

## 빅데이터 분석을 통한 한국관광 인식에 관한 연구\*

A Study on Awareness of Korea Tourism through Big Data Analysis

오익근\*\* · 이태숙\*\*\* · 전채남\*\*\*\*

Oh, Ick-Keun · Lee, Tae-Sook · Chon, Chae-Nam

**요약** : 본 연구에서는 관광객들의 관광인식을 SNS 빅데이터를 통해 분석하고자 2014년부터 1월 1일- 12월 31일까지 1년간의 네이버, 다음, 구글의 '관광/여행' 키워드를 포함하고 있는 웹페이지, 개인들이 운영하는 블로그, 공신력 있는 정보를 제공하는 뉴스 등을 수집하여 분석하였다. 자료수집은 데이터 수집 및 처리 프로그램인 TEXTOM을 활용하였고, 데이터는 UCINET와 함께 패키지화된 NetDraw를 활용하여 연결정도 중심성, 근접 중심성, 매개 중심성, 그리고 아이겐벡터 중심성 등을 분석하였다. 분석결과 해외 주요 관광국가 및 도시, 국내 주요 관광도시 등의 웹가시성이 높은 것으로 파악되었으며, 가족여행, 자유여행, 해외여행, 신혼여행 등의 관광행태와 숙박 관광과 2박 3일, 3박 4일 등의 여행일정도 높게 나타났다. 이 외에도 안전, 관광주간, 관광두레, 해미읍성이 비중있게 포함되어 2014년의 사회 환경적 특성을 확인할 수 있었다. 이러한 소셜미디어를 활용한 국내 관광 환경 변화에 대한 관광인식 진단은 실현가능한 관광마케팅 전략 수립에 유용한 기초자료가 될 수 있을 것으로 기대한다.

**핵심어** : 관광인식, 사회네트워크 분석, 빅 데이터, 시멘틱네트워크 분석

**ABSTRACT** : The objective of this study is to examine tourists' awareness of tourism through SNS big data. For this, this study collected data containing 'tourism/travel' keywords from web pages, blogs and news on Naver, Daum and Google during one year from January 1 through December 31, 2014. Data were collected by using TEXTOM, a data collecting and processing program and from those data, degree centrality, closeness centrality, betweenness centrality and eigenvector centrality were analyzed by utilizing packaged NetDraw along with UCINET. As a result of the analysis, it was found that the web visibility of foreign and domestic tourist destinations was high. In addition, the result showed that the web visibility was also high for tourism types such as family trip, free-style trip, overseas trip and honeymoon; overnight tourism; and tour itinerary such as two nights and three days, three nights and four days. Except this, tourism safety, tourism week, Tour Doorae(tourism cooperative group) and Haemi-eup Fortress were included and social and environmental tourism characteristics in 2014 were identified. It is expected that this diagnosis on awareness of tourism according to changes in domestic tourism environment by using these social media will be a foundation of baseline data useful for establishing tourism marketing strategies.

**Key words** : Tourism perceptions, Social network analysis, Big data, Semantic network analysis

\* 이 논문은 2014년 정부(교육부) 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2014S1A5A2A03065863)

\*\* 계명대학교 호텔관광학과 관광경영전공 교수, e-mail: ickoh@kmu.ac.kr

\*\*\* 순천제일대학교 호텔관광과 교수, e-mail: tseyi1010@hanmail.net

\*\*\*\* (주)더아이엠씨(THE IMC) 대표, 경영학 박사, e-mail: cnjun@theimc.co.kr

## I. 서 론

국내 관광 트렌드는 공급 측면이든 수요 측면이든 모두 양(quantity) 중심에서 질(quality) 중심으로 변화가 일고 있으며, 이에 따른 가치중심 관광을 위한 종합적인 지원체계가 필요하다(한국문화관광연구원, 2012). 뿐만 아니라 끊임 없는 미래관광 트렌드 변화를 파악하고 이를 위한 방안마련의 노력도 필요하다. 하지만, 지금까지는 현안대응에 급급한 나머지 미래 트렌드 변화에 기반한 장기적이고 미래지향적인 전략방향 제시가 미흡한 것으로 파악된다. 이를 위해서는 환경변화를 고려한 관광발전 패러다임 모색이 필요하며 그 일환으로 과학적인 증거 기반의 관광 서비스 R&D 토대가 뒷받침되어야 한다. 그동안의 관광시장 조사 프레임이나 분석 접근체계로는 관광객의 다양한 성향이나 행태들을 파악하는데 한계가 따랐다. 이에 본 연구에서는 한국이 세계 최고의 정보통신 기술력(ICT)과 인프라를 보유하고 많은 데이터를 생산, 소비하고 있음에 착안하여 관광 관련 대용량 데이터에 주목하였다. 관광산업에서도 빅 데이터에 대한 연구와 활용, 그리고 관련 기업과 정부까지 큰 관심을 보이고 있다(이양환, 2013). 우리가 가진 가장 큰 장점이 초고속 인터넷 환경과 높은 스마트폰 보급률에 기인한 데이터 생성과 보유에 있음을 고려할 때 관광 빅 데이터에 대한 접근은 시의적절하다고 보여진다. 현재 쌓아 놓은 정보를 어떻게 활용하는가에 따라 미래 관광산업의 흐름이 달라진다는 인식이 비즈니스계 뿐 만 아니라 공적영역에까지 퍼짐으로서 빅 데이터는 관광기업의 경쟁원리나 정부 거버넌스(governance) 차원에서 경쟁력 확보의 필수적 과제로 보여진다(이양환, 2013).

이처럼 수없이 널려있는 정보를 분석해 새로운 정보를 만들고 현재 상황을 분석하며, 앞으로의 트렌드를 예측하는데 유용한 빅 데이터에 체

계적으로 접근한다면, 관광객들의 감성에 다가갈 수 있는 새로운 길을 찾아낼 수 있지 않을까? 빅 데이터는 이러한 가능성에 대한 해답을 줄 수 있을 것으로 보여진다. 이러한 맥락에서 본 연구에서는 국내 관광환경에 대한 관광객들의 관광인식을 빅데이터를 통해 분석하고자 한다. 구체적으로는 인터넷 포털 사이트와 SNS를 중심으로 한 한국관광에 대한 인식을 살펴보고자 한다. 현대 사회는 1년을 예측할 수 없을 정도로 빠르게 진화하고 변화하고 있다. 관광산업을 둘러싼 국내외 환경 또한 급격히 변화하고 있다. 이러한 측면에서 소셜미디어를 통한 관광환경 변화에 대한 관광인식 진단은 현실성이 제고된 관광마케팅 전략 수립에 기초자료의 근간이 될 수 있을 것으로 본다. 뿐만 아니라 온라인의 다양한 자료원에서 생산된 빅 데이터를 관광연구에서 어떻게 활용할 것인지에 대한 방향성 제시와 한국어 최적화 빅 데이터 분석 솔루션 적용을 통해 기술적인 한계로 빅 데이터를 수집·분석하지 못하던 문제를 해결할 수 있는 지침이 될 수 있을 것으로 기대한다. 더불어 국내관광에 대한 인식조사를 빅 데이터를 통해 규명하는 것은 관광 데이터 분석 패러다임의 변화를 도모하고, 데이터 활용의 혁신적 단초를 제공하며, 나아가 관광학 연구의 학문적 지평 또한 넓힐 수 있을 것으로 기대한다.

## II. 이론적 고찰

### 1. 관광정책에 따른 관광인식 패러다임

1960년대 이후부터 현재까지의 국내 관광정책에 대해 한국문화관광연구원(2012)에서는 발전 단계별 발아기(1960-1979), 형성기(1980-1996), 확충기(1997-2004), 성장기(2005-2012)로 분류하였다. 시기별 내용을 살펴보면, 먼저, 발아기에는 수출 위주의 경제성장 중심의

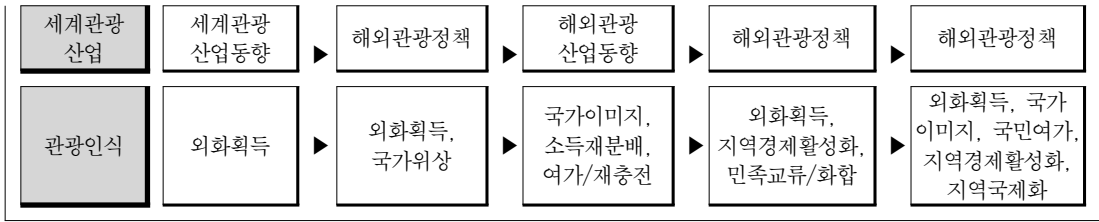
국정가치에 영향을 받아 외화획득을 위한 정책이 수립되었고, 관광을 국제관계 개선을 위한 정책 수단으로 파악하여 국제회의를 적극적으로 개최하였다. 다음으로 형성기의 경우 정치·경제·사회적인 급변속에서 국제관광 진흥정책에서 국민관광 진흥정책으로 정책기조가 변경되었으며, 확장기에는 관광의 경제적 효과뿐 만 아니라 사회·문화적인 효과가 부각되었고 국민관광 활성화, 남북관광, 지역관광 활성화 등의 관광정책 대상 범위가 확대되었으며, 관광을 신성장 동력산업으로 육성하여 관광산업의 경쟁력 강화를 위한 제도개선을 시작하는 등의 관광산업 환경개선이 이루어졌다. 끝으로 성장기에는 관광산업의 외연적 성장뿐만 아니라 관광산업 환경의 질적개선이 이루어졌으며, 국제관광정책과 국민관광정책이 균형

적으로 추진되었다(한국문화관광연구원, 2012). 결과적으로 한국 관광정책은 발아기에서 성장기를 거치면서 크게 경제적 가치 중심에서 경제·사회·문화적 가치로, 소극적 국제관광 진흥정책에서 적극적 국제관광 발전정책으로, 중앙중심의 관광정책에서 지방·분권정책으로, 규제중심 정책에서 탈규제·지원정책으로 변화한 것으로 나타났다(한국문화관광연구원, 2012). 이외에도, 1960년대부터 2000년대까지의 한국관광 이슈 및 인식에 대해 분석한 연구결과가 눈길을 끈다. 민경익·이태숙(2012)은 신문기사를 소재로 1960년대부터 2000년대까지의 한국관광 인식에 대해 내용분석을 하였으며, 그 결과에 대한 흐름을 다음과 같이 제시하였다.

