 분석되었다. 이는 앞서 설명한대로 TF-IDF는 하나의 텍스트 그룹에서 특정 단어의 중요도는 해당 텍스트 내의 출현 빈도에 대해 비례하고, 전 체 데이터에 출현 빈도와 반비례하도록 가중치를 산출하는 방법으로 TF-IDF값이 높은 단어는 해당 문서 내에서 핵심적인 메시지를 담고 있을 확률이 높아 어떤 문서에서의 특정 단어의 중요도를 측정할 수 있는 방법으로 활용되고 있다. TF에 비해 TF-IDF가 높은 텍스트는 관광사업, 관광지, 가능, 도시, 관광특구 등으로 나타났으며, TF에 비해 TF-IDF가 낮은 텍스트는 김정은, 도라산, 백두산, 투자, 현대, 자원, 보유, 명소, 강원도 등으로 나타났다. 즉, TF에 비해 TF-IDF가 높게 나타난 텍스트는 전체 데이터에서 차지하는 중요도가 높은 것으로 판단 할 수 있다. 반면, TF에 비해 TF-IDF가 낮게 나타난 텍스트는 출현빈도가 높은 텍스트로 판단할 수 있다.

<표 1> TF 및 TF-IDF 결과

No	텍스트	TF	비율	No	텍스트	TF-IDF	비율
1	금강산	303	6.82	1	금강산	241	8.55
2	남한	178	4.00	2	남한	158	5.60
3	중국	162	3.64	3	중국	115	4.08
4	관련주	144	3.24	4	남북	112	3.97
5	여행	139	3.13	5	관련주	102	3.62
6	남북	128	2.88	6	여행	101	3.58
7	관광객	111	2.50	7	관광객	98	3.48
8	김정은	89	2.00	8	개발	76	2.70
9	개발	85	1.91	9	미국	70	2.48
10	미국	84	1.89	10	평양	65	2.30
11	도라산	81	1.82	11	개성공단	64	2.27
12	평양	79	1.78	12	상품	64	2.27
13	상품	72	1.62	13	통일	62	2.20
14	개성공단	70	1.57	14	김정은	61	2.16
15	통일	69	1.55	15	관광산업	57	2.02
16	여행사	69	1.55	16	여행사	56	1.99
17	관광산업	64	1.44	17	국가	54	1.91
18	재개	63	1.42	18	재개	53	1.88
19	사업	63	1.42	19	사업	52	1.84
20	국가	63	1.42	20	평화	52	1.84
21	평화	62	1.39	21	지역	51	1.81

No	텍스트	TF	비율	No	텍스트	TF-IDF	비율
22	지역	57	1.28	22	방문	48	1.70
23	제재	57	1.28	23	도라산	47	1.67
24	백두산	57	1.28	24	제재	47	1.67
25	방문	54	1.21	25	경제	44	1.56
26	경제	52	1.17	26	관광사업	43	1.52
27	투자	48	1.08	27	사진	42	1.49
28	국제	48	1.08	28	원산	41	1.45
29	현대	47	1.06	29	경협	40	1.42
30	원산	47	1.06	30	관광지	40	1.42
31	경협	47	1.06	31	국제	40	1.42
32	사진	46	1.03	32	사람	40	1.42
33	사람	46	1.03	33	백두산	39	1.38
34	관광사업	46	1.03	34	안보	39	1.38
35	안보	43	0.97	35	문화	38	1.35
36	외국인	42	0.94	36	투자	37	1.31
37	관광지	42	0.94	37	외국인	36	1.28
38	문화	41	0.92	38	현대	36	1.28
39	자원	38	0.85	39	관심	34	1.21
40	철도	37	0.83	40	철도	34	1.21
41	중단	37	0.83	41	중단	33	1.17
42	관심	37	0.83	42	개성	32	1.13
43	단체	35	0.79	43	단체	32	1.13
44	개성	35	0.79	44	자원	32	1.13
45	보유	32	0.72	45	세계	30	1.06
46	세계	31	0.70	46	가능	28	0.99
47	명소	30	0.67	47	대통령	27	0.96
48	기업	30	0.67	48	기업	26	0.92
49	강원도	30	0.67	49	도시	26	0.92
50	대통령	29	0.65	50	관광특구	25	0.89

