

텍스트마이닝 기법을 활용한 유스호스텔의 활성화 방안 에 관한 연구: 부산 유스호스텔 아르피나를 중심으로

이 승 후¹ · 김 학 선²

논문초록

유스호스텔은 청소년을 위한 숙박시설로서 출발하여 호텔 등과 더불어 세계적으로 대표 숙박 시설로 알려져 있다. 유스호스텔은 최근 다양한 형태의 숙박시설이 발달하고, 저가의 숙박업이 활성화되어 숙박시설로서의 입지가 줄어들고 있는 실정이다. 따라서 본 연구는 침체된 유스호스텔의 현재 문제점을 파악하기 위하여 2015년부터 2017년까지 3년간 아르피나와 관련된 신문 기사를 활용하여 텍스트 마이닝 분석을 실시하였고, 고객의 만족도를 파악하고 문제점 해결방안을 도출하기 위해 최근 1년간의 유스호스텔 방문자의 페이스북과 블로그의 리뷰를 추출하여 오피니언 마이닝 분석을 실시하였다. 분석결과 타 숙박시설에 비해 SNS마케팅이 현저히 부족하며, 주 고객이 단체로 한정되어 있었다. 또한, 유스호스텔에 관한 감성분석 결과 저렴한 가격, 깨끗한 시설, 안전함 등의 긍정이 47.3%로 가장 많았으며, 시설의 노후, 위치적 입지, 공용화장실 등의 부정이 15.2%로 나타났다. 이러한 결과를 토대로 아르피나의 마케팅 전략과 매출 증대를 위한 활성화 방안을 제시하였다.

주제어: 오피니언 마이닝, 유스호스텔, 의미연결망분석, 텍스트 마이닝, 아르피나

-
- 1 경성대학교 호텔관광외식경영학과 박사과정. 48434 부산광역시 남구 수영로 309(대연동)
tel: 051-627-6830 e-mail: pingklo@naver.com (주저자)
- 2 경성대학교 호텔관광외식경영학과 부교수. 48434 부산광역시 남구 수영로 309(대연동)
tel: 051-627-6830 e-mail: kims@ks.ac.kr (교신저자)

I. 서 론

유스호스텔이란 설립배경부터 청소년을 위한 숙박시설로서 출발하여, 독일에서 미래 세대인 청소년들에게 도시화 및 산업화에 따른 생활환경의 황폐화에서 벗어나 친자연적인 심신연마를 위한 여행길에서 학교를 숙소로 빌려 사용한 것에서 유래되었다(한국 유스호스텔연맹, 2011). 1932년 네덜란드에서 국제유스호스텔연맹(IYHF)이 발족되었으며, 국내의 경우 1967년에 한국유스호스텔협회가 발족되어 유스호스텔이 활성화되기 시작했다(최영준·오홍철, 2010).

전 세계적으로 다양한 종류의 숙박시설이 존재하고 있다. 하지만 시대에 맞지 않는 숙박시설은 퇴출되고, 그 시대가 요구하는 새로운 형태의 숙박시설이 등장하게 된다(전인호·강명석, 2013). 우리나라의 경우도 서비스와 시설, 명칭에 있어 많은 변화가 이루어졌다. 간단한 예를 들면, 민박이나 여관이라 불리던 시설은 줄어들고 모텔이나 펜션, 게스트하우스란 이름으로 현 시대에 맞는 형태의 숙박업이 증가하고 있다. 숙박시설의 변화하는 이유는 숙박시설을 이용하는 소비자들의 외부환경과 소비자의 의식 및 행동 등이 급격하게 변화 하면서, 소비자들의 라이프스타일이 변화하기 때문이다(손병모, 2011). 한국유스호스텔연맹에 따르면 국내 유스호스텔 수는 1998년 20개, 2008년 79개, 2018년 현재 86개로 최근 성장률은 아주 낮은 상태이다.

유스호스텔에 관한 선행연구를 학술연구정보서비스(RISS)에서 검색하면 국내논문으로는 민덕기(1968)의 ‘유스호스텔의 특성과 한국에 있어서의 전망’이란 논문이 최초의 논문으로 시작하여 현재(2018년 4월)까지 총 36편의 학술지논문이 검색된다. 해외에서는 유스호스텔의 기원이라 할 수 있는 1912년 이후 1931년에 최초의 논문을 검색할 수 있다. 최근 연구로는 이희열·권점옥(2011)이 대학생들을 대상으로 유스호스텔에 대한 이미지가 이용에 관한 만족도와 행동의도에 관한 연구를 하였다. 이와 같이 유스호스텔에 관한 연구는 아주 미비한 실정이며, 특히 2000년 이후의 논문은 거의 찾아보기 힘든 실정이다. 이는 1960년도 국내에 유스호스텔이 들어온 이후 발전이 없었다고 볼 수 있으며, 시대 흐름에 따른 변화가 많이 없었다고 판단할 수 있다. 또한 숙박시설의 급격한 변화에도 불구하고 유스호스텔은 큰 발전 없이 학교 및 기타 단체고객위주로 운영되고 있다.

따라서 본 연구에서는 빅데이터의 텍스트 마이닝 기법을 활용하여 부산 유일의 유스호스텔인 ‘아르피나’에 대한 의미연결망분석과 이를 시각화하고, ‘유스호스텔’에 관한

리뷰정보를 추출하여 감성분석을 실시하여 유스호스텔의 현 실태를 파악하고 마케팅 방향과 매출 증대를 위한 활성화 방안을 도출하고자 한다. 더 나아가 관광산업의 활성화로 부산 지역경제성장에 이바지함을 목적으로 한다.

II. 이론적 배경

2.1 부산 유스호스텔 아르피나

2.1.1 아르피나 소개

아르피나는 부산관광공사 산하의 유스호스텔로 2004년 7월에 개관하여 현재까지 운영 중에 있는 복합시설이다. 유스호스텔의 숙박시설뿐만 아니라 연수시설, 체육시설 및 청소년 수련시설을 갖추고 있어 청소년을 비롯하여 일반 단체 및 가족 여행객 등 다양한 형태의 방문객들을 대상으로 운영되고 있다. 개관 이후 2004년에는 유스포럼(Youth Forum)을 개최하였고, 2005년 APEC 지원단 숙소, 2008년 이후 부산국제영화제(BIFF) 관광 숙소로 지정되어 왔다.