이 연구의 결과를 연대별 살펴보면, 먼저,

〈표 1〉 한국관광 인식 변화(1960~2000년대)

	1960년대	1970년대	1980년대	1990년대	2000년대
국제관광	방문단활동	방문단활동, 기생파티	관광형태 다양화	관광교류/협정, 해외여행 자유화	테마관광, 인센티브관광, 의료관광
국민 국내관광	단체여행	휴가철/연휴 관광	명소관광	선심관광, 관광인파	테마관광
관광개발	관광기반 기초인프라	관광단지/ 관광시설	관광단지 (테마관광지)	관광단지, 관광시설, 관광상품	테마관광지, 테마관광상품
관광조사	출입국/ 관광수지	출입국/ 관광수지	출입국/ 관광수지	출입국/ 휴가계획/ 관광산업매출	지자체 관광통계
관광정보	관광 행사 유치/개최	관광관련행사, 패키지상품	관광지/ 관광상품소개	관광지/관광 상품 다양화	테마관광상품 (체험관광)
관광행정	외화획득위한 관광정책	외화획득위한 관광정책	외화획득/ 국민관광 활성화 정책	외래관광객 유치, 국가이미지	관광상품 브랜드화
관광산업	관광산업 발전과제 (인프라추진)	관광마케팅, 관광산업 발전과제 (인프라추진)	관광산업 발전과제 (서비스품질)	관광기업, 관광산업실태	관광마케팅 (지자체 전략산업)
북한관광		대표단교류	북한관광 개방/정책	금강산관광, 북한관광정책	금강산/ 개성관광



자료: 민경익·이태숙(2012). 미디어 분석을 통한 한국관광 인식 변화에 관한 연구: 신문기사를 중심으로. 관광학연구, 36(10), 81-104.

1960년대는 관광산업을 외국인 유치를 통한 외화획득산업으로 간주하였고, 따라서 대부분의 관광정책이나 개발 목적이 외국인 유치에 초점이 맞추어졌다. 해외 주요인사의 방한 및 내국인의 해외여행 활동이 특별한 '사건'으로 보도되기도 하였다. 1970년대도 마찬가지로 관광산업의 목적 및 관광산업에 대한 인식은 외국인 관광객 유치를 통한 외화획득이라는 관점이 지배적이었다. 특히 관광요금 등을 정부의 규제 하에 일원화하기도 하였다. 1980년대에는 관광 관점이 보다 다양화 된 시기로, 국가이미지 및 국가위상이라는 측면에서 관광을 바라보며 국민의식을 촉구하였으며, 특히 관광개발에 있어 외래관광객을 위한 관광지, 수용시설 등에서 국민 국내관광 수용 인프라 구축에 관심이 높아져, 테마관광지 등의 개발계획이 시작되었다. 한편 국민 국내관광 활성화가 시작되면서 사치 및 향락이라는 관광산업에 대한 부정적 인식도 발생하였다. 1990년대는 지방자치제도가 시행됨에 따라 지역별로 테마관광지 조성, 문화관광상품 개발 등의 지역경제 활성화를 위한 관광 역할이 급부상하였으며, 특히, 금강산관광사업이 시작됨에 따라 관광을 통한 '민족교류' 및 '민족화합'이라는 남북관계의 특수적 상황에 배경을 둔 인식이 나타났다. 2000년대는 지역경제 활성화 및 지역의 국제화라는 관점에서의 지역전략산업으로서의 관광산업에 대한 접근이 나타나기 시작하였다. 특히 2000년대에는 한류관광, 문화관광 등과 더불어 MICE, 의료관광 등을 집중 육성하려는 노력이 전개되었

다. 이렇듯 관광산업이 지역의 전략산업으로 부상함에 따라 각 지자체의 관광마케팅 활동 및 관광객 유치를 위한 경쟁이 심화되고 있음이 나타났다.

결과적으로 이 연구에서 분석된 한국관광 이슈 및 인식도 결국 국내 관광정책의 흐름과 그 맥을 같이 함을 알 수 있다. 다양한 관광정책들은 많은 관광환경 변화를 도모하며, 관광 환경변화는 결국 국내관광에 대한 인식과 개념 변화에 영향을 주고 있음이 확인된다. 뿐만 아니라 이제는 보는 관광을 뛰어 넘고 체험형 관광을 지나 여러 장르와 융복합으로 가는 관광으로 변모하고 있다. 이처럼 시대별 변화를 거듭하고 있는 관광 인식에 대한 파악은 중요하며, 필요하다고 보여진다. 이에 본 연구에서는 최근의 관광환경에 대한 관광인식에 대해서도 조사해 보고자하며, 이를 위해 빅 데이터를 활용하고자 한다.

## 2. 빅 데이터 개요

최근 빅 데이터가 IT 트렌드의 메인 이슈로 부상하고 있다. Gartner는 빅 데이터를 '미래 경쟁력을 좌우하는 21세기의 원유'라고까지 표현할 정도로, 전 세계는 빅 데이터의 성장 잠재력과 가치에 주목하고 있다(배동민·박현수·오기환, 2012). 빅 데이터는 규모가 방대한 자료(data)를 의미하는 용어로 '대용량 데이터를 활용, 분석하여 가치 있는 정보를 추출하고, 생성된 지식을 바탕으로 능동적으로 대응하거나 변화를 예측하기 위한 정보화 기술'로 정의되고 있

나 아직 합의된 내용은 없는 실정이다(강만모·김상락·박상무, 2012; 이양환, 2013). 하지만, 빅 데이터라는 개념이 최근에 갑자기 나타난 것은 아니다. 대용량 데이터는 예전부터 있었으나 약 10년 사이에 검색엔진의 급격한 진화와 소셜 네트워킹 사이트(social networking site)의 확산, 클라우드(cloud) 서비스의 본격화 등이 빅 데이터의 분석과 활용을 서두르게 하였다(이양환, 2013). 이러한 빅 데이터 개념에 대한 논의를 살펴보면, Gartner(2012)는 '향상된 시사점(insight)과 더 나은 의사결정을 위해 사용되는 비용 효율이 높고, 혁신적이며, 대용량, 고속 및 다양성의 특성을 가진 정보 자산'이라고 하였고, 맥킨지(2011)는 데이터 베이스의 규모에 초점을 맞추어 '일반적인 데이터베이스 SW가 저장, 관리, 분석할 수 있는 범위를 초과하는 규모의 데이터'라고 하였으며, IDC(Internet Data Center, 2011)는 데이터베이스가 아닌 업무 수행에 초점을 맞추어 '다양한 종류의 대규모 데이터로부터 저렴한 비용으로 가치를 추출하고 데이터의 초고속 수집, 발굴, 분석을 지원하도록 고안된 차세대 기술 및 아키텍처'로 개념화하였다. 이외에도 Douglas(2012)는 Gartner Report를 인용해 다양한 형태의 대량 데이터가 실시간으로 형성되어 초대용량(high-volume), 빠른 생성속도(high-velocity), 높은 다양성(high-variety)을 갖는 정보자산이라 정의하였고, O'eilly(2012)는 자료가 너무 크고, 빠르게 변화하여 전통 DB에서 처리하기 힘든 데이터라고 정의하였다. 최근에는 빅 데이터가 소셜미디어와 정보기술의 발달로 더욱 빠르게 전

파되고 변화되며, 비정형적이고, 일정한 패턴의 발견이 어렵게 되면서 가치 창출의 중요성이 더욱 강조되었다. 이에 Douglas(2012)가 제시한 정의에 최고의 가치(high-value) 특성이 추가된 정의도 사용되고 있다(이영진·윤지환, 2014). 국내에서는 삼성경제연구소(2012)에서 과거의 관리 및 분석 시스템으로는 처리할 수 없는 방대한 데이터의 집합이라 하였으며, 함유근·채승병(2012)은 구체적인 수치를 활용해 수 천 테라바이트(TB)의 크기를 가지고 있는 비정형 데이터로, 생성과 이용이 짧게는 몇 초, 길게는 몇 시간 단위로 발생하여 기존의 방식으로 처리하기 어려운 데이터 집합이라 정의하였다.

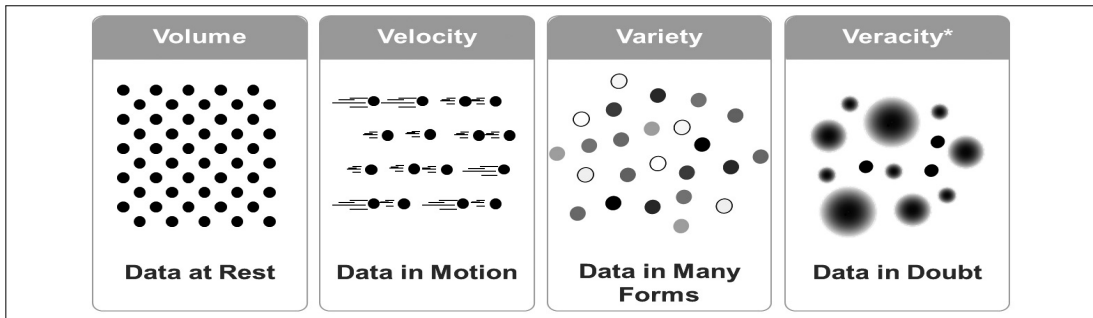
IBM(2011)은 최근 빅 데이터의 특징으로 모호함을 의미하는 Veracity(진실성)의 'V'를 추가하여 [그림 1]에서 보는 바와 같이 4가지 'V'로 규정하였다. 여기서 진실성(Veracity)은 데이터의 모호성, 불완전성 및 데이터간의 불일치 등으로 인한 데이터의 불확실성(uncertainty)을 의미한다고 한다.