## 2. 토픽모델링 결과

LDA(Latent dirichlet allocation) 알고리즘을 활용하여 발생확률을 기준으로 연관어를 추출하였으며, 토픽모델링을 위한 alpha 값은 50/토픽수로 8.33, beta 값은 0.1로 지정하였다(Griffiths & Steyvers, 2004). 또한, 샘플링 반복횟수를 1,000번 반복하도록 설정하였으며, 유의미한 결과를 도출하기 위해서는 토픽수의 결정이 매우 중요하기 때문에 토픽 수 결정을 위하여 토픽의 해석가능성과 타당도 및 연구목적과 관련된 유용성에 따라(Nahm, 2016) 4개부터 15개까지 설정하여 비교분석 하는 과정을 거쳐 최종 6개의 토픽으로 분류하였다. 토픽별 비중값이 높은 단어 10개를 추출한 결과 <표 2>와 같으며, 이를 도식화 하면 <그림 2>와 같다.

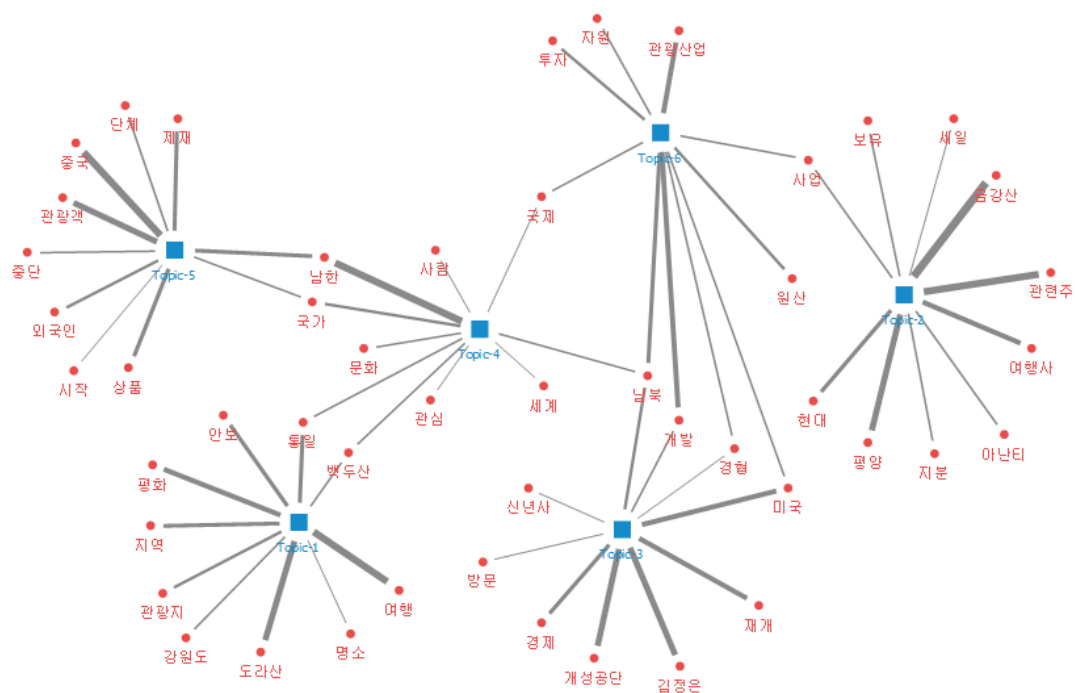
〈표 2〉 토픽모델링 결과

No	Topic-1 남북연계관광	Topic-2 대북관광투자	Topic-3 정치적변화	Topic-4 통일환경변화	Topic-5 관광시장변화	Topic-6 대북자원투자
1	여행	금강산	김정은	남한	중국	개발
	0.057	0.115	0.034	0.051	0.067	0.024
2	도라산	관련주	개성공단	국가	관광객	관광산업
	0.034	0.058	0.028	0.014	0.044	0.021
3	평화	평양	재개	남북	남한	남북
	0.025	0.03	0.025	0.012	0.019	0.02
4	통일	여행사	미국	문화	상품	투자
	0.019	0.026	0.021	0.011	0.017	0.016