2.1.2. 유스호스텔(Youth Hostel)의 정의와 선행연구

독일에서 발생한 유스호스텔은 교사인 리하르트 쉬르만(Richard Schirrmann)이 1909년 여름 8일간의 독일 알테나(Altena)에서 아헨(Aachen)까지의 여정을 통해 착안하여 창시되었다. 독일어로 청소년이라는 Jugend와 숙박이라는 Herberge의 의미를 포함하고 있어 청소년숙박시설 이라 할 수 있다(한국유스호스텔연맹, 2011). 유스호스텔이란 ‘젊은이의 집’이란 뜻으로 국내 또는 세계 곳곳을 여행하는 청소년들이 매우 저렴한 가격으로 안전하고 편안하게 머물 수 있는 여행숙박시설이다. 유스호스텔의 기능은 단순한 숙박시설이 아닌 다양한 부대시설과 유익한 프로그램을 가지고 교육, 세미나, 연수는 물론 개인 및 가족에게 레저, 휴식공간을 제공함에 있다(전명기, 1998).

유스호스텔에 관한 선행연구를 학술연구정보서비스(RISS)에서 검색하면 국내논문으로는 민덕기(1968)의 ‘유스호스텔의 특성과 한국에 있어서의 전망’이란 논문이 최초의 논문으로 시작하여 현재(2018년 4월)까지 총 36편의 학술지논문이 검색된다. 해외에서

는 유스호스텔의 기원이라 할 수 있는 1912년 이후 1931년에 최초의 논문을 검색할 수 있다. 최근 연구로는 이희열·권점옥(2011)이 대학생을 대상으로 유스호스텔에 대한 이미지가 이용에 관한 만족도와 행동의도에 관한 연구를 하였다. 국외논문으로 Biesanz(1941)는 독일 유스호스텔이 나치의 출현 전 국제화 촉진에 어떠한 기여를 했는지, 나치 운동 이후 독일 내 호스텔 시스템이 이데올로기 변화에 부합하기 위해 어떻게 정치적 이데올로기가 변화되었는지에 대한 연구가 있다. Nash, Thyne and Davies(2006)는 스코틀랜드 유스호스텔 협회가 배낭여행객들을 대상으로 실시한 연구에서 유스호스텔 이용객들은 활동적인 것을 선호한다는 것과 오래 머무른다는 결과를 도출했다. 국내의 경우, 최근 게스트하우스, 비즈니스호텔 등의 저가 숙박시설들의 등장으로 유스호스텔의 입지가 하락하고 이용 고객이 급감하고 있다. 그에 따라 유스호스텔에 관한 연구 또한 매우 찾기 힘든 실정이다.

<표 2> 한국 유스호스텔 현황

순번	지역	유스호스텔 명	수
1	서울	서울국제, 하이서울, 서울올림픽파크텔, 국제청소년센터, 호스텔코리아 창덕궁	5개
2	인천	강화, 강화로얄, 서해, 오션빌	4개
3	경기	가평, 고양, 공립, 그린캠프, 베르아텔, 베어스타운, 안양블루몬테, 양지파인, 양평SN, 자연과별, 캠프리그브스, 한국민속촌, 한울유스센터, 홈브리지케빈, 홈비리지힐사이드	15개
4	강원	강촌, 동서울, 보광휘닉스빌, 북한강기화, 비발디파크, 설악산, 세계, 오투, 용평, 유토피아, 치악산황둔, 평창, 신안웰리힐리	13개
5	대전	유성	1개
6	충남	계룡산갑사, 공주, 자연부여, 천안상록, 합덕	5개
7	충북	소백산, 수안보사조마을, 열림원, 증평, 청풍, 충주호월악산	6개
8	부산	부산관광공사아르파나	1개
9	경남	거제, 거제해와나루, 파라다이스힐, 우포생태촌, 청학동명륜학당, 청학동고목당, 통영동원	7개
10	경북	문경새재, 경주황룡, 계림, 동궁, 리멤버, 신라, 토함산, 포시즌, 한국관,	10개

		해누리	
11	광주	오아시타호텔	1개
12	전남	보성, 순천시, 우수영, 진도군, 해남	5개
13	전북	격포채석장, 고창군선운산, 익산, 전주JK, 갤러리스테이호텔, 진스호텔, 지리산	7개
14	제주	메이폴, 제주조이빌, 제주C&P, 마이테르, 호텔코리아 제주시탐동, 제주시명도암	6개
총		86 개	

주: 한국유스호스텔연맹의 자료를 바탕으로 본 연구자가 재구성

2.2 텍스트 마이닝(Text Mining)

텍스트 마이닝은 비정형화된 방대한 텍스트 자료로부터 의미 있는 정보 및 유용한 패턴과 관계 등을 추출하는 기술을 의미한다(Chiang, Lin & Chen, 2011). 또한 게시글, 댓글 등의 비정형 텍스트 데이터에서 사용자가 관심을 나타내는 정보를 키워드의 수준이 아니라 맥락으로 분석하고 데이터 속에서 숨겨진 패턴을 발견함으로써 가치 있는 정보를 추출하는 기수를 의미한다(전채남·서일원, 2013).

대중들의 게시글에서 나타나는 정보를 수집하고 내용을 분석하기 위해서는 자동화된 텍스트 분석 기법의 활용이 요구된다. 자동화된 텍스트 분석은 자연어처리를 통해 사람들 간의 현실의 언어를 컴퓨터가 인식할 수 있는 형태로 전환하고, 의미 있는 텍스트를 추출하는 텍스트 마이닝으로 이루어진다(양지윤, 2018).

텍스트 마이닝을 활용한 연구로는 김학선(2017)이 빅데이터를 활용한 음식관광에 대한 텍스트 마이닝분석을 통해 국내 음식관광의 과학적 마케팅의 단초를 제공하였다. 또한 김주영·김동수(2017)는 텍스트마이닝 기반의 온라인 상품 리뷰 추출을 통한 목적별 맞춤형 정보 도출에 관한연구를 하였고, 김도경·김인신(2017)은 텍스트 마이닝을 이용한 온라인 리뷰의 호텔 선택속성 분석에 관하여 연구하였다.