이렇게 다양하게 정의되고 있는 빅 데이터의 활용가치가 주목받게 된 것은 World Economic Forum (2012)에서 빅 데이터를 국제개발의 새로운 가능성을 여는 중요 기술의 하나로 지목하면서 부터라 회자된다(이영진·윤지환, 2014). 빅 데이터의 활용가치가 주목받게 되면서 세계 각국의 정부와 기업들은 국가와 기업의 성패를 가름할 새로운 경제적 가치의 원천이 될 것으로 기대하며 다양한 부문에서 빅 데이터의 활용을 시도하고 있다(한국정보화진흥원, 2012). 해외

〈표 2〉 빅 데이터의 정량적 요건

구분	내용
양(Volume)	수십 테라바이트 이상
종류(Variety)	형식을 갖추고 정리된 구조적 데이터보다 구조화되어 있지 않은 비정형 데이터가 90% 이상
속도(Velocity)	데이터의 생성, 처리, 유통, 소비되는 속도가 수 초에서 수 시간 단위

자료 : 한형상·이창호(2014). 빅데이터 분석의 현황과 발전 전략. PD 이슈리포트, 6. KEIT



자료 : 한형상·이창호(2014). 빅데이터 분석의 현황과 발전 전략. PD 이슈리포트, 6. KEIT

〈그림 1〉 IBM의 Big Data 정의

주요국보다는 늦지만 우리나라도 2011년 말 부터 정부 차원에서 빅 데이터에 관한 논의가 시작 되었다. 관련 정부부처에서 정책과제로 빅 데이터를 채택하였으며, 연구소 및 포럼 등이 활발히 설립되었다. 2012년 11월에는 '빅 데이터 마스터 플랜' 안이 발표되면서 빅 데이터 진흥의 청사진을 마련하였다. 이러한 국내 빅 데이터의 시장은 국민들의 신기술 수용력이 높고, IT 인프라가 잘 구축되어 데이터 생산과 1차 활용이 활발하며, 사회 정보화를 추진하면서 데이터 축적 기반이 잘 조성되어 있는 실정이다. 하지만, 방대한 공공 데이터 공개가 선진국 대비 상대적으로 늦은 편이고 실제 활용을 위한 인력과 인프라(SW, HW), 전문성 축적이 부족하며, 데이터 활용에 대한 인식도 부족하고, 빅 데이터 활용을 저해할 수 있는 법/제도의 정비도 미진한 실정이다.

### 3. 관광 분야에서의 빅 데이터 논의

맥킨지(McKinsey)는 공공행정, 소매, 제조, 개인정보 등 다양한 분야에서 빅 데이터가 적용될 것이라 추정하였다(송태민, 2012). 이러한 추세 속에서 하버드 경영대학원의 Davenport Thomas 교수는 여행 업계가 빅 데이터에서 혜택을 얻을 가능성이 아주 크다고 밝혔다.

빅 데이터는 특성상 로그기록, 소셜, 위치, 소비, 현실 데이터 등 비정형화된 데이터 유형의

다양화로 맞춤형 관광서비스를 제공할 수 있는 여건을 마련해 준다. 자신의 관광경험에 대해 표출된 감정, 인식, 체험결과 등의 수집과 분석이 가능해져 이용자별 특성과 선호를 고려한 맞춤형 서비스 제공이 가능하다. 고객의 트렌드를 정확히 안다는 것은 무엇보다도 중요하며, 정보들의 대부분은 모바일을 통한 데이터에서 비롯된다고 보아도 과언이 아니다. 이미 유용하게 사용하고 있는 빅 데이터의 처리, 분석 등이 관광부문에 도입되고 있다(변정우, 2013). 적용 사례가 많지는 않지만, 빅 데이터의 미래 추세 예측과 마케팅 활용도가 높기 때문에 그 사례들이 속속 제시되고 있다. 그 사례를 살펴보면, 먼저 인터넷 의존 비중의 높은 분야가 여행 정보임에 착안한 여행정보 사이트인 '트립 어드바이저(tripadvisor)'는 개인의 텍스트와 이미지 리뷰 등 방대한 비정형 데이터를 분석하였다. 그 결과 여행 목적별(유적지 여행, 휴양지 여행, 도심형 휴양 여행, 쇼핑여행 등), 지역별(근거리 여행, 원거리 여행), 취향별(시간이 걸리더라도 저렴한 여행, 비싸더라도 편한 여행, 버스나 기차 선호여행, 짧은 시간에 많은 곳 방문하는 여행, 한 곳에서 오래 머무르는 여행) 개인 선호 스타일의 결과를 도출하였다. 이를 통해 비슷한 지역의 비슷한 연령대가 선호하는 여행형태를 파악할 수 있었으며, 또한 각 여행자의 구미에 맞는 여행상품

을 제안하고 자세한 여행정보 함께 제공하였다. 이는 여행자들이 현지에 인력을 파견하여 얻는 한정적인 여행정보와 전 세계인들이 매일 쏟아붓는 정보의 양과 질에 분명한 차이가 있음을 시사해 주는 결과이다(삼정 KPMG 경제연구원, 2012).

국내의 경우 한국관광공사에서는 '빅 데이터'를 활용하여 전국 문화관광축제의 개최효과를 심층 분석하였다(한국관광공사, 2014). 기존 관광분야 조사·연구의 한계를 보완하고 지방 축제와 관련한 국민들의 관광행태 및 특성을 구체적으로 파악하기 위해 추진된 '빅 데이터' 분석은 2013년 개최된 전국 16개 우수 지방축제를 시범 대상으로 2013년 11월부터 진행되었다. 특히, 이 분석에서 이동통신사의 '빅 데이터'를 관광산업에 접목한 것은 세계적으로도 전례를 찾기 힘든 시도일 뿐만 아니라 유동인구 분석부터 매출분석, 소셜 미디어 분석 등 국내에서 시도되고 있는 다양한 '빅 데이터' 분석기법을 일시에 적용했다는 점에서 높이 평가받았다. 이 분석은 이동통신사가 보유한 이동통신망 데이터 및 빅 데이터 분석 인프라를 기반으로 진행되었으며, 분석에 활용된 전체 '빅' 데이터는 DVD 51만장(2,200 테라바이트) 분량에 달하는 것으로 나타났다. 이를 통해 시간대별, 성·연령별, 지역별 등 다각적인 외부 유입인구 분석과 매출 변동 현황, 소셜 데이터 등의 분석도 병행되어 상세한 정보까지 보여줌으로서 세부적인 축제 개최효과를 제시하였다.

또한 부산 해운대구에서는 '빅 데이터'를 접목한 관광정책을 실시하였다. '레디안 6'라는 자료 수집기를 통해 5개월간 트위터와 페이스북, 블로그 등 소셜미디어에 '해운대'라는 키워드가 언급된 소셜 정보를 수집·분석하였다. 약 3만 8,500여개의 글 가운데 관련성이 높은 1만 여건을 분석한 결과, '해운대 해수욕장'과 '달맞이길' 등 자연경관에 만족 수준이 높음이 발견되었으

며, 반면, '비싼 숙박요금', '교통 체증', '복잡한 버스노선' 등이 부정적인 요소임이 확인되었다(NSP 통신, 2013. 8. 22; 문정욱, 2013). 이를 통해 해운대구는 관광객의 정확한 수요와 장애요인 등을 파악하여 보다 과학적인 관광정책과 서비스 개선을 추진하는 계기를 마련하였다. 이외에도 제주도는 '빅 데이터를 활용한 스마트 관광정보 서비스' 추진을 통해 빅 데이터 서비스 플랫폼을 구축하였다(연합뉴스, 2013. 9. 12; 문정욱, 2013). 제주도의 빅 데이터 기반 관광서비스는 관광객과 관광지, 기상 등 관련 빅 데이터를 수집, 분석하여 실시간으로 이용자에게 제공하는 시스템으로, 관광객이 모바일에 스마트여행 길잡이 애플리케이션을 설치하면 해당 관광객의 동선과 정보가 자동으로 빅 데이터 분석 시스템에 전송되도록 하여 보다 풍부한 분석결과를 얻게 된다. 현실적인 관점에서는 관광산업체 보다 카드사에서 관광객 정보를 활용하여 컨설팅에 활용하는 경우가 있다. 신용카드사는 한국문화정보센터와 제휴를 맺고 국내외 관광객의 이용행태 등에 대한 광범위한 자료를 분석해 관광정책 수립의 기초자료를 제공하기도 하였다(이영진·윤지환, 2014).

한편, 빅 데이터를 활용한 학술적 논의도 이어지고 있다. 하지만, 관련 연구들이 외식이나 호텔(공효순·송은지, 2013; 도해용·류기상, 2013; 도해용·이애주·이창원, 2013), 온라인 관광정보의 공공성 분석(이영진·정호권·부석현, 2010), 관광 분야에서의 SNS 빅 데이터 활용방법 모색(이영진·윤지환, 2014) 등에 그치고 있어 여전히 빅 데이터의 활용도는 낮은 것으로 파악된다. 국내 관광 부문에서의 빅 데이터 활용도가 낮은 이유는 그동안 대량의 자료를 수집해 온 정부 등 공공기관에서 축적한 2차 자료 사용이 부족하였을 뿐만 아니라(국내 공공기관의 원천 데이터 공개는 최근 시행), 빅 데이터 분석을 위한 솔루션이나 소프트웨어 활용 기술이 부