No	Topic-1 남북연계관광	Topic-2 대북관광투자	Topic-3 정치적변화	Topic-4 통일환경변화	Topic-5 관광시장변화	Topic-6 대북자원투자
5	지역	현대	경제	백두산	제재	원산
	0.018	0.017	0.017	0.01	0.017	0.015
6	안보	보유	남북	통일	외국인	사업
	0.017	0.012	0.015	0.009	0.014	0.011
7	관광지	사업	개발	사람	단체	미국
	0.013	0.011	0.009	0.007	0.011	0.011
8	강원도	아난티	방문	관심	국가	경험
	0.01	0.011	0.008	0.005	0.011	0.011
9	백두산	지분	경험	세계	중단	자원
	0.009	0.01	0.007	0.005	0.009	0.01
10	명소	세일	신년사	국제	시작	국제
	0.007	0.008	0.006	0.005	0.005	0.009

토픽1은 비중값이 높은 10개의 텍스트를 도출한 결과 여행, 도라산, 평화, 통일, 지역, 안보, 관광지, 강원도, 백두산, 명소로 나타났으며 관광자원으로 명명하였다. 토픽2는 금강산, 관련주, 평양, 여행사, 현대, 보유, 사업, 아난티, 지분, 세일 순으로 나타났으며 대북관광투자로 명명하였다. 토픽3은 김정은, 개성공단, 재개, 미국, 경제, 남북, 개발, 방문, 경험, 신년사 순으로 나타났으며 정치적변화로 명명하였다. 토픽4는 남한, 국가, 남북, 문화, 백두산, 통일, 사람, 관심, 세계, 국제 순으로 나타났으며 통일 환경 변화로 명명하였다. 토픽5는 중국, 관광객, 남한, 상품, 시작, 중단, 외국인, 단체, 제재, 관광객 순으로 나타났으며 관광시장변화로 명명하였다. 마지막으로 토픽6은 개발, 관광산업, 남북, 투자, 원산, 사업, 미국, 경험, 자원, 국제 순으로 나타났으며 대북자원투자로 명명하였다. 토픽이 도출된 텍스트를 도식화 하면 <그림 2>와 같다.



<그림 2> 토픽별 상위 10개 텍스트 도식화



### 3. 중심성 분석 결과

출현빈도 기준으로 선정된 100개의 텍스트가 서로 어떻게 연결되어 있는지 계량화하고 이 텍스트 중 핵심적 역할을 하는 텍스트를 확인하기 위하여 중심성 분석을 실시하였다. 중심성을 측정하는 방법은 연결정도 중심성(degree centrality), 근접 중심성(closeness centrality), 매개 중심성(betweenness centrality), 위세 중심성(eigenvector centrality)을 분석하였다. 연결정도 중심성은 개체 간의 직접적인 관계를 나타내는 지표로 금강산, 남한, 남북, 중국, 여행, 관련주, 개발, 원산, 미국, 평양순으로 나타났다. 근접 중심성은 네트워크의 개체 간 관계에서 가까이 위치하여 개체를 연결하는 준거점의 역할을 측정하는 지표로 금강산, 여행, 관련주, 개발, 중국, 남한, 남북, 관광객, 평양, 방문 순으로 나타났다. 매개 중심성은 네트워크 내에서 개체들 간의 중개 역할을 측정하는 지표로 금강산, 남한, 여행, 중국, 남북, 개발, 관련주, 미국, 상품, 김정은 순으로 분석되었다. 마지막으로 위세 중심성은 네트워크의 개체 간 연결 개수 뿐만 아니라 연결된 노드의 중요도를 측정하는 지표로 금강산, 개성공단, 평양, 원산, 관련주, 지분, 현대, 국제, 관광특구, 관광사업 순으로 나타났다.