2.3 의미연결망분석(SAN : Semantic Network Analysis)

의미연결망 분석은 사회의 시스템 구조를 파악할 때 행위자 사이의 관계를 파악할

때 행위자 사이의 관계를 중시하는 사회연결망 분석을 커뮤니케이션 메시지에 적용시킨 방법이다(Wang & Rada, 1998; Wasserman & Faust, 1994). 사회연결망 분석은 ‘노드(node)’라는 개인과 이들 간의 ‘연결 관계(link)’로 구성되며 개인 간의 상호 관계를 종합적으로 분석 한다(김상국, 2011). 의미연결망 분석은 개별 노드가 어떠한 상호 관련이 있는지 연결 관계로 보여주며, 그 과정에서 어떠한 노드를 중심으로 담론을 구축하고 있는지를 가시적으로 보여준다. 이렇게 형성된 단어 간 연결 구조, 즉 특정 이슈에 대한 의미연결망 구조는 담론이 어떠한 방식으로 어떻게 구축되고 있는지 조직적인 분석을 가능케 한다(Hwang & Hwang, 2016). 이처럼 의미연결망 분석에서는 특정 유목을 가정하지 않은 상태에서 단어들의 빈도와 한 문장 안에서 동시에 사용되는 단어들의 관계를 통해 메시지의 의미와 패턴을 분석할 수 있다(Wang & Rada, 1998). 분석 기준이 되는 지표로는 분석 수준에 따라 네트워크에 대한 밀도, 포괄성, 집중도의 거시 수준의 분석 지표, 개별 노드에 대한 연결정도, 강도 및 거리 등의 미시수준의 연결성 분석 지표가 활용되고 있다. 또한 네트워크 내 각 노드의 영향력을 나타내는 분석 지표로는 중심성분석이 있으며, 노드 간 유사성을 기준으로 하위 집단을 분류하는 지표로 구조적 등위성에 바탕 한 분석이 이루어지고 있다(이수상, 2014).

이러한 배경으로 의미연결망 분석을 활용한 분석 연구들이 여러 학문으로 다양화 되며 점차 늘고 있다. 이승후·김학선(2018)은 ‘조리학원’과 관련된 빅데이터를 웹에서 수집하여 텍스트 마이닝 작업을 통해 상위 노출빈도를 도출하여, 의미연결망 분석을 실시하였다. Shuting과 Kim(2017)은 크루즈 여행에 관한 중국과 미국의 비교를 의미연결망 분석을 통해 실시하였다. 또한 최윤정·권상희(2014)는 세계적 이슈인 빅데이터와 관련된 신문기사를 바탕으로 의미연결망 분석을 실시하였다. 이런 선행연구를 통해 주요 단어 사이의 연결 상태, 네트워크 연결구조, 사용빈도 등을 통해 텍스트에 담긴 의미의 흐름을 파악하는데 유용한 방법론으로 활용됨을 알 수 있다.

2.4 오피니언 마이닝(Opinion Mining)

감성분석은 오피니언 마이닝(Opinion Mining)과 이모션 마이닝(Emotion mining)으로 구분한다. 이모션 마이닝은 텍스트 내에 반영된 슬픔, 행복, 놀람 등의 감정에 관한 분석방법이며(Yadollahi, Gan and Tan, 2017), 오피니언 마이닝은 사회적 현상에 다른 감성들을 시각화하기 위한 텍스트 마이닝의 한 기법으로 감성분석을 의미하며 텍스트

를 작성한 사람들의 태도, 의견, 성향과 같은 주관적인 데이터를 특정 주제에 대하여 긍정인지, 중립인지 또는 부정인지를 분류하는 기술이다(Turney, 2002). 또한 정형 및 비정형 텍스트 데이터의 주관적 정보를 식별하기 위해 자연언어처리(Natural Language Processing), 텍스트분석(Text Analysis) 기술 및 전산 언어학(Computational Linguistics)을 활용하는 것을 말한다. 일반적으로 오피니언 마이닝은 토픽(Topic) 또는 문서의 전반적인 문맥 극성(Polarity)과 관련하여 작성자의 태도를 판단하게 된다(Alrababah, Gan and Tan, 2017).

오피니언 마이닝 관련 연구로는 서민송·유환희(2017)가 오피니언 마이닝 기법을 이용한 사회적 재난의 시민 감성도를 분석하였으며, 최은정·김동근(2017)은 오피니언마이닝을 이용한 사용자 맞춤 장소 추천 시스템을 개발하였다. 그 이외에 여러 분야에서 리뷰 데이터로 오피니언 마이닝기법을 활용하여 고객의 반응을 분석하고 만족도를 알아보는 연구가 진행되고 있다.

III. 연구방법

본 연구는 부산 유스호스텔 ‘아르피나’와 관련된 2015년 1월부터 2017년 12월까지의 신문기사를 수집하고, 정제하는 과정은 SCTM(Smart Crawling & Text Mining)을 활용하여 탐색 키워드를 수집하였다. SCTM은 크롤링(crawling)방식을 이용해 web상의 데이터를 채널별로 자동 수집하여 처리해주는 솔루션으로, 연관성과 정확도에 따라 추출된 데이터를 수집하며 다양한 주제에 대한 오피니언 분석(Opinion acalysis)이 가능하며 텍스트 마이닝을 가능하게 해 주는 빅 데이터 솔루션이다. 수집된 비정형데이터를 분석용 데이터로 활용하기 위해 텍스트의 정제과정을 수행하였으며, 정제된 데이터를 텍스트 마이닝을 통해 빈도를 계산하여 주요단어들을 선별하였다. 선정된 단어들의 범주화를 통해 매트릭스 데이터로 만들어 시맨틱 네트워크 분석에 이용하였으며, UCINET6.0 프로그램을 사용하여 텍스트의 네트워크 중심성과 구조적 등위성(CONCOR)분석을 실시하였다. UCINET6.0과 함께 패키지화된 NetDraw를 활용하여 분석결과를 시각화함으로써 텍스트의 관계를 구체적으로 제시하고자 한다. 이를 통해 연결망의 구조와 연결강도를 파악하였으며 각 노드(단어)가 연결망에서 차지하는 구체적인 역할을 가늠하기 위해 본 연구에서는 Freeman(1978)의 중심성 개념 정리와 측정방법을 토대로 계량화한 지표인 연결정도 중심성(degree centrality)과 매개중심성

(betweenness centrality), 근접중심성(Closeness Centrality), 위세중심성(Eigenvector Centrality)을 살펴보았다.