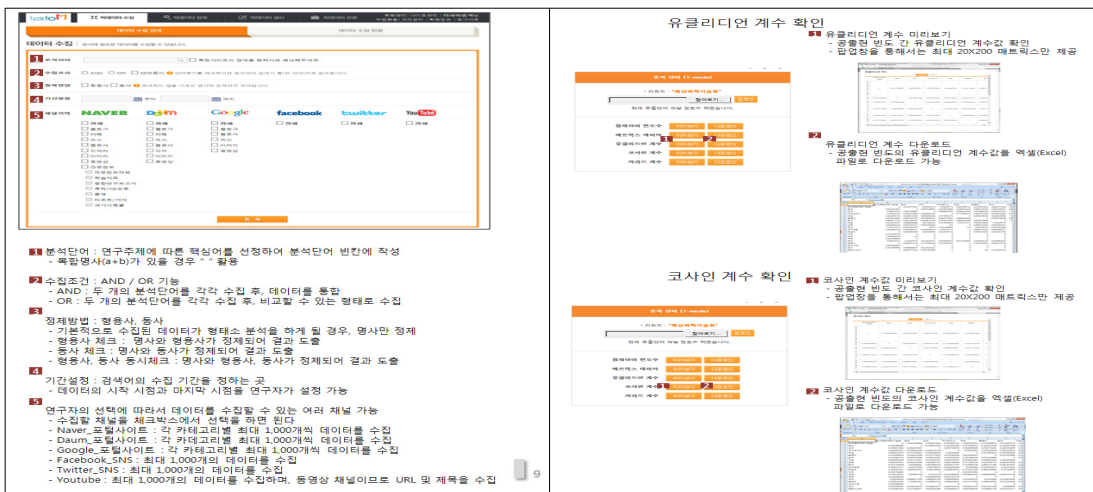
족했기 때문에 판단된다(이영진·윤지환, 2014). 하지만 최근 정부 뿐 만 아니라 포털사이트 등에서도 한정적이지만 SNS 등에서 수집한 자료를 부분적으로나마 공개하고 있다. 따라서 이제는 데이터 수집의 문제가 아니라 수많은 데이터를 어떻게 의미있는 정보로 전환할지에 대한 고민이 필요하다(이영진·윤지환, 2014). 이러한 맥락에서 본 연구에서는 기술적 측면에서의 빅 데이터 활용 가능성에 대한 탐색과 분석 가능한 소프트웨어 개발을 통해 관광분야 데이터 활용의 가치에 토대를 둔 데이터 분석기반의 학술적 논의에 비중을 두고자 한다.

### III. 연구방법

#### 1. 조사설계 및 자료 수집 방법

한국관광에 대한 인식 조사를 위해 국내 관광

관련 정보를 제공하는 각 포털을 활용하여 관련된 텍스트를 수집하였다. 구체적으로는 인터넷에 분산저장 되어 있는 관광정보 즉, 웹페이지, 개인들이 운영하는 블로그, 공신력 있는 정보를 제공하는 뉴스 등에서 자료를 수집하였다. 자료수집은 데이터 수집 및 처리 프로그램인 텍스트롬(TEXTOM: Text to Metrix)<sup>1)</sup>을 활용하였다. TEXTOM은 web상의 다양한 데이터를 채널별로 자동 수집하여 정제, 매트릭스 생산까지 일괄 처리해 주는 데이터 처리 솔루션으로 크롤링(crawling) 방식을 이용한다. 또한 연관성과 정확도에 따라 추출된 데이터를 수집하고 다양한 주제에 대한 감성분석(sentiment analysis)과 오피니언 분석(opinion analysis)이 가능하여 텍스트 마이닝을 가능하게 해주는 한국어 최적화 빅 데이터 분석 솔루션이다. 그동안의 전통적인 조사의 한계를 극복하고 인터넷의 데이터를 자동으로 수집하고 수집된 데이터를 정제한 후 결과



〈그림 2〉 텍스트롬(Textom) 처리 과정 예시(텍스트롬 홈페이지 참조)

1) TEXTOM은 (주)더아이앤씨(THE IMC)가 자체 개발한 빅데이터 처리 솔루션임. 웹과 소셜미디어 상의 데이터를 수집에서 텍스트 마이닝과 매트릭스 데이터 생성까지 일괄 처리되는 방식으로 빅데이터 활용의 핵심기술인 빅데이터 큐레이션 및 분석 데이터를 생성하는 솔루션임. TEXTOM은 국내외 포털사이트와 소셜 미디어 등 다양한 채널에서 데이터를 대량으로 수집한 뒤 자동 정제 과정을 통하여 네트워크분석이 가능한 데이터 마이닝이 가능하며, 정제된 데이터는 유클리디언 계수, 코사인 계수, 자카드 계수 등 다양한 값으로 변환하여 다운로드 할 수 있어 타 분석 프로그램과의 확장에도 효율적임.



를 정리 분석할 수 있는 조사도구의 필요성에 의해(박한우, 2012) TEXTOM은 개발되었으며, 이후 전채남(2014), 김해원·전채남(2014) 등에 의해 학술적 논의로써 과학적인 방법으로 검증되어지고 있다.

본 연구에서는 2014년부터 1월 1일-12월 31일까지 1년간의 네이버, 다음, 구글에서 '관광, 여행' 키워드를 포함하고 있는 자료를 수집하였다. 수집된 데이터를 통해 핵심어를 도출하고, 이후 분석데이터의 탐색을 진행하여 데이터 수정 및 변환 과정을 거친 후 텍스트마이닝(text-mining)과 시맨틱네트워크(semantic network) 분석을 진행하였다.

텍스트마이닝은 구조화되지 않은 대규모의 텍스트 집합으로부터 새로운 지식을 발견하는 과정이며 문서 내에서 유용한 정보를 찾는 기술로 대량 정보의 문서 데이터에서 의미를 도출해내는 일련의 과정을 설명한다. 또한 텍스트마이닝은 비정형의 텍스트데이터를 자연어처리와 형태소

분석 기술에 기반하여 정제하고 유용한 단어를 추출해 빈도수를 추출하여 맥락(context) 수준의 의미를 찾아내는 프로세스를 의미한다(정근하, 2010). 시맨틱네트워크 분석은 문장 속의 개념들과 상징들에 대한 구조를 밝히기 위해 사용되었다는 점에서 소셜네트워크 분석의 한 형태이며(Chung & Park, 2010), 텍스트에서 자주 사용하는 단어들은 그 사람의 정체성을 나타내므로 담론의 성격을 드러내는 하나의 지표로도 작용한다.

## 2. 사회네트워크 분석(Social Network Analysis)

네트워크적인 성격을 가지기 시작한 사회변동 속에서 네트워크 분석에 대한 중요도가 커지게 되면서 이에 대한 학문적인 주목도도 높아지고 있다. 이 분석법은 많은 학문분야에서 빠른 속도로 보급되었으며, 이제는 대부분의 사회과학 분

〈표 3〉 중심성(centrality) 개요

구분	내용	본 연구에서의 적용
연결정도 중심성	네트워크의 노드들이 얼마나 많은 연결을 가지고 있는지를 중심으로 측정하는 개념. 한 노드에 연결된 다른 노드의 수로 측정. 연결된 노드의 수가 많을수록 연결정도 중심성은 높아지게 되며, 가장 연결이 많은 단어는 그 네트워크의 중요한 역할을 하는 허브가 됨	단어 간 연결 정도를 측정하여 연결정도 중심성이 높은 단어는 다른 단어와 많이 연결되어 있는 것으로 해석
근접 중심성	하나의 노드가 다른 노드에 얼마나 가깝게 있는가를 나타내는 개념. 두 노드 사이의 거리가 가깝게 있다면 그 노드와 쉽게 연결된다고 보고 그만큼 중심적인 역할을 한다고 간주	근접 중심성이 높은 단어는 다른 단어와 쉽게 연결되어 나타난다고 해석
매개 중심성	하나의 노드가 다른 노드들과의 네트워크를 구축할 때 중개자 혹은 다리 역할을 얼마나 수행하는가를 측정하는 개념. 중개역할에 초점을 맞추어 '중심성'을 측정.	매개 중심성이 높은 단어는 다른 단어들이 나타날 때 중개 역할이 상대적으로 많다고 해석
아이겐벡터 중심성	직접적으로 관계를 맺고 있는 노드들뿐만 아니라 연결된 이웃노드들의 간접적인 관계까지 포함하여 영향력을 계산하는 개념. 위계중심성(prestige centrality) 중 하나로써 중요한 노드와 연결될수록 해당 노드의 중심성은 높게 측정	아이겐벡터 중심성이 높은 단어는 중요하게 등장하거나 언급된 단어와 같이 연결된 정도가 많다고 해석
CONCOR	구조적 등위성 분석 중에 가장 보편적으로 사용되는 분석방법임. CONCOR는 단어 간의 동시 발생적 매트릭스를 통하여 의미 구조 내 텍스트를 분석하기 위해 사용됨. 복잡한 네트워크 집단사이에 숨겨진 하위집단들의 관계의 구조적 분석에 활용.	네트워크 분석을 위해 흔히 사용되는 소프트웨어 프로그램인 UCINET를 사용하여 실행
에고 네트워크 분석	전체 네트워크에서 단 하나의 주체(노드, 단어)를 중심으로 네트워크를 보고자 할 때 분석하는 방법.	관광에 대한 전체 네트워크에서 '한국관광공사'와 관련된 단어를 중심으로 어떠한 단어들이 연결되어 있는지를 살펴볼 때 활용

야, 물리학, 생물학 등의 자연과학분야 등에서 연구가 이루어지고 있다(김용학, 2011). 특히, 인터넷 사회의 등장과 정보의 폭발과 초연결사회의 도래로 복잡한 네트워크에 숨겨져 있는 정보의 비밀을 탐색하는 목적으로 정치, 경제, 조직 등의 분야에서도 활용되고 있다.

이러한 맥락에서 본 연구는 사회 네트워크 분석에서 나온 기법을 이용하고자 한다. 네트워크 분석 방법은 시스템을 구성하는 개체들 간의 관계구조를 연구대상으로 하는 실증적 연구 방법론으로 분석 단위가 개별 요소들이 아니라 개별 요소들 간의 관계성을 가진 연결망, 즉 구조(structure)를 살피는 것이 주요한 분석 틀이라고 할 수 있다. 이러한 관계적인 속성 분석을 위해 여러 가지 지표들이 발달되어 있다. 그 중 중앙성(centrality)은 네트워크 내에서 중앙에 얼마나 가까이 위치하고 있는지를 측정하는 지표로서, 사회네트워크 분석법에서 가장 많이 쓰이고 있다(김정하, 2012). 대표적인 개념으로서는 연결정도 중앙성(degree centrality), 매개중앙성(betweenness centrality), 인접 중앙성(closeness centrality) 등이 있으며(김용학, 2011). 본 연구에서는 연결정도 중심성(degree

centrality), 근접 중심성(closeness centrality), 매개 중심성(betweenness centrality), 그리고 아이겐벡터 중심성(eigenvector centrality) 등을 분석하고자 한다.