〈표 3〉 중심성 분석 결과

순위	연결정도중심성		근접중심성		매개중심성		위세중심성	
1	금강산	0.275	금강산	0.462	금강산	0.310	금강산	0.624
2	남한	0.165	여행	0.414	남한	0.151	개성공단	0.347
3	남북	0.165	관련주	0.414	여행	0.141	평양	0.279
4	중국	0.132	개발	0.414	중국	0.137	원산	0.217
5	여행	0.132	중국	0.410	남북	0.135	관련주	0.204
6	관련주	0.121	남한	0.403	개발	0.101	지분	0.204
7	개발	0.110	남북	0.396	관련주	0.096	현대	0.170
8	원산	0.088	관광객	0.391	미국	0.079	국제	0.159
9	미국	0.088	평양	0.387	상품	0.060	관광특구	0.156
10	평양	0.077	방문	0.382	김정은	0.050	관광사업	0.154
11	사업	0.077	정부	0.379	방문	0.048	여행사	0.145
12	철도	0.066	상품	0.373	공동	0.047	남북	0.136
13	상품	0.066	백두산	0.373	평양	0.046	관광객	0.125
14	방문	0.066	관광산업	0.373	사업	0.045	경협	0.121
15	관광객	0.066	재개	0.371	관광객	0.042	아난티	0.118
16	평화	0.055	경제	0.368	원산	0.042	개성	0.108
17	통일	0.055	기업	0.367	관광산업	0.039	중국	0.103
18	외국인	0.055	시작	0.363	철도	0.037	관광지구	0.082
19	여행사	0.055	미국	0.363	여행사	0.035	재개	0.080
20	김정은	0.055	철도	0.361	국제	0.034	중단	0.078
21	국제	0.055	개성공단	0.361	평화	0.034	개발	0.069
22	관광산업	0.055	원산	0.358	국가	0.033	정상회담	0.067
23	경제	0.055	과거	0.357	통일	0.032	보유	0.063
24	건설	0.055	현대	0.355	경제	0.030	갈마	0.062
25	현대	0.044	통일	0.351	백두산	0.029	과거	0.061
26	투자	0.044	경협	0.351	투자	0.028	세일	0.060
27	지분	0.044	정상회담	0.347	홈페이지	0.022	사회	0.052
28	제재	0.044	외국인	0.345	지역	0.022	방문	0.050
29	재개	0.044	평화	0.343	소개	0.022	지역	0.047

순위	연결정도중심성		근접중심성		매개중심성		위세중심성	
30	안보	0.044	지분	0.343	선언	0.022	백두산	0.045
31	백두산	0.044	아난티	0.343	비핵화	0.022	남한	0.044
32	문화	0.044	지역	0.342	재개	0.018	외국인	0.042
33	국가	0.044	국제	0.338	제재	0.015	투자	0.041
34	관광특구	0.044	관광특구	0.336	대통령	0.013	건설	0.037
35	공동	0.044	공동	0.336	중단	0.013	철도	0.035
36	경협	0.044	개성	0.336	지분	0.013	정부	0.035
37	개성공단	0.044	사업	0.335	시작	0.012	기업	0.035
38	크루즈	0.033	건설	0.335	기업	0.012	여행	0.034
39	지역	0.033	투자	0.333	정상회담	0.011	시작	0.032
40	중단	0.033	크루즈	0.330	경협	0.011	공동	0.031
41	정상회담	0.033	진행	0.329	크루즈	0.010	사업	0.030
42	정부	0.033	여행사	0.329	교류	0.010	관계	0.030
43	전문	0.033	관광지구	0.329	정부	0.010	전문	0.029
44	입진각	0.033	전문	0.327	문화	0.010	진행	0.029
45	아난티	0.033	중단	0.326	관광지	0.010	평화	0.022
46	시작	0.033	강원도	0.325	사회	0.009	상품	0.020
47	사회	0.033	김정은	0.324	현대	0.009	교류	0.018
48	보유	0.033	관광사업	0.324	건설	0.009	김정은	0.017
49	도라산	0.033	국민	0.323	외국인	0.008	통일	0.016
50	대통령	0.033	사회	0.322	전문	0.006	관광산업	0.016