또한, 2017년 5월 25일부터 2018년 5월 24일까지 1년간 네이버 블로그, 다음 블로그, 페이스북에서 ‘유스호스텔’관련 리뷰를 추출하여 ‘긍정’, ‘중립’, ‘부정’ 3단계로 감성분석을 실시하여 현재 유스호스텔 이용자들의 만족도를 확인하고자 한다.

따라서, 아르피나와 관련하여 어떤 단어가 빈번하게 사용되었으며, 단어들 사이의 관련성이 어떠한지를 밝힘으로써 유스호스텔의 현재 이미지와 앞으로 발전방향에 관한 시사점을 도출하고자 한다.

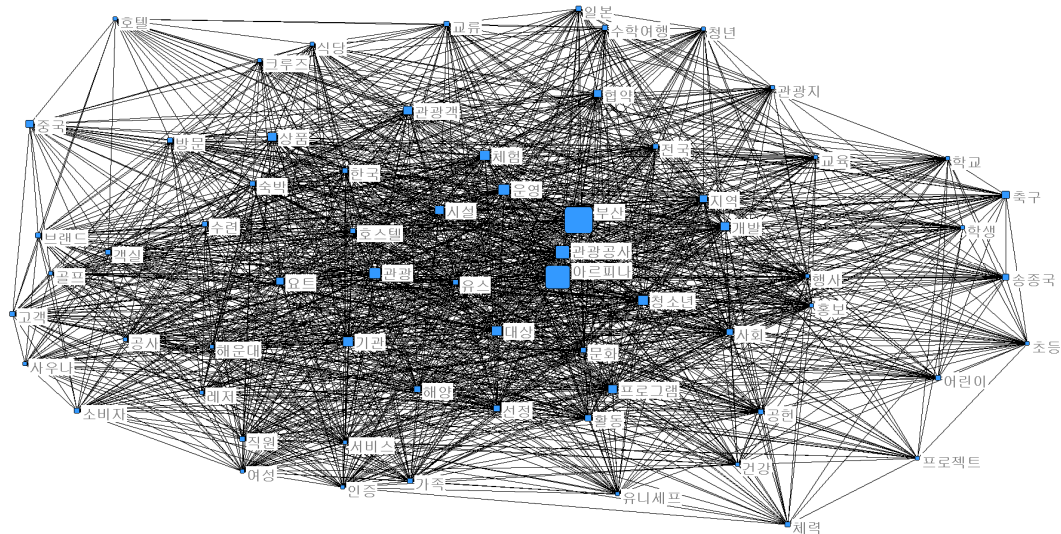
IV. 연구결과

4.1 데이터 빈도 (상위 주요 빈도)

‘아르피나’와 관련된 2015년 1월부터 2017년 12월까지의 신문 및 뉴스 보도 자료를 텍스트 마이닝 한 결과 총 12,760개의 단어가 출현되었다. 그 중 상위 빈도 200개의 단어를 추출하여 ‘을’, ‘를’, ‘에’ 등의 불필요한 조사 및 단어를 삭제하여 총 66개로 <표 1>과 같이 정리하였다.

조사결과, 키워드인 ‘부산’, ‘아르피나’를 제외한 상위 10개의 단어 관광공사, 관광, 운영, 청소년, 대상, 체험 등의 단어들이 누적빈도 28.4%로 웹 가시성(web visibility)이 높은 것으로 파악되었다. 또한 관광, 청소년, 호스텔, 수련, 숙박 등의 유스호스텔의 주목적인 청소년숙박시설과 관련된 단어들이 크게 부각되었으며, 프로그램, 축구, 활동, 문화 등의 다양한 프로그램 부분이 가시성이 높은 것으로 파악되었다. 66개의 주요 상위단어의 출현 빈도를 시각화한 분석결과는 <그림 1>과 같다.

텍스트마이닝 기법을 활용한 유스호스텔의 활성화 방안에 관한 연구



<그림 1> 주요 상위단어 출현빈도 시각화

<표 1> 상위 주요 빈도

순위	단어	빈도	%	순위	단어	빈도	%	순위	단어	빈도	%
1	부산	875	9.39	23	선정	119	1.28	45	행사	70	0.76
2	아르피나	765	8.21	24	송중국	117	1.26	46	인증	70	0.76
3	관광공사	365	3.92	25	어린이	109	1.17	47	서비스	69	0.75
4	관광	311	3.34	26	브랜드	107	1.15	48	식당	68	0.73
5	운영	293	3.15	27	전국	101	1.09	49	교육	65	0.7
6	청소년	274	2.94	28	숙박	99	1.07	50	학생	65	0.7
7	대상	264	2.84	29	유스	96	1.03	51	공사	65	0.7
8	체험	247	2.65	30	호스텔	96	1.03	52	홍보	63	0.68
9	기관	236	2.54	31	문화	94	1.01	53	여성	63	0.68
10	시설	227	2.44	32	수련	93	1	54	학교	62	0.67
11	관광객	219	2.35	33	고객	92	0.99	55	청년	61	0.66
12	프로그램	203	2.18	34	체력	90	0.97	56	유니세프	61	0.66
13	상품	198	2.13	35	한국	89	0.96	57	해운대	60	0.65
14	개발	195	2.1	36	일본	86	0.93	58	호텔	60	0.65
15	지역	192	2.06	37	방문	85	0.92	59	크루즈	57	0.62
16	축구	176	1.89	38	객실	82	0.88	60	건강	57	0.62
17	요트	162	1.74	39	수학여행	81	0.87	61	골프	56	0.61
18	협약	162	1.74	40	소비자	81	0.87	62	초등	56	0.61
19	중국	161	1.73	41	직원	79	0.85	63	프로젝트	56	0.61
20	해양	152	1.64	42	가족	78	0.84	64	레저	54	0.58
21	활동	149	1.6	43	공헌	78	0.84	65	관광지	52	0.56
22	사회	132	1.42	44	교류	73	0.79	66	사우나	50	0.54

4.2 주요 단어 중심성 분석 결과

연관 단어의 연결정도 중심성(Degree Centrality), 매개중심성(Betweenness Centrality), 근접중심성(Closeness Centrality), 위세중심성(Eigenvector Centrality)을 분석한 결과는 <표 2>와 같다. 아르피나에 대한 주요 단어의 중심성을 확인한 결과 관광공사, 운영, 관광, 청소년, 대상, 체험, 시설 등이 높음이 확인 되었으며, 이는 주요 단어의 빈도 결과와 비슷하게 나타난 것을 알 수 있다.