### 3. 분석설계

관광객이 웹이나 소셜미디어에 남긴 텍스트 자료로부터 구조화된 형태의 정보를 추출하여 패턴과 의미를 시각적으로 나타내고 의미 체계를 파악하고자 함이 본 연구의 목적이다. 이를 위해 웹과 소셜미디어에 작성된 텍스트를 수집-정제-분석-시각화의 과정을 통해 유용한 정보를 도출하고자 한다. 즉, 수집 데이터인 비정형데이터를 분석용 데이터로 활용하기 위하여 텍스트의 정제 과정을 수행하고, 정제된 데이터를 텍스트 마이닝을 통해 빈도를 계산하여 주요 단어들을 선정한다. 선정된 단어들의 범주화를 통해 매트릭스 데이터로 만들어 시멘틱네트워크 분석에 이용한다. 분석에는 Ucinet6<sup>2)</sup>을 활용하여 텍스트의 네트워크 중심성과 구조적 등위성(CONCOR: CONvergence of iterated CORrelations)<sup>3)</sup>, 에고 네트워크(ego-network) 분석을 진행한다.

〈표 4〉 분석데이터 정보

구분	내용
수집 범위	Naver(블로그, 뉴스, 카페, 웹, 지식인), Daum(블로그, 뉴스, 카페, 웹, 지식인), google(블로그, 뉴스, 웹)
수집 기간	2014년 1월 - 2014년 12월
수집 도구	TEXTOM
검색어	관광/ 여행
분석 도구	UciNet 6.0, NetDraw
수집 데이터	TEXTOM을 이용해 분석단어인 '관광/ 여행' 데이터를 수집

- Freeman이 개발한 다양한 연결망 분석기법을 활용할 수 있는 종합적인 프로그램으로 네트워크분석에서 주로 사용하고 있으며, 무료로 이용 가능하여 많은 연구자들이 사용하고 있다. 또한 Netdraw 기능이 포함되어 있어 결과 분석 뿐만 아니라 시각화 도구까지 포함하고 유용한 도구이다. 관련 프로그램은 다음의 홈페이지에서 다운로드가 가능(<https://sites.google.com/site/ucinetsoftware/home>)
- CONCOR 분석은 단어 간의 동시출현 매트릭스의 피어슨 상관관계를 분석하고 이를 토대로 노드들의 블록을 식별해내고 블록들 간의 관계를 파악하는 방법으로, 주로 복잡한 네트워크 집단의 숨겨진 하위집단들 관계의 구조적 분석에 활용(Wasserman & Faust, 1994)

UCINET와 함께 패키지화된 NetDraw를 활용하여 분석결과를 시각화함으로써 텍스트의 관계를 구체적으로 제시하고자 한다.

#### IV. 분석결과

##### 1. 데이터 빈도(상위 주요 빈도)

관광 및 여행 관련 내용을 텍스트마이닝 한 결과 총 31,735개의 단어가 출현되었으며, 그 중 상위 100개의 결과를 정리하면 다음과 같다 (100개 중 부사(예: 이제 now 등)와 유사어(한국관광공사/관광공사, 정보/관광정보 등) 등과 같은 단어는 삭제되어 81개로 정리됨).

조사결과, 해외 주요 국가 및 도시, 국내 주요 도시 등의 웹가시성(Web visibility)이 높은 것으로 파악되며, 가족여행, 자유여행, 해외여행, 신혼여행 등의 관광행태가 크게 부각되어 나타났고, 숙박 관광과 2박 3일, 3박 4일 등의 여행일 정도 높은 가시성을 나타내었다. 뿐만 아니라 각 교통수단들에 대한 내용들도 비중있게 제시되었다. 한편, 제주도의 경우 상기 표 외에도 제주여행, 제주도 여행, 제주관광, 제주도 관광, 제주 홀로여행, 제주 가족여행, 제주관광공사, 제주 공항, 제주 항공, 제주 맛집, 제주 둘레길, 제주 관광지도 등으로 제시되어 이들을 통합한다면 전체 결과물 중 제주도의 웹가시성이 가장 높은 것으로 파악된다. 이 외에도 1일, 1인 등의 빈도가

〈표 5〉 관광/ 여행 데이터 상위 주요 단어 빈도

순위	단어	빈도	%	순위	단어	빈도	%	순위	단어	빈도	%
1	여행	4981	3.59%	28	가족여행	183	0.13%	55	호치민	109	0.08%
2	관광	3385	2.44%	29	베트남	180	0.13%	56	홍콩	106	0.08%
3	여행사	660	0.48%	30	문화관광	179	0.13%	57	시내	105	0.08%
4	제주도	610	0.44%	31	신혼여행	178	0.13%	58	캐나다	104	0.07%
5	추천	543	0.39%	32	서비스	167	0.12%	59	강원도	104	0.07%
6	중국	416	0.30%	33	관광버스	165	0.12%	60	하나투어 여행사	104	0.07%
7	한국관광공사	356	0.26%	34	신청	158	0.11%	61	관광산업	103	0.07%
8	일본	326	0.24%	35	체험	157	0.11%	62	호주	102	0.07%
9	여행코스	316	0.23%	36	수학여행	156	0.11%	63	대구	101	0.07%
10	자유여행	309	0.22%	37	사진	155	0.11%	64	쇼핑	100	0.07%
11	호텔	300	0.22%	38	가을	155	0.11%	65	겨울	99	0.07%
12	미국	282	0.20%	39	무료	149	0.11%	66	홈페이지	98	0.07%
13	여행상품	267	0.19%	40	필리핀	148	0.11%	67	10월	97	0.07%
14	관광명소	263	0.19%	41	포함	145	0.10%	68	하와이	97	0.07%
15	관광비	262	0.19%	42	맛집	141	0.10%	69	교통정보	97	0.07%
16	정보	258	0.19%	43	지도	140	0.10%	70	관광열차	96	0.07%
17	서울	248	0.18%	44	비자	139	0.10%	71	공항	95	0.07%
18	롯데관광	248	0.18%	45	3박4일	136	0.10%	72	안전	95	0.07%
19	해외여행	240	0.17%	46	할인	131	0.09%	73	울릉도	95	0.07%
20	오사카	235	0.17%	47	숙박	128	0.09%	74	경주	95	0.07%
21	시간	229	0.17%	48	예약	127	0.09%	75	비	94	0.07%
22	계획	221	0.16%	49	기차여행	126	0.09%	76	크루즈	93	0.07%
23	부산	216	0.16%	50	2014	126	0.09%	77	홍보	92	0.07%
24	2박3일	209	0.15%	51	이벤트	124	0.09%	78	친구	92	0.07%
25	가이드	202	0.15%	52	휴양	115	0.08%	79	대한민국	91	0.07%
26	패키지	191	0.14%	53	모집	114	0.08%	80	문의	91	0.07%
27	항공	183	0.13%	54	5월	109	0.08%	81	대만	89	0.06%



10	자유여행	0.014	13	여행상품	45.856	28	가족여행	0.789	20	오사카	8.792
16	정보	0.014	19	해외여행	45.355	57	5월	0.782	9	여행코스	8.756
20	오사카	0.013	8	일본	45.109	10	자유여행	0.768	21	시간	8.491
27	항공	0.012	57	5월	45.109	7	한국관광공사	0.751	10	자유여행	8.435
13	여행상품	0.012	34	안내	44.865	70	10월	0.705	23	부산	8.097
29	베트남	0.011	4	제주도	44.865	21	시간	0.674	25	가이드	7.449
7	한국관광공사	0.011	7	한국관광공사	44.865	19	해외여행	0.659	15	관광비	7.407
24	2박3	0.011	21	시간	44.865	52	2014	0.628	26	패키지	7.240
22	계획	0.010	70	10월	44.865	17	서울	0.618	24	2박3	6.733
14	관광명소	0.010	48	숙박	44.385	8	일본	0.617	32	서비스	6.607
21	시간	0.010	14	관광명소	44.385	4	제주도	0.616	22	계획	6.578
25	가이드	0.010	23	부산	44.385	27	항공	0.609	17	서울	6.446
26	패키지	0.010	9	여행코스	44.385	43	맛집	0.594	36	체험	6.424
23	부산	0.010	49	예약	44.149	38	사진	0.582	14	관광명소	5.912
18	롯데관광	0.010	26	패키지	44.149	23	부산	0.571	45	비자	5.752
17	서울	0.009	38	사진	44.149	81	친구	0.561	29	베트남	5.605
45	비자	0.009	81	친구	44.149	9	여행코스	0.560	19	해외여행	5.500
35	신청	0.009	52	2014	44.149	25	가이드	0.554	44	지도	5.418
19	해외여행	0.009	67	쇼핑	43.915	14	관광명소	0.537	55	휴양	5.388
32	서비스	0.009	25	가이드	43.915	48	숙박	0.533	65	호주	5.377
34	안내	0.008	40	무료	43.455	49	예약	0.514	48	숙박	5.220
36	체험	0.008	43	맛집	43.455	36	체험	0.505	7	한국관광공사	4.997
58	호치민	0.008	36	체험	43.455	26	패키지	0.489	34	안내	4.781
28	가족여행	0.008	17	서울	43.229	74	공항	0.487	46	3박4	4.738
31	신혼여행	0.008	74	공항	43.005	35	신청	0.486	41	필리핀	4.665
48	숙박	0.007	78	비	43.005	78	비	0.462	31	신혼여행	4.643
46	3박4	0.007	27	항공	43.005	67	쇼핑	0.459	38	사진	4.551
44	지도	0.007	32	서비스	42.784	40	무료	40	66	대구	4.551
40	무료	0.007	83	문의	42.784	32	서비스	32	60	시내	4.527
43	맛집	0.007	24	2박3	42.564	18	롯데관광	18	43	맛집	4.465
47	할인	0.007	18	롯데관광	42.564	68	겨울	68	35	신청	4.423
49	예약	0.007	46	3박4	42.347	83	문의	83	42	포함	4.267
55	휴양	0.007	35	신청	42.132	69	홈페이지	69	54	버스	4.160
39	가을	0.007	31	신혼여행	42.132	39	가을	0.344	49	예약	4.151
60	시내	0.006	44	지도	42.132	44	지도	0.341	28	가족여행	4.150
42	포함	0.006	69	홈페이지	42.132	24	2박3	0.337	39	가을	4.088
61	캐나다	0.006	12	미국	41.709	54	버스	0.333	64	관광산업	4.070
38	사진	0.006	68	겨울	41.709	31	신혼여행	0.318	61	캐나다	3.945
52	2014	0.006	39	가을	41.5	46	3박4	0.301	13	여행상품	3.800
67	쇼핑	0.006	54	버스	41.5	12	미국	0.287	77	울릉도	3.776
65	호주	0.005	75	안전	41.294	33	관광버스	0.235	52	2014	3.659
78	비	0.005	60	시내	41.089	75	안전	0.226	68	겨울	3.485
59	홍콩	0.005	15	관광비	40.686	82	대한민국	0.224	67	쇼핑	3.481
74	공항	0.005	53	이벤트	40.686	53	이벤트	0.219	33	관광버스	3.446
77	경주	0.005	33	관광버스	40.488	15	관광비	0.219	40	무료	3.400
54	버스	0.005	37	수학여행	40.488	37	수학여행	0.215	78	비	3.397
41	필리핀	0.005	59	홍콩	40.291	62	강원도	0.211	79	크루즈	3.356
33	관광버스	0.005	20	오사카	40.097	60	시내	0.184	59	홍콩	3.345
53	이벤트	0.005	47	할인	40.097	59	홍콩	0.179	81	친구	3.271