주: 출현빈도 상위 100개의 텍스트에서 연결정도 중심성 순위 50위를 기준으로 제시

## V. 결 론

본 연구는 소셜미디어를 대상으로 빅데이터 분석기법을 활용하여 북한관광에 대한 인식의 핵심요소를 파악하고 구성 개념 간의 관계를 기술적, 탐색적으로 고찰하고자 하였다. 이를 위해 TF 및 TF-IDF값 산출, 토픽분석을 실시하였으며, 중심성 분석을 통해 북한관광 인식에 있어 주요 텍스트 및 관계를 도출하였다. TF 및 TF-IDF값 산출 결과 금강산, 남한, 중국 순으로 텍스트의 빈도 및 TF-IDF값이 높게 나타났으며, 토픽모델링 결과 남북연계관광, 대북관광투자, 정치적 변화, 통일 환경 변화, 관광시장변화, 대북자원투자 총 6개의 토픽으로 분류되어 북한관광에 대한 인식 특성을 확인할 수 있었다. 마지막으로 연결정도 중심성, 근접 중심성, 매개 중심성, 위세 중심성 분석을 통해 핵심 역할을 하는 텍스트를 파악하였다.

이용자의 생각과 의견, 경험, 정보 등을 생산하고 확산시키기 위해 사용하는 소셜미디어를 활용하여 북한관광에 대한 인식을 확인한 결과 다음과 같은 시사점을 도출하였다. 첫째, 남한, 중국, 미국 등 국제적 영향에 대한 인식의 TF 및 TF-IDF값이 높게 산출되었다. 즉, 북한관광을 인식함에 있어 국제적 영향관계에 대해 중요하게 생각하고 있음을 확인할 수 있다. 이는 김도희·허석재(2018)의 연구에서와 같이 국민들은 북한관광에 대해 인식할 때 국제 정치적 상황과 연관지어 생각하고 있음을 확인하였다. 따라서 남북관광교류 재개 및 활성화에 대비하여 남·북, 북·미, 북·중 등 주변국과의 관계에 대한 지속적 노력이 필요할 것으로 판단된다. 또한, 관련주, 관광객, 개발, 여행, 상품 등의 텍스트 TF 및 TF-IDF값이 높게 산출된 것을 보면 북한관광을 인식함에 있어 경제적 효과에 대해 중요하게 인식하고 있음을 확인할 수 있었다. 이는 이상현(2011), 김병로·최영희(2012), 노희경(2019)의 연구와 같이 국민들은 북한관광 활성화 및 통일의 효과로 경제적 효과에 대해 중요하게 인식하고 있음을 재확인하였다.

둘째, 토픽 모델링 결과 남북연계관광, 대북관광투자, 정치적 변화, 통일 환경 변화, 관광시장변화, 대북자원투자를 주제로 북한관광에 대해 인식하고 있음을 확인할 수 있었다. 북한관광에 대해 인식할 때 남북연계관광을 중심으로 인

식하고 있다는 것은 향후 남북 연계 관광 상품 개발 및 활성화를 위해 정부 차원뿐만 아니라 지자체 및 민간 차원의 사전 준비가 필요함을 시사한다. 또한, 핵심요소로 대북관광 및 자원 투자를 주제로 인식하고 있음을 확인할 수 있었으며, 이는 노희경(2019)의 연구와 같은 결과이다. 따라서 북한의 시장개방에 대비하여 남한자본이 차별받지 않고 시장을 선점하기 위한 철저한 준비가 필요하다. 그리고 정치적 변화, 통일 환경 변화, 관광시장변화를 중심 주제로 북한관광에 대해 인식하고 있음을 확인하였다. 이는 북한은 방문이 제한된 목적지로 정치적 영향과 국내외적인 통일 환경의 변화, 관광시장의 변화에 민감하게 반응함을 유추해 볼 수 있다.