검색어인 부산과 아르피나를 배제한 주요 단어 연결정도 중심성을 확인한 결과, 운영, 관광, 청소년, 관광객, 관광공사, 중국, 대상, 프로그램, 기관, 시설 등의 순으로 도출되었으며, 매개중심성은 관광공사, 대상, 운영, 청소년, 지역, 관광, 체험, 시설, 개발, 문화 등의 순으로 나타났다. 또한 근접중심성 분석결과 관광공사, 운영, 청소년, 대상, 관광, 체험, 시설, 지역, 유스, 호텔 등의 순으로 도출되었고, 위세중심성은 운영, 관광공사, 관광, 관광객, 중국, 청소년, 축구, 대상, 프로그램, 활동 등의 순으로 나타났다. 빈도 순 상위 20개 단어의 연결정도중심성, 위세중심성, 근접중심성, 위세중심성을 동시에 비교한 결과 관광공사, 운영, 청소년, 관광 등의 단어들이 노출평균 대비 중요한 역할을 하는 것으로 분석되었다.

<표 2-1> 주요단어 중심성 분석 결과

연결정도 중심성			매개 중심성		
순위	단어	값	순위	단어	값
1	부산	0.082	1	부산	50.013
2	아르피나	0.073	2	아르피나	50.013
5	운영	0.043	3	관광공사	50.013
4	관광	0.035	7	대상	43.729
6	청소년	0.033	5	운영	41.295
11	관광객	0.033	6	청소년	40.779
3	관광공사	0.032	15	지역	40.406
19	중국	0.029	4	관광	39.038
7	대상	0.025	8	체험	36.08
12	프로그램	0.025	10	시설	30.987
9	기관	0.024	14	개발	30.894
10	시설	0.023	31	문화	30.805

텍스트마이닝 기법을 활용한 유스호스텔의 활성화 방안에 관한 연구

21	활동	0.023	12	프로그램	30.309
16	축구	0.022	22	사회	29.103
14	개발	0.021	21	활동	27.669
13	상품	0.018	29	유스	27.342
8	체험	0.017	30	호스텔	27.342
37	방문	0.017	9	기관	27.015
15	지역	0.016	28	숙박	26.654
24	송종국	0.016	17	요트	26.259

<표 2-2> 주요단어 중심성 분석 결과

근접 중심성			위세 중심성		
순위	단어	값	순위	단어	값
1	부산	100	1	부산	62.52
2	아르피나	100	2	아르피나	58.04
3	관광공사	100	5	운영	36.88
5	운영	97.015	3	관광공사	31.83
6	청소년	97.015	4	관광	30.62
7	대상	97.015	11	관광객	30.08
4	관광	95.588	19	중국	27.37
8	체험	92.857	6	청소년	27.36
10	시설	92.857	16	축구	22.02
15	지역	91.549	7	대상	21.50
29	유스	91.549	12	프로그램	20.92
30	호스텔	91.549	21	활동	19.74
31	문화	91.549	9	기관	19.50
9	기관	90.278	10	시설	19.41
12	프로그램	90.278	14	개발	18.49
28	숙박	90.278	24	송종국	16.39
14	개발	89.041	37	방문	15.91
17	요트	89.041	13	상품	15.77
20	해양	89.041	8	체험	14.52
23	선정	87.838	15	지역	14.38

4.3 CONCOR(CONvergence of iterated CORrelations) 분석

이수상(2014)은 CONCOR 분석이란 전체 키워드를 중심으로 유사점을 가진 키워드들의 클러스터(군집)를 찾아내어 키워드 동시출현 노드들 간의 형성된 관계를 파악하는 방법이라고 하였다. 도출된 단어들의 구조적 등위성을 분석하고자 텐드로그램을 통하여 3단계의 위계적 집단을 확인하였으며(그림 3), 그 중 2단계의 군집을 토대로 CONCOR분석을 수행하여 <그림 4>와 같은 분석결과를 도출하였다.

CONCOR 분석결과를 <표 3>과 같이 정리하였다. ‘고객 및 숙박시설’ 그룹은 청소년, 관광, 관광객, 숙박, 학생, 호스텔, 호텔 등의 단어들이 포함되어 있어, 아르피나의 이용고객층이 청소년, 학생인 것을 알 수 있었으며, ‘부대시설’ 그룹은 식당, 시설, 기관, 사우나, 골프 등의 단어들이 포함되어, 아르피나 내부에 스포츠시설이 활성화 되어 있음을 알 수 있다. ‘사회공헌 및 프로그램’ 그룹은 프로그램, 운영, 체험, 축구, 어린이, 행사 등의 단어들이 포함되어, 숙박시설로써의 역할 이외에 다양한 프로그램과 사회공헌 활동을 한다고 볼 수 있다. 마지막으로 ‘국제교류’ 그룹은 상품, 일본, 중국, 한국, 해운대, 크루즈 등의 단어들이 포함되어 있어 근접 국가와의 교류가 활발함을 알 수 있다.

<표 3> 군집분석 결과

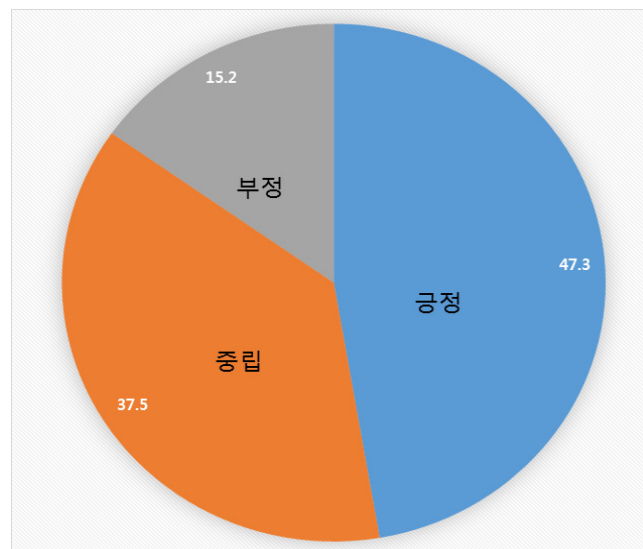
	주요 단어	군집명
1	청소년, 관광, 관광객, 숙박, 학생, 청년, 호스텔, 호텔, 등	고객 및 숙박시설
2	관광공사, 식당, 시설, 기관, 사우나, 골프, 요트, 등	부대시설
3	프로그램, 운영, 체험, 축구, 어린이, 행사, 등	사회 공헌 및 프로그램
4	상품, 일본, 중국, 한국, 교류, 해운대, 크루즈, 등	국제교류