82	대한민국	0.005	29	베트남	40.097	20	오사카	0.167	82	대한민국	3.259
71	하와이	0.005	80	홍보	39.713	80	홍보	0.165	76	울릉도	3.208
68	겨울	0.005	82	대한민국	39.524	29	베트남	0.154	74	공항	3.035
81	친구	0.005	41	필리핀	39.524	30	문화관광	0.148	58	호치민	2.985
75	안전	0.005	56	모집	39.336	56	모집	0.144	83	대만	2.927
66	대구	0.005	84	대만	39.151	47	할인	0.140	62	강원도	2.853
83	문의	0.004	79	크루즈	39.151	77	경주	0.120	47	할인	2.849
57	5월	0.004	62	강원도	39.151	45	비자	0.117	70	10월	2.755
79	크루즈	0.004	50	기차여행	38.967	79	크루즈	0.112	56	모집	2.712
37	수학여행	0.004	77	경주	38.967	41	필리핀	0.111	57	5월	2.548
56	모집	0.004	30	문화관광	38.785	84	대만	0.109	69	홈페이지	2.492
70	10월	0.004	71	하와이	38.605	61	캐나다	0.094	71	하와이	2.426
30	문화관광	0.004	61	캐나다	38.605	65	호주	0.092	53	이벤트	2.405
69	홈페이지	0.004	65	호주	38.605	71	하와이	0.089	51	관광가이드	2.400
84	대만	0.004	55	휴양	38.426	50	기차여행	0.081	75	안전	2.376
50	기차여행	0.004	45	비자	38.249	64	관광산업	0.081	18	롯데관광	2.348
64	관광산업	0.004	51	관광가이드	38.249	51	관광가이드	0.075	83	문의	2.320
62	강원도	0.004	64	관광산업	37.9	66	대구	0.066	37	수학여행	2.066
51	관광가이드	0.003	66	대구	37.727	55	휴양	0.060	80	홍보	2.039
76	울릉도	0.003	73	관광열차	37.387	73	관광열차	0.060	50	기차여행	1.970
80	홍보	0.003	58	호치민	37.22	58	호치민	0.045	73	관광열차	1.618
73	관광열차	0.003	76	울릉도	37.054	76	울릉도	0.037	30	문화관광	1.297
72	교통정보	0.003	72	교통정보	35.47	72	교통정보	0.016	72	교통정보	0.323
63	하나투어 여행사	0.000	63	하나투어 여행사	0.000	63	하나투어 여행사	0.000	63	하나투어 여행사	0.000

〈표 7〉 주요 단어 빈도와 각 중심성 비교

구분	빈도		연결정도중심성		근접중심성		매개중심성		위세중심성	
	빈도	순위	빈도	순위	빈도	순위	빈도	순위	빈도	순위
여행	4981	1	0.196	1	50	1	1.717	1	93.279	1
관광	3385	2	0.147	2	50	2	1.717	2	90.506	2
여행사	660	3	0.025	4	48.824	3	1.330	3	13.772	5
제주도	610	4	0.025	5	44.865	17	0.616	22	<b>17.831</b>	<b>3</b>
추천	543	5	0.026	3	48.538	4	1.275	4	17.994	4
중국	416	6	0.015	8	46.369	8	1.015	6	9.728	10
한국관광공사	356	7	0.011	18	44.865	18	0.751	15	4.997	33
일본	326	8	0.016	7	45.109	14	0.617	21	11.798	7
여행코스	316	9	0.015	9	44.385	24	0.560	28	8.756	13
자유여행	309	10	0.014	12	46.111	9	0.768	14	8.435	15
호텔	300	11	0.017	6	46.893	7	0.897	9	12.413	6
미국	282	12	0.015	11	41.709	48	0.287	51	9.519	11
여행상품	267	13	0.012	16	45.856	12	1.010	7	3.800	50
관광명소	263	14	0.010	21	44.385	22	0.537	30	5.912	25
관광비	262	15	0.015	10	40.686	54	0.219	56	7.407	18
정보	258	16	0.014	13	47.977	5	1.269	5	9.753	9
서울	248	17	0.009	27	43.229	35	0.618	20	6.446	23
롯데관광	248	18	0.010	26	42.564	42	18	41	2.348	76
해외여행	240	19	0.009	30	45.355	13	0.659	18	5.500	28
오사카	235	20	0.013	14	40.097	59	0.167	61	8.792	12

가장 영향력 있는 중심 노드를 찾기 위해 주요 단어의 위세중심성을 계산한 결과, 제주도, 추천, 여행사, 호텔, 일본, 항공, 정보, 중국, 미국 등이 높게 나타났다. 여기에서 제주도와 해외 국가들의 비중도가 큼을 알 수 있다.

단어의 빈도보다 연결정도 중심성이 높거나 위세중심성이 높은 단어를 찾아 상대적으로 다른 단어나 이슈와 보다 밀접하게 연결되어 있는 것을 확인하고자 이들 값을 비교하였다. 빈도 순위 상위 20개의 단어를 비교한 결과 다른 단어나 이슈와 보다 밀접하게 연결된 단어는 추천, 호텔, 정보로 나타났다.

### 3. CONCOR(CONvergence of iterated CORrelations) 분석

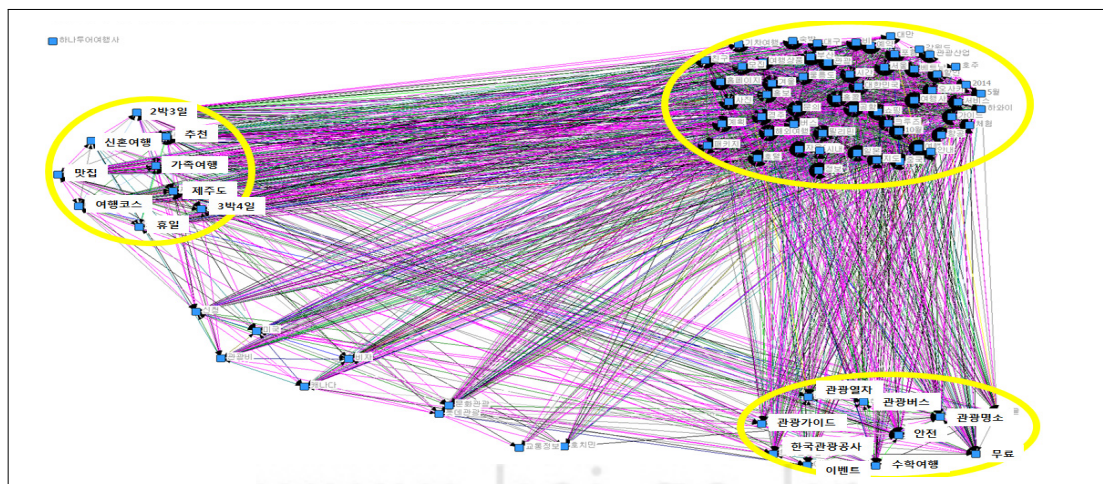
CONCOR는 상관관계 분석을 반복적으로 수행하여 적절한 수준의 유사성 집단을 찾아내는 방법으로 본 연구에서는 의미망 분석결과를 바탕으로 CONCOR 분석을 실시하였다. 그 결과, 3개의 클러스터(cluster)가 생성되었고 그 중 서울, 경주, 부산, 강원도, 호주, 하와이, 홍콩, 오사카, 홈페이지, 여행상품 등의 노드가 포함된 클러스터가 가장 크게 이루어졌으며, 이를 ‘관광

대상지 중심 클러스터’로 명명하였고, 그 다음의 2개 클러스터는 비슷한 크기로 생성되었으며, 그 중 하나는 제주도, 2박 3일, 3박 4일, 가족여행, 여행코스, 맛집 등이 포함되어 있어 ‘제주 중심 여행 클러스터’로 명명하였고, 다른 하나는 관광열차, 관광버스, 관광가이드, 관광명소, 안전, 무료, 한국관광공사 등이 포함되어 있어 ‘관광활동 중심 클러스터’로 명명하였다.