셋째, 중심성 분석 결과를 토대로 다음과 같은 시사점을 도출 할 수 있었다. 연결정도중심성이 높게 나타난 금강산, 남한, 남북, 중국, 여행 등은 북한관광 인식에 직접적인 영향을 주는 요소이다. 즉, 남북관광교류 재개 및 활성화에 대한 국민들의 적극적 관심을 이끌어 내기 위해서는 금강산 관광에 초점을 두어야 한다는 사실이다. 근접중심성이 높게 나타난 금강산, 여행, 관련주, 개발, 중국 등은 북한관광 인식에서 다른 구성개념과 어우러져 시너지 효과를 발휘하는 요소가 될 수 있다. 이는 북한관광에 대해 인식 할 때 투자 및 개발 등과 함께 인식이 구성되어 진다는 것을 의미한다. 따라서 국민들의 북한관광에 대한 관심을 이끌어 내기 위해서는 투자 및 개발의 측면도 함께 논의 되어져야 한다는 것이다. 매개중심성이 높게 나타난 금강산, 남한, 여행, 중국, 남북 등은 북한관광을 인식하는데 있어 연결고리로서의 역할 큰 요소로 해석해 볼 수 있다. 위세중심성이 높게 나타난 금강산, 개성공단, 평양, 원산, 관련주는 북한관광 인식에 파급력을 지니는 요소로 해석할 수 있다. 북한관광에 대해 인식할 때 여행적인 요소뿐만 아니라 개성공단, 도시, 경제적 측면까지 함께 인식함을 알 수 있다.

따라서 남북관광교류 재개 및 활성화에 대비하여 다음과 같은 국민들의 인식 반영이 이루어져야 할 것이다. 국민들의 북한관광에 대한 관심을 이끌어 내기 위해서는 투자 및 개발의 측면도 함께 논의 되어져야 하며, 북한의 시장개방에 대비하여 남한자본이 차별받지 않고 시장을 선점하기 위한 철저한 준비가 필요하다. 또한, 북한은 방문이 제한된 목적지로 정치적 영향과 국내외적인 통일 환경의 변화, 관광시장의 변화에 민감하게 반응하므로 내대외적인 정책적 노력이 필요하다. 그리고 국민들은 북한관광을 인식함에 있어 남북연계관광도 함께 인식하고 있다는 것은 향후 남북 연계 관광 상품 개발 및 활성화를 위해 정부 차원뿐만 아니라 지자체 및 민간 차원의 사전 준비가 필요함을 시사한다. 또한, 관광분야에서 관광대상지에 대한 인식 연구를 진행함에 있어 소셜빅데이터를 활용하여 비정형데이터를 객관적·계량적인 정형데이터로 변환하여 연구에 활용할 수 있는 방안을 제시하는 학문적 기여를 하였다. 본 연구의 한계점은 데이터 수집의 대상을 블로그만을 대상으로 하여 향후 연구에서는 다양한 소셜미디어를 분석 대상으로 활용할 필요가 있다. 또한, 빅데이터 분석을 통해 추출된 텍스트를 통해 긍정 또는 부정적 인식을 파악하는 데는 한계가 있다. 향후 연구에서는 감정적 측면 확인할 수 있는 분석을 통해 인식의 긍·부정 관계를 파악해 볼 필요가 있으며, 빅데이터 분석을 통해 추출된 다양한 텍스트를 활용하여 계량적 연구와 연계한다면 명확한 실증연구가 가능할 것으로 판단된다.

## 참 고 문 헌

- 강인원(2006). 관광교류협력사업과 연계지역의 관광발전 영향요인에 관한 연구: 금강산관광사업과 강원도 고성군의 발전 방안을 중심으로. *마케팅과학연구*, 16(4), 53-71.
- 김경희·한학진·김재석(2016). 사회연결망을 이용한 서울지역 관광상품 만족도에 관한 연구: 재방문한 중국과 일본관광객을 중심으로. *호텔리조트연구*, 15(3), 163-182.
- 김도희·허석재(2018). 남·북관계 및 외교·안보현안에 대한 대국민 인식조사. 국회입법조사처.
- 김병로·최경희(2012). 남북한 주민의 통일의식 비교 분석. *통일과평화*, 4(1), 101-139.
- 김보경·김미경(2015). 크루즈 관광 관련 이슈에 대한 키워드 네트워크 분석. *Tourism Research*, 40(4), 1-20.
- 김상태(2018). 한반도 신경제지도 구상과 관광정책. *한국관광정책*, 73, 79-84.
- 김용진·손용훈(2018). 뉴스데이터의 LDA 토픽 분석을 통한 장수군 농촌지역 활성화 사업의 특징: 관광·생활 키워드를 중심으로. *농촌계획*, 24(4), 69-80.
- 김정태·정지형·이승민(2013). ECOsight 기반의 미래기술 전망: 기술, 인물, 사회 통합적 기술예측. *Insight Report*, 13(2), 1-77.
- 김정현·손호기·김상범·채혜성(2017). 사회 연결망 분석을 통한 축제 연계관광 프로그램 개발: 유유누에마을 참뽕 축제를