1. 고객 및 숙박시설



4.4 오피니언 마이닝(Opinion Mining)

오피니언 마이닝은 주로 다양한 소셜 미디어 콘텐츠로부터 상품 및 서비스의 선호도, 사회적 사건이나 정치 이슈 등에 대한 대중들의 의견을 분석하는 데 적용되어 왔고, 관련 연구들 또한 이러한 의견들의 추출, 분류, 이해, 평가를 위한 방법론에 관한 논의를 주로 다루고 있다(Zhao, Qin and Liu, 2010). 본 연구에서는 ‘아르피나’에 관한 감성분석을 하려하였으나, 관련 데이터 양의 한계로 ‘유스호스텔’에 관련된 자료를 사용하여 감성분석을 실시하였다. 2017년 5월 25일부터 2018년 5월 24일까지 1년간 네이버 블로그, 다음 블로그, 페이스북에서 ‘유스호스텔’ 관련 리뷰 및 게시글 20,912를 추출하여 긍정, 중립, 부정 3단계로 감성분석한 결과 긍정이 9,891개(47.3%), 중립이 7,842개(37.5%), 부정이 3,179(15.2%)로 긍정적인 의견이 가장 많았다. 긍정적인 의견에는 저렴한 가격, 깨끗한 시설, 안전함 등의 내용이 대부분을 차지했고, 부정적 의견으로는 불친절함, 시설의 노후, 주변 환경, 위치적 입지, 공용화장실 등의 문제점이 다소 나타났다. <그림 4>는 감성분석 결과를 나타내는 표이다. <그림 5>은 소셜메트릭스(Social metrics)에서 제공하는 감성분석 프로그램으로 ‘아르피나’와 ‘유스호스텔’에 관한 감성분석을 비교한 것이다. 두 결과의 공통점은 다른 숙박시설보다 가격이 저렴하다는 것에 긍정적인 반응을 나타내었다.



<그림 4> 유스호스텔의 감성분석 결과



<그림 5> 아르피나와 유스호스텔의 감성분석 비교

V. 결론 및 시사점

본 연구는 ‘부산 아르피나’ 키워드를 웹 크롤링 및 텍스트 마이닝을 통해 분석하고자 데이터 수집 및 처리 프로그램인 SCTM을 활용하여 2015년 1월부터 2017년 12월까지의 신문 및 뉴스 보도 자료를 수집하여 분석하였다. 3년간의 수집된 키워드는 총 12,760개가 수집되었으며, 그 중 상위 빈도 200개 단어를 추출하여 중복된 의미나 불필요한 조사 및 단어를 제거하여 총 66개로 도출 분석하였다. 정제된 데이터는 UCINET6.0과 패키지화된 NetDraw를 활용하여 중심성분석(Centrality Analysis)과 도출된 단어들의 구조적 등위성을 분석하고자 CONCOR분석을 수행하였다. 또한 이용고객의 만족도를 알아보기 위해 2017년 5월 25일부터 2018년 5월 24일까지 1년간 네이버 블로그, 다음 블로그, 페이스북에서 ‘유스호스텔’ 관련 리뷰 및 게시글 20,912를 추출하여 긍정, 중립, 부정 3단계의 감성분석을 실시하였다.

분석결과 관광공사, 관광, 운영, 청소년, 대상, 체험 등의 단어들이 웹 가시성(web visibility)이 높은 것으로 파악되었다. 군집분석 결과 청소년, 관광, 관광객, 숙박, 학생, 호스텔, 호텔 등은 ‘고객 및 숙박시설’그룹이 형성되어 아르피나의 이용고객층이 청소년, 학생인 것을 알 수 있었으며, ‘부대시설’ 그룹은 식당, 시설, 기관, 사우나, 골프 등의 단어들이 포함되어, 아르피나 내부에 스포츠시설이 활성화 되어있음을 알 수 있었다. ‘사회공헌 및 프로그램’ 그룹은 프로그램, 운영, 체험, 축구, 어린이, 행사 등의 단어들이 포함되어, 숙박시설로써의 역할 이외에 다양한 프로그램과 사회공헌 활동을 한다

고 볼 수 있다. 마지막으로 ‘국제교류’ 그룹은 상품, 일본, 중국, 한국, 해운대, 크루즈 등의 단어들이 포함되어 있어 근접 국가와의 교류가 있음을 알 수 있었다.

검색어인 ‘부산’과 ‘아르피나’를 배제한 주요 단어 연결정도 중심성을 확인한 결과, 운영, 관광, 청소년, 관광객, 관광공사, 중국, 대상, 프로그램, 기관, 시설 등의 순으로 도출되었으며, 매개중심성은 관광공사, 대상, 운영, 청소년, 지역, 관광, 체험, 시설, 개발, 문화 등의 순으로 나타났다. 또한 근접중심성 분석결과 관광공사, 운영, 청소년, 대상, 관광, 체험, 시설, 지역, 유스, 호스텔 등의 순으로 도출되었고, 위세중심성은 운영, 관광공사, 관광, 관광객, 중국, 청소년, 축구, 대상, 프로그램, 활동 등의 순으로 나타났다. 빈도 순 상위 20개 단어의 연결정도중심성, 위세중심성, 근접중심성, 위세중심성을 동시에 비교한 결과 관광공사, 운영, 청소년, 관광 등의 단어들이 노출평균 대비 중요한 역할을 하는 것으로 분석되었다.