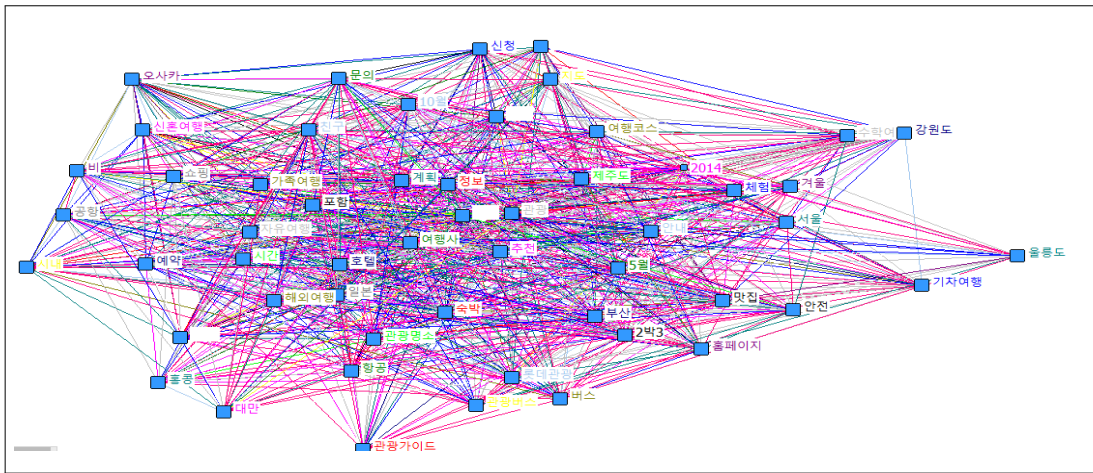
### 4. 주요 단어 에고네트워크 분석(Ego Network Analysis)

에고네트워크 분석은 특정한 노드를 중심으로 관련된 다른 노드간의 연결을 표현한 네트워크로써 이를 통해 내재된 구성노드들 간의 특성을 기술하고 파악하는데 사용한다. 본 연구에서는 일정별 ‘2박 3일’과 ‘3박 4일’의 에고네트워크 분석을 실시하였다.

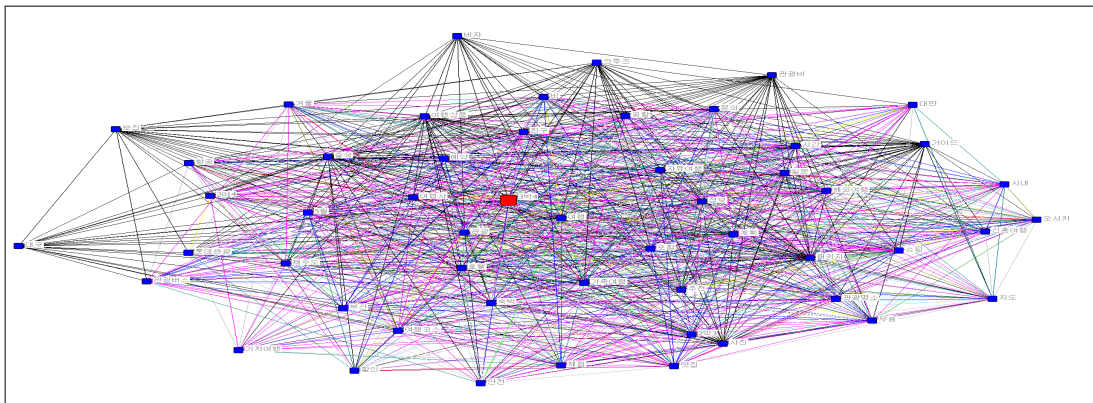
분석결과 2박 3일은 맛 집, 추천, 체험, 5월, 홈페이지, 가족여행, 정보, 숙박, 안전, 포함, 관광명소, 무료, 쇼핑, 일본 등과의 연결이 강한 것으로 나타났으며, 3박 4일은 호텔, 친구, 5월, 예약, 중국, 제주도, 여행사, 여행상품, 자유여행 등이 강한 것으로 도출되었다. 이러한 결과는 2



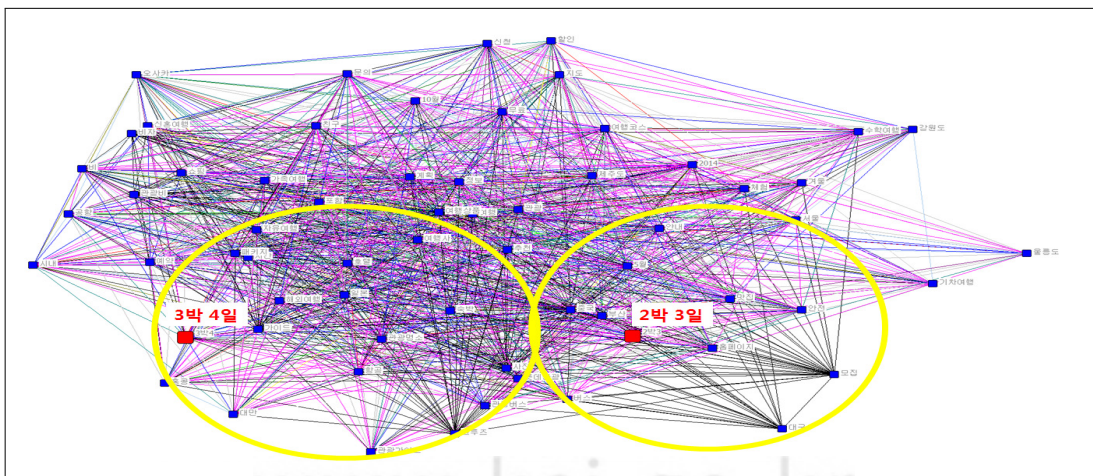
〈그림 4〉 CONCOR 분석 결과 도식화



〈그림 5〉 2박 3일 에고네트워크



〈그림 6〉 3박 4일의 에고네트워크



〈그림 7〉 일정별(2박 3일/ 3박4일) 에고네트워크 분석 비교



박 3일의 경우 가족 여행의 유형이 강하며, 주로 홈페이지 등을 통해 정보를 얻고, 관광명소나 맛집 등을 추천을 통해 즐기며, 무료, 체험을 선호 하면서 안전도 고려하는 것으로 풀이된다. 3박 4일의 경우 가까운 중국이나 홍콩과 같은 해외 여행을 위한 내용들이 강하며, 여행사의 여행상품을 많이 고려하는 것으로 확인된다. 하지만, 여행상품 패키지 못지 않게 예약 및 자유여행도 강하게 나타나고 있음이 확인된다.

## V. 결론

본 연구에서는 국내관광에 대한 관광인식을 빅데이터를 통해 분석하고자 2014년부터 1월 1일- 12월 31일까지 1년간의 네이버, 다음, 구글을 활용하여 '관광/여행' 키워드를 포함하고 있는 웹페이지, 개인들이 운영하는 블로그, 공신력 있는 정보를 제공하는 뉴스 등을 수집하여 분석하였다. 데이터는 UCINET와 함께 패키지화된 NetDraw를 활용하여 연결정도 중심성, 근접 중심성, 매개 중심성, 그리고 아이겐벡터 중심성 등을 분석하였다.

분석결과 전반적으로 해외 주요 국가 및 도시, 국내 주요 도시 등의 웹가시성(Web visibility)이 높은 것으로 파악되었으며, 가족여행, 자유여행, 해외여행, 신혼여행 등의 관광행태와 숙박관광과 2박 3일, 3박 4일 등의 여행일정도 높게 나타났다. 특히, 제주도의 경우 주요 단어에서도 높게 나타났지만, 유사한 데이터(제주여행, 제주도 여행, 제주도 관광, 제주 홀로여행, 제주 가족여행, 제주관광공사, 제주 공항, 제주 맛집, 제주 둘레길, 제주 관광지도 등)로도 많이 제시되어 이들을 통합한다면 전체 결과물 중 제주도의 웹가시성은 가장 독보적인 수준이다. 이 외에도 1일, 1인 등의 빈도가 부각되었고, 안전, 관광주간, 관광두레, 해미읍성도 포함되어 있어 2014

년의 사회 환경적 특성을 반영하고 있음이 확인된다. 그러나 2014년도에는 정부의 관광정책이나 관광개발과 관련된 내용은 미미한 것으로 파악된다. 단, 수학여행이 높게 출현되어 세월호 사고로 인한 수학여행 폐지 논란이 이어지면서 기사화 된 것이 출현된 경우가 있긴 하다.

생성된 데이터들은 적절한 수준의 유사성으로 3개의 클러스터(cluster)가 생성되었고 그 중 서울, 경주, 부산, 강원도, 호주, 하와이, 홍콩, 오사카, 홈페이지, 여행상품 등의 노드가 포함된 '관광 대상지 중심 클러스터'와 제주도, 2박 3일, 3박 4일, 가족여행, 여행코스, 맛집 등이 포함되어 있는 '제주 중심 여행 클러스터', 관광열차, 관광버스, 관광가이드, 관광명소, 안전, 무료, 한국관광공사 등이 포함되어 있는 '관광활동 중심 클러스터'로 구분되었다.

이 외에도 '2박 3일'과 '3박 4일'에 대한 예고 네트워크 분석에서 2박 3일은 맛집, 추천, 체험, 홈페이지, 가족여행, 정보, 숙박, 안전, 포함, 관광명소, 무료, 쇼핑, 일본 등과의 연결이 강하게 나타났으며, 3박 4일은 호텔, 친구, 5일, 예약, 중국, 제주도, 여행사, 여행상품, 자유여행 등이 강하게 도출되었다. 이러한 분석결과를 통해 본 연구에서는 다음과 같은 시사점을 제시한다.