- 중심으로. *동북아관광연구*, 13(3), 103-122.
- 김해원·전재남(2014). 빅데이터를 활용한 콘텐츠 제작방안에 관한 탐색적 연구: TV홈쇼핑을 중심으로. *사이버커뮤니케이션 학보*, 31(3), 5-51.
- 김홍태(2018). 남북한 평화조성과 체육교류협력의 추진전략. *북한연구학회 하계학술발표논문집*.
- 남인용·박한우(2007). 대권 예비후보자 관련 신문기사의 네트워크 분석과 홍보전략. *한국정당학회보*, 6(1), 79-107.
- 노희경(2019). 질적 접근을 통한 한반도 통일관광에 대한 인식 연구. *호텔관광연구*, 21(1), 70-82.
- 박재현·이강철·전현진(2017). 감성원형모형 필터 기반의 토픽모델링 분석 체계 개발. *한국감성과학회 춘계학술대회*, 9-9.
- 박창수·윤유식·심상진(2006). 남북관광교류정책에 따른 통일의식과 북한관광 만족도 및 방문가치, 재방문의사에 관한 연구. *호텔관광연구*, 8(2), 128-140.
- 배준호·이재형(2018). 한반도관광 활성화를 위한 북한관광 상품 분석: 한국국민의 인식을 중심으로. *관광레저연구*, 30(10), 459-475.
- 백승현·김기탁(2018). 소셜 네트워크의 빅데이터 분석을 이용한 골프연습장 인식변화. *골프연구*, 12(2), 101-117.
- 서정아·오익근(2017). 사회연결망 분석을 활용한 대구의 관광지 이미지 분석: 온라인 빅데이터를 중심으로. *한국콘텐츠학회 논문지*, 17(8), 443-454.
- 송태민(2015). 소셜빅데이터를 활용한 국민 통일인식 동향 분석. *보건·복지 Issue&Focus*, 269(1), 1-8.
- 신용석·김영운·김영운·동용승·이상준(2005). *남북관광 교류 협력 활성화 방안*. 한국문화관광정책연구원.
- 안재섭(2009). 서울시 거주 중국 조선족의 사회·공간적 연결망: 기술적 분석을 중심으로. *한국사지리학회지*, 19(4), 215-223.
- 안효선·이인성(2016). 디자인분야에서 빅데이터를 활용한 감성평가방법 모색: 한복 연관 디자인 요소, 감성적 반응, 평가 어휘를 중심으로. *한국의류학회지*, 40(6), 1034-1044.
- 윤영일(2016). 소셜네트워크분석을 통한 캠핑에 대한 인식 변화 연구. *관광경영연구*, 20(2), 265-288.
- 윤영일·하동현(2017). 소셜미디어 빅데이터 분석을 통한 비즈니스호텔의 인식 연구. *경영교육연구*, 32(3), 379-398.
- 이대은·강상훈·박득희(2017). 사회연결망분석(SNA)기법을 활용한 중국인 개별관광객의 다목적지 방문 패턴 분석. *관광연구저널*, 31(5), 37-48.
- 이동준·원종운·권용장·김미례(2016). 소셜 네트워크 빅데이터 기반 택배업체 고객만족도에 관한 연구. *한국전자거래 학회지*, 21(4), 55-67.
- 이동훈·이민훈·박성민·이준환(2010). 확산되는 소셜미디어와 기업의 新소통 전략. *CEO 인포메이션*, 764(1).
- 이미경(2018). 소셜 빅데이터를 이용한 한국인의 여행트렌드 분석: 가족여행과 나홀로여행을 중심으로. *관광학연구*, 42(10), 111-134.
- 이상현(2011). 남북관광교류에 대한 지역주민의 영향인식이 감정 성향, 만족과 애착도에 미치는 영향. *동국대학교 대학원, 박사학위논문*.
- 이수상(2014). 언어 네트워크 분석 방법을 활용한 학술논문의 내용분석. *정보관리학회지*, 31(4), 49-68.
- 이용화·이해정(2017). 남북관계 전문가 통일의식 설문조사. *현안과과제*, 2017(28), 1-8.
- 이해정·이용화(2018). 남북사회문화협력의 비전과 과제. *현안과과제*, 2018(3), 1-15.
- 장동석(2018). 통일의식과 북한관광 인식에 대한 연구. *관광연구저널*, 32(9), 111-122.
- 장동석·고계성(2014). 북한관광 재개 및 활성화를 위한 전문가 조사 연구. *관광연구저널*, 28(7), 75-86.
- 장미화·윤영일(2016). 소셜 미디어 빅데이터 분석을 통한 캠핑에 대한 정부 정책과 국민들의 인식 변화 연구. *관광연구*, 31(1), 91-112.
- 정동준·김선·김희정·나용우·문인철·송영훈·최규빈·임경훈·이정옥(2019). *2018통일의식조사*. 서울대학교 통일평화연구원 통일학연구.
- 정순여·김대훈·양인하(2018). 제주도 들불축제에 대한 소셜미디어 사용자들의 인식에 관한 연구: 빅데이터 분석을 통한 접근. *탐라문화*, 59(1), 157-186.
- 차민경·권상희(2015). '창조경제'에 대한 네트워크 의제 설정 경향 의미연결망 분석: 2008년부터 2014년까지 소셜 미디어 데이터 분석을 중심으로. *한국소통학회 학술대회논문집*, 10월, 90-95.
- 최의철(2000). *남북한 교류·협력 활성화 방안*. 통일연구원.
- 통일부(2018). *통일백서*. 통일부.
- 통일부 홈페이지(2018). <https://www.unikorea.go.kr/unikorea/business/cooperation/status/history>