감성분석 결과는 2017년 5월 25일부터 2018년 5월 24일까지 1년간 네이버 블로그, 다음 블로그, 페이스북에서 ‘유스호스텔’ 관련 리뷰 및 게시글 20,912를 추출하여 긍정, 중립, 부정 3단계로 감성분석 한 결과 긍정이 9,891개(47.3%), 중립이 7,842개(37.5%), 부정이 3,179(15.2%)로 긍정적 의견이 가장 많았다. 긍정적 의견에는 저렴한 가격, 깨끗한 시설, 안전함 등의 내용이 대부분을 차지했고, 부정적 의견으로는 불친절함, 시설의 노후, 주변 환경, 위치적 입지, 공용화장실 등의 문제점이 다소 나타났다.

이러한 분석결과를 통해 본 연구에서는 다음과 같은 학문적 시사점을 도출 하였다.

첫째, 2000년대 이후 유스호스텔의 뚜렷한 발전이 없었으며, 관련연구 또한 미흡한 상황에 최근 주목 받고 있는 빅데이터를 활용하여 현재 문제점과 만족도를 분석하여 향후 발전 방향을 시사하는 것에 의미가 있다. 둘째, 유스호스텔을 대상으로 고객의 감정을 파악하는 오피니언 마이닝 기법을 이용하여 고객의 만족도를 파악했다는 것에 이론적 시사점이 있다.

다음 실무적 시사점은 첫째, 오피니언 마이닝 분석에 사용된 리뷰 정보에서 다소 노출된 부정적 의견은 ‘직원의 불친절’, ‘가격대비 불만족’ 등이 나타났다. 숙박시설이 다양화 되고, 저가의 시설이 증가하고 있는 시점 유스호스텔은 호텔만큼의 서비스와 인지도에 미치지 못하는 것이 현실이다. 따라서 유스호스텔이 활성화되기 위해서는 단체뿐만 아니라 개인에게도 조금 더 저렴한 숙박비를 측정해야 할 것이다. 또한 직원의 서비스교육에 투자해 저가와의 차별화를 둘 필요가 있다. 둘째, 유스호스텔의 가장 부족한 부분은 홍보 전략이다. 유스호스텔의 시작은 청소년 수련시설이지만 인구의 급격

한 감소와 해외 수학여행이 증가하는 현 시대에 맞게 청소년뿐만 아니라 국내 여행객에게 홍보할 필요가 있다. 또한 외국 학교와의 교류로 청소년 단체 고객을 확보 한다면 국내고객 감소를 보완할 수 있을 것이다. 셋째, 텍스트 마이닝에서 군집분석 결과 국제적 교류활동에 관한 군집이 형성되었다. 즉, 현재에도 국제적 교류가 있지만 중국, 일본 등 아시아지역으로 한정되어 있음을 알 수 있다. 또한 국내·외 고객을 확보하기에는 타 숙박시설과 비교해 마케팅 활동이 많이 부족하다. 따라서 전 세계적으로 알릴 수 있는 SNS마케팅이 활발히 이루어질 필요가 있다. 아르피나의 큰 장점인 다양한 스포츠센터와 부대시설을 앞세워 경쟁우위를 점할 수 있는 마케팅과 해운대라는 입지적 조건을 강조한다면 고객유치에 큰 도움이 될 것이다. 또한, 이승후·김연순·김학선(2018)의 SNS특성과 소비자태도에 관한 연구에서 SNS정보를 고객에게 즐겁고 쉽게 이해할 수 있도록 제공한다면 긍정적인 소비자태도를 나타낸다고 하였다. 따라서 SNS마케팅 시 아르피나의 소개 및 광고뿐만 아니라 부산의 관광지과 맛집을 소개한다면 SNS마케팅을 극대화할 수 있을 것이다. 마지막으로, 아르피나는 부산 관광공사에서 운영하는 공기업이란 특성을 활용하여 부산의 태마관광과 음식관광을 패키지화 하여 상품으로 발전한다면 아르피나의 홍보효과를 넘어 부산관광을 발전시킬 수 있는 계기가 될 것이다.

본 연구는 아르피나에 관한 데이터부족으로 유스호스텔 전체에 관한 감성분석을 실시하여, 아르피나 고객의 만족도를 나타내지 못한 것에 한계점이 있다. 또한 주 이용고객층인 청소년인데 반해 추출된 데이터는 연령을 구분하지 않고 분석되어 아르피나의 만족도를 대표할 수 없다. 이러한 연구의 한계점을 보완하기 위하여 아르피나 이용고객을 대상으로 설문조사를 실시하여 빅데이터와 비교분석하는 종단연구가 필요할 것으로 사료된다.

참 고 문 헌

- 김도경·김인신(2017), “텍스트 마이닝을 이용한 온라인 리뷰의 호텔 선택속성 분석,” 관광학연구, 제41권, 제9호, pp. 109-127.
- 김상국(2011), “사회 연결망분석(SNA)과 산업공학,” ie 매거진, 제18권 제1호, pp. 24-32.
- 김주영·김동수(2017), “텍스트 마이닝 기반의 온라인 상품 리뷰 추출을 통한 목적별 맞춤화

산업혁신연구 (제34권 2호), 2018. 6. pp. 267~288.