첫째, 관광인식 범주가 다양함을 확인할 수 있다. 특히 여행행태에 있어서 자유여행, 가족여행, 문화관광, 기차여행, 수학여행, 해외여행, 여행상품 등이 주요 인식 단어로 도출되었다. 이중 자유여행은 2014년도의 미디어의 영향이 큰 것으로 풀이된다. TV 방송에 '꽃보다 할배'와 같은 여행 관련 프로그램의 방영으로 해외여행의 경우 패키지 상품 못지않게 자유여행의 선호도가 높아지고 있으며, 이러한 영향이 사회 트렌드로 반영된 결과로 볼 수 있다.

둘째, 관광목적지로서의 제주도의 위상이 확인되는 결과이다. 앞서서도 제시하였지만, 제주도의 경우 제주여행 또는 관광 외에도 제주도 모

바일 할인쿠폰, 제주도 여행코스, 제주도 관광지, 제주관광공사, 제주관광협회, 제주도 겨울 여행 코스, 제주관광센터, 드라이브 제주, 제주도 효도관광, 제주 특급호텔 무료 이용, 제주도청, 제주도 택시관광, 제주관광 정보, 제주중문관광단지, 제주도 버스관광 패키지, 제주관광홍보, 제주여행 할인 소품물, 제주여행 계획, 제주도 가족여행, 제주항, 제주 숙박, 제주관광안내 등의 단어들로 제시되어 데이터 양이 가장 크며, 또한 다양한 형태로 접근되었음이 확인된다. 이처럼 제주도가 국내에서 최고의 관광목적지로서 위상이 높은 것은 긍정적 현상이나 전체적인 국내관광 시장을 고려해 볼 때, 제주도뿐만 아니라 다양한 지역으로의 관심을 분산시킬 수 있는 대안이 필요하다. 이는 각기 지자체들의 자발적인 노력도 중요하겠지만 정부에서도 방안을 강구하여 특정 관광목적지로의 집중 현상보다는 분산을 통해 지역의 균형 발전 프레임을 구축할 필요가 있다.

셋째, 1인 관광, 홀로 관광의 데이터도 부각되어 관광행태의 다양성을 더해 주고 있다. 직접 정보를 검색하고, 여행코스를 찾으며, 예약, 할인, 무료, 지도, 체험거리 등을 추구하는 행태들이 높아져 부각된 현상이라 해석된다. 이는 타인이 바라보는 시선에서 벗어나고 싶어하는 요즘의 싱글 라이프와 그 맥을 같이 한다고 보여진다. 점차 혼자에 익숙해지는 사회현상을 고려하여 관광시장에서도 싱글 시장의 잠재성을 고려해야 할 것이다.

넷째, 정부정책이나 관광개발과 관련된 데이터는 상대적으로 미미한 것으로 확인된다. 대체로 특수 형태의 고급관광 보다 일반 보편적 관광활동을 추구하는 욕구들이 높아 일상에서 즐길 수 있는 소소한 것들이 많이 도출된 만큼, 관광정책 방향 수립시 좀 더 생활밀착형 관광정책 수립에 관심을 가지는 것도 성과제고에 도움이 될 것으로 풀이된다.

소셜미디어 빅데이터는 양(量)에 제한이 없다. 무제한으로 수집한 정보라 하더라도 텍스트 마이닝 툴(text-mining tool)을 이용하면 다양한 반응들을 정량화해서 분석할 수 있다(신승중, 2012). SNS에서는 개인의 취향이 보다 직접적으로 반영되고, 진실성과 진정성, 관련성이 증가되므로 데이터로서의 가치가 높다. 또한 실시간성과 가속성이라는 특징을 지녔기 때문에 어떠한 매체보다도 이슈의 확산 속도가 빠르다. 즉 SNS상의 빅데이터를 관광 인식의 측정 기준으로 삼는다는 것은 해당 시점에 어떤 방법론보다도 시의적절하고 유용하며 정확한 평가를 얻는데 도움이 될 것으로 기대한다. 이러한 맥락에서 본 연구는 디지털 경제의 확산으로 관광산업에서도 빅 데이터에 기반한 과학적인 마케팅을 도입할 수 있는 단초를 제공하고, 나아가 통계분석에서 융복합적 측면에서의 학문적 성과를 제고하는 논의로서 그 의의를 가진다. 하지만, 본 연구는 국내 관광 인식 조사에 있어서 한국어 영역에 대해서만 분석이 이루어져 영어권역에 대한 의견은 포함하지 못한 아쉬움이 있다. 즉, 외국어 분석 솔루션에 대한 기술적 한계로 외국인들의 인식은 분석되지 못하였다. 향후에는 이를 극복하여 국내 관광에 대한 세계 각국 관광객들의 반응을 살펴보고, 나아가 숙박시설, 관광 인프라 등 세부 영역별 인식 변화 등을 살펴봄으로서 관광 마케팅 전략 수립에 유용한 토대를 마련해야 할 것이다.

#### 참고문헌

- 강만모 · 김상락 · 박상무(2012). 빅데이터의 분석과 활용. 『정보과학회지』, 30(6), 25-32.  
공효순 · 송은지(2013). 빅데이터를 이용한 호텔기업 CRM 및 보안에 관한 연구. 『융합보안논문지』, 13(4), 69-75.  
김용학(2011). 『사회 연결망 분석』. 서울: 박영사.

- 김해원·전채남(2014). 빅데이터를 활용한 콘텐츠 제작방안에 관한 탐색적 연구: TV홈쇼핑을 중심으로. 『사이버커뮤니케이션 학보』, 31(3), 5-51.
- 김정하(2012). 『사회연결망분석을 이용한 지역의료 관광산업의 이해관계자 네트워크 분석: 서울시 강남구 지역을 대상으로』. 한양대학교대학원 석사학위논문.
- 네이버 지식백과(2014). <http://terms.naver.com>.
- 도해용·류기상(2013). 빅데이터 분석기법을 이용한 레스토랑 품질과 고객만족이 추천행동에 미치는 영향에 관한 연구. 『외식경영연구』, 16(2), 233-255.
- 도해용·이애주·이창원(2013). 빅데이터를 이용한 레스토랑품질이 고객만족에 미치는 영향에 관한 연구: 국가의 조절효과 중심으로. 『외식경영연구』, 16(6), 49-70.
- 문정욱(2013). 관광서비스 발전을 위한 ICT 활용 현황과 시사점: 국내외 사례분석을 중심으로. 『방송통신정책』, 25(20), 1-35.
- 민경익·이태숙(2012). 미디어 분석을 통한 한국관광 인식 변화에 관한 연구: 신문기사를 중심으로. 『관광학연구』, 36(10), 81-104.
- 변정우(2013). 빅 데이터 관광에서 얼마나 활용될 수 있을까?. 『웹진 문화관광』, 3월호.
- 배동민·박현수·오기환(2012). 빅데이터 동향 및 정책 시사점. 『방송통신정책』, 25(10), 37-74.
- 삼성경제연구소(2012). 빅데이터: 산업 지각변동의 진원. 『CEO Information』, 851.
- 삼정 KPMG경제연구원(2012). 『빅 데이터 분석을 통한 기업 미래 가치 창출』.
- 신승중(2012). 빅데이터를 이용한 SNS 활용방안 연구. 『한국인터넷방송통신학회논문지』, 12(6), 267-272.
- 송태민(2012). 빅 데이터를 활용한 자살요인 다변량 분석: Google 검색트렌드 적용. 『보건·복지 Issue & Focus』, 168호.
- 연합뉴스(2013. 9. 12). 제주 관광정보 빅데이터 플랫폼 구축... 실시간 서비스.
- 이영진·정호권·부석현(2010). 온라인 관광 (eTourism)정보의 공공성평가: 포털 사이트를 대상으로. 『관광·레저연구』, 22(3), 389-408.
- 이영진·윤지환(2014). 관광분야에서 SNS 빅데이터 활용 방법 모색: 관광정보 검색 키워드 분석을 중심으로. 『관광연구저널』, 28(3), 5-14.
- 이양환(2013). 문화산업에서 빅 데이터의 필요성. 『웹진 문화관광』, 3월호.
- 전채남(2014). 텍스트를 이용한 SNA 분석. 『한국데이터사이언스학회 2014 학술대회 발표자료』.
- 정근하(2011). 『텍스트마이닝과 네트워크 분석을 활용한 미래예측 방법 연구』. 한국과학기술기획 평가원.
- 한국관광공사(2014). 『빅데이터 활용 관광사업 성과 시범 분석: 2013 문화관광축제를 중심으로』. 한국관광공사.
- 한국문화관광연구원(2012). 『미래 관광환경 변화 전망과 新관광정책 방향』. 한국문화관광연구원.
- 한국정보화진흥원(2012). 『대한민국 사회현안과 빅데이터 전략 세미나』. 한국정보화진흥원.
- 한형상·이창호(2014). 빅데이터 분석의 현황과 발전 전략. 『PD 이슈리포트』, 6.
- 함유근·채승병(2012). 『빅데이터, 경영을 바꾸다』. 삼성경제연구소.
- NSP 통신(2013.08.22). 부산 해운대구, 전국 최초 '빅데이터' 접목 관광정책 선보여.
- Chung, C., & Park, H. W. (2010). Textual analysis of a political message: The inaugural addresses of two Korean presidents. *Social Science Information*, 49(2), 215-239.
- Douglas, L. (2012). The importance of 'Big Data': A definition. Gartner, *Retrieved*, 21, June.
- IBM. (2011). *Understanding big data*.
- Gartner. (2012). *The importance of 'Big Data': A definition*.
- IDC. (2011. 6). *Extracting value from chaos*.

IPK International. (2011). *ITB World Travel Report 2010/2011*.

McKinsey Global Institute. (2011. 6). *Big data: The next frontier for innovation, competition, and productivity*.

O'eilly. (2012). *Big data now: Current perspectives from O'eilly Radar*. O'eilly Media.

UCINET Softwqre. (2015). <https://sites.google.com/site/ucinetsoftware/home>

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and appli-*

*cations*. New York, Cambridge University Press.

World Economic Forum. (2012). *Big data, big impact: New possibilities for international development*.

2015년 7월 10일 최초투고논문 접수

2015년 11월 17일 최종심사완료 및 게재확정 통보

2015년 11월 19일 최종논문 도착

3인 익명심사 畢