- 홍순직(2014). 남북관광 협력 추진 방향. *한국관광정책*, 71, 60-64.
- 황수환(2018). 평화와 관광: 금강산 관광사업의 평가. *북한연구학회 하계학술발표논문집*, 6월, 191-203.
- 황인하(2009). 북한관광의 서비스품질이 관광객 만족 및 행동의도에 미치는 영향. *호텔관광연구*, 34(1), 279-296.
- Battisti, F. D., Ferrara, A., & Salini, S.(2015). A decade of research in statistics: a topic model approach. *Scientometrics*, 103, 413-433.
- Blei, D. M., Ng, A. Y., & Jordan, M. I.(2003). Latent dirichlet allocation. *Journal of Machine Learning and Research*, 3, 993-1022.
- Blei, D. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84.
- Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G.(2009). Network analysis in the social sciences. *Science*, 323(5916), 892-895.
- Borgatti, S. P., Everett, M. G., & Johnson, J. C.(2013). *Analyzing Social Networks*, Thousand Oaks, California : SAGE Publications Inc.
- Griffiths, T., & Steyvers, M. (2004). Finding scientific topics. *Proceedings of the National Academy of Sciences*.
- Han, J. & M. Kamber,(2011). *Data Mining: Concepts and Techniques, 3rd edition*. Morgan Kaufmann Publishers.
- Hansen, D., Shneiderman, B. & Smith, M.A.(2010). *Analyzing social media networks with NodeXL: Insights from a connected World*. Morgan. Morgan Kaufmann.
- Jockers, M. L.(2014). *Text analysis with R for students of literature. Switzerland*: Springer International Publishing.
- Kim, J., & Baek, S.(2016). Analysis of Issues on the College and University Structural Reform Evaluation Using Text Big Data Analytics. *Asian Journal of Education*, 17(3), 409-436.
- Nahm, C. H.(2016). An illustrative application of topic modeling method to a farmer's diary. *Institute of Cross Cultural Studies*, 22(1), 89-135.
- Steyvers, M. & Griffiths, T.(2007). Probabilistic Topic Models. *Handbook of latent semantic analysis*, 427(7), 424-440.
- Wiedemann, G. (2013). Opening up to big data: computer-assisted analysis of textual data in social sciences. *Historical Social Research*, 38(4), 332-358.