- 정보 도출 방법론 연구,” 한국전자거래학회지, 제21권, 제2호, pp. 151-161.
- 김학선(2017), “빅데이터를 활용한 음식관광관련 의미연결망 분석의 탐색적 적용,” 한국조리학회지, 제23권, 제4호, pp. 22-32.
- 민덕기(1968), “유스호스텔의 특성과 한국에 있어서의 전망,” 관광학회, 제2권, pp. 93-94.
- 서민송·유환희(2017), “오피니언 마이닝 기법을 이용한 사회적 재난의 시민 감성도 분석,” 한국지형공간 정보학회지, 제25권, 제1호, pp. 37-46.
- 손병모(2011), “관광지 선택속성이 관광객의 지각된 가치와 충성도에 미치는 영향,” 대한관광경영학회지, 제26권, 제1호, pp. 201-228.
- 양지윤(2018), “텍스트 마이닝을 통한 아이코닉 건축과 장소 브랜딩의 상관성에 관한 연구,” 홍익대학교 박사학위논문.
- 이수상(2014), “언어 네트워크 분석 방법을 활용한 학술논문의 내용분석,” 정보관리학회지, 제31권, 제4호 pp. 49-68.
- 이승후·김연순·김학선(2018), “수정된 기술수용모델을 활용한 일반음식점 SNS특성이 소비자태도에 미치는 영향에 관한 연구,” 한국조리학회지, 제24권, 제4호, pp. 22-34.
- 이승후·김학선(2018), “빅데이터를 활용한 조리학원의 의미연결망 분석에 관한 연구,” 한국조리학회지, 제24권, 제3호, pp. 164-173.
- 이희열·권점옥(2011), “대학생들의 유스호스텔에 대한 이미지가 이용에 관한 만족도 및 행동의도에 미치는 영향분석,” 외식경영연구, 제14권, 제4호, pp. 337-358.
- 전명기(1998), “청소년 수련시설 특성화 방안 연구,” 한국청소년정책연구원 연구보고서, pp. 1-164.
- 전인오·강명석(2013), “숙박시설의 지각된 가치가 관광만족도에 관한 연구,” 디지털융복합연구, 제11권, 제10호, pp. 79-94.
- 전채남·서일원(2013), “빅데이터 분석의 기술마케팅 활용에 관한 연구: 잠재 수요기업 발굴을 중심으로,” 마케팅논집, 제21권 제2호, pp. 181-203.
- 최영준·오홍철(2010), “국내유스호스텔 활성화를 위한 탐색적 연구: 부산 아르피나 유스호스텔 사례를 중심으로,” 호텔관광연구, 제12권, 제2호, pp. 162-176.
- 최윤정·권상희(2014), “빅데이터관련 신문기사의 의미연결망 분석,” 사이버커뮤니케이션학보, 제31권 제1호, pp. 241-286.
- 최은정·김동근(2017), “오피니언마이닝을 이용한 사용자 맞춤 장소 추천 시스템,” 한국정보통신학회논문지, 제21권, 제11호, pp. 2043-2051.

한국유스호스텔연명(2011), <http://www.kyha.or.kr/default.asp>

Alrababah, S. A. A., Gan, K. H. and Tan, T. P.(2017), “Mining opinionated product features using WordNet lexicographer files,” *Journal of Information Science*, Vol. 43, No. 6, pp. 769-785.

Biesanz, J.(1941), “Nazi Influence on German Youth Hostels,” *Social Forces*, pp. 554-559.

Chiang, D. M. H., Lin, C. P. and Chen, M. C.(2011), “The adaptive approach for storage assignment by mining data of warehouse management system for distribution centres,” *Enterprise Information Systems*, Vol. 5, No. 2, pp. 219-234.

Freeman, L. C.(1978), “Centrality in social networks conceptual clarification,” *Social Network*, Vol. 1, No. 2, pp. 215-239.

Hwang, D. Y. and Hwang, G. E.(2016), “Examining of semantic map of humanities contents through semantic network analysis,” *Korea Humanities Content Society*, Vol. 43, pp. 229-255.

Nash, R., Thyne, M. and Davies, S.(2006), “An investigation into customer satisfaction levels in the budget accommodation sector in Scotland: A case study of backpacker tourists and the Scottish Youth Hostels Association,” *Tourism Management*, Vol. 27, No. 3, pp. 525-532.

Shuting, T. and Kim, H. S.(2017), “A study of comparison between cruise Tours in China and USA through big data analytics,” *Culinary Science & Hospitality Research*, Vol. 23, No. 6, pp. 1-11.

Turney, P. D.(2002), “Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews,” *Proceedings of the 40th annual meeting on association for computational linguistics*, Association for Computational Linguistics, pp. 417-424.

Wang, W. and Rada, R.(1998), “Structured hypertext with domain semantics,” *ACM Transactions on Information Systems (TOIS)*, Vol. 16, No. 4, pp. 372-412.

Wasserman, S. and Faust, K.(1994), “Social network analysis: Methods and applications,” (Vol. 8). *Cambridge university press*.

Yadollahi, A., Shahraki, A. G. and Zaiane, O. R.(2017), “Current state of text sentiment analysis from opinion to emotion mining,” *ACM Computing Surveys (CSUR)*,

산업혁신연구 (제34권 2호), 2018. 6. pp. 267~288.

Vol. 50, No. 2, pp. 25.

Zhao, Y. Y., Qin, B. and Liu, T.(2010), "Sentiment analysis," *Journal of Software*, Vol. 21, No. 8, pp. 1834-1848.

05월/30일/2018 접수, 06월/12일/2018 1차 수정, 06월/19일/2018 2차 수정,
06월/30일/2018 게재

A Study on Activation Plan of Youth Hostel Using Text Mining Technique

: Focusing on Busan Youth Hostel ARPINA

Lee, Seung-Hoo¹ · Kim, Hak-Seon²

Abstract

Arpina is a youth hostel in Busan operated by Busan Tourist Corporation and opened in July 2004. It is an institution equipped with physical facilities and training facilities as well as accommodation facilities. Currently, various types of accommodation are developed and low-priced lodging business is being activated. Therefore, in order to identify the current problems of stagnant youth hostels, this study conducted text mining analysis using ARPINA related newspaper articles for three years from 2015 to 2017, A review of Facebook and blogs of youth hostel visitors for the past one year was conducted and Opinion Mining Analysis was conducted. As a result, SNS marketing was significantly lacking compared to other accommodation facilities, and main customers were limited to groups. The marketing strategy and the activation plan for the sales increase were suggested.

Keywords: youth hostels, text mining, meaning network, opinion Mining, arpina

1 Ph D candidate. School of Hospitality & Tourism Management, Kyungsoo University. 314-79 Daeyeon-dong, Nam-gu, Busan 48434 tel: 051-663-4473, e-mail: pingklo@naver.com (First Author)

2 Associate Professor. School of Hospitality & Tourism Management, Kyungsoo University. 314-79 Daeyeon-dong, Nam-gu, Busan 48434 tel: 051-663-4473, e-mail: kims@ks.ac.kr (Corresponding Author)

산업혁신연구 (제34권 2호), 2018. 6. pp. 267~288.

저자소개

이승후



현재 경성대학교 호텔관광외식경영학과 박사과정
최종학력: 경성대학교 외식서비스경영전공 석사
관심분야: 관광 빅데이터, 음식관광, 외식서비스

김학선



현재 경성대학교 호텔관광외식경영학과 부교수
최종학력: Texas Tech university 박사
관심분야: 관광 빅데이터, 음식관광, 응용영양