

Comparative Analysis of Issues of Consumer Interest in Busan Osiria Tourism Complex Using Big Data: Focusing on COVID-19 During and After Social Distancing

빅데이터를 활용한 부산 오시리아 관광단지의 소비자 관심 이슈 비교분석: COVID-19 사회적 거리두기 해제 전·후 비교를 중심으로

Yong-Hwa Hong¹, Ha-Kyun Kim²

홍용화¹, 김하균²

¹ Doctoral Student, Graduate School of Information Systems, Pukyong National University, Korea, dic301@naver.com

² Professor, Division of Business, Pukyong National University, Korea, kimhk@pknu.ac.kr

Corresponding author: Ha-Kyun Kim

Abstract: The Osiria Tourism Complex is a project to develop a 3.66 million square meter beach area in Daeyeon-ri and Sirang-ri, Gijang-gun, Busan, into a luxury multi-tourism complex that includes accommodation, leisure, and shopping facilities. In order to suggest future directions by examining the development trend of the Osiria Tourism Complex, which has entered the completion stage, and identifying how consumers' perceptions are changing after the COVID-19 social distancing period and lifting through keywords shared on social media do. Data collection was carried out using Textom on the condition that 'Osiria' and 'Tourism Complex' were included at the same time, on March 22, 2020, during the COVID-19 social distancing period, for Naver, a domestic portal site, and Daum. It was collected in two groups: ~ October 31, 2021 (589 days) and November 1, 2021 ~ October 31, 2022 (365 days) after COVID-19 social distancing was lifted. In order to extract meaningful information, words were edited using N-gram and Excel. Data on the top 50 words was generated through search frequency analysis, and network analysis was performed using this data. NetDraw were used for network visualization. CONCOR analysis was performed using UCINET 6 and NetDraw to understand the characteristics of groups composed of similar words. As a result of the study, after the COVID-19 social distancing was lifted, new words such as Lotte World, cafe, wedding, ocean view, party, restaurant, and lodging appeared, confirming that tourism activities were activated.

Keywords: Big Data, Text Mining, Osiris Tourism Complex, Network Analysis, CONCOR Analysis, COVID-19, Social Distancing

요약: 오시리아 관광단지는 부산시 기장군 대변리·시랑리 일대 해변 366만㎡를 숙박·레저·쇼핑 시설 등이 포함된 사계절 체류형 명품 복합관광단지로 조성하는 사업이다. 완성단계에 접어든 오시리아 관광단지에 대해 발전 동향을 살펴보고, SNS로 공유되는 키워드를 통해 COVID-19 사회적 거리두기 해제 전·후에 소비자의 인식이 어떻게 변화되고 있는지를 파악

Received: November 07, 2022; 1st Review Result: December 23, 2022; 2nd Review Result: January 23, 2023
Accepted: February 28, 2023

하여 향후 방향성을 제안하고자 하였다. 데이터 수집은 텍스트롬(Textom)을 이용하여 ‘오시리아’와 ‘관광단지’가 동시에 포함된 조건으로 국내 포털사이트인 네이버, 다음을 대상으로 COVID-19 사회적 거리두기 기간인 2020년 3월 22일~2021년 10월 31일(589일)과 COVID-19 사회적 거리두기 해제 후 2021년 11월 1일~2022년 10월 31일(365일) 두 그룹으로 수집하였다. 유의미한 정보를 추출하기 위하여 N-gram과 엑셀을 이용하여 단어의 편집을 진행하였다. 검색 빈도 분석을 통해 상위 50개 단어에 대한 매트릭스 데이터를 생성하였고, 이를 이용하여 네트워크 분석을 수행하였다. 네트워크 시각화에는 UCINET 6 및 NetDraw를 사용하였다. 유사한 단어들로 구성된 그룹의 특성을 파악하기 위해 UCINET 6와 NetDraw를 활용하여 CONCOR 분석을 수행하였다. 연구 결과, COVID-19 사회적 거리두기 해제 후에는 롯데월드, 카페, 웨딩, 오션뷰, 파티, 맛집, 숙박 등의 새로운 단어들 등장하여 관광 활동이 활성화된 것을 확인하였다.

핵심어: 빅데이터, 텍스트 마이닝, 오시리아 관광단지, 네트워크 분석, CONCOR 분석, COVID 19, 사회적 거리두기

1. 서론

오시리아 관광단지는 4조 원을 투입해 부산시 기장군 대변리·시랑리 일대 해변 366만㎡를 숙박·레저·쇼핑 시설 등이 포함된 사계절 체류형 명품 복합관광단지 조성하는 사업이다. 1999년 제4차 국토종합계획에 해당 사업 계획이 반영되면서 추진되기 시작한 오시리아 관광단지는 지난 10년간 계속 투자자가 선정되었다가 투자를 포기하는 등 여러 가지 굴곡이 있었던 사업 계획이다. 현재 테마파크의 롯데월드 어드벤처 부산이 2019년 5월에 착공하여 2022년 3월 31일에 개장하였다[1].

2019년 12월 중국 우한에서 처음 발생하여 전 세계로 확산된 새로운 바이러스(이하 COVID-19)로 인하여 실내외에서 마스크를 쓰고, 사회적 거리두기를 하는 새로운 일상이 시작되었다. 2020년 1월 20일 국내에서 첫 확진자가 발생한 이후로 COVID-19의 대응을 위해 2020년 3월 23일부터 2021년 10월 31일까지 1년 8개월간 개인별 방역 수칙 준수, 다인 이용시설의 운영 제한, 사적 모임 금지, 재택 근무 등의 사회적 거리두기를 시행하였다[2].

관광은 감염병, 천재지변, 사회환경 등의 사회적 재난 요인들에 영향을 많이 받는 산업이라 할 수 있다. 2015년 메르스 발병 시 한국을 방문한 외래 관광객의 26.5%~53.5%의 감소와 국내 관광객의 22.2%가 감소한 현상은 감염병이 관광산업에 부정적인 영향을 끼치는지 보여주는 대표적인 사례이다[3]. 세계관광기구(UNWTO)는 2020년 국가 간 관광객은 전년 보다 80% 감소할 것이며, 관광객 통계를 시작한 1950년 이후 최악의 상황이라고 진단했다. 실제로 2020년 해외 여행객 수는 전년 대비 74% 감소하였으며 이번 위기는 빨라도 2022년 여름이 지난 후에야 회복될 것으로 예상하였다. 또한, 여행제한조치를 점진적으로 해제하고 COVID-19 백신이나 치료제가 개발되어 원활하게 보급된다고 해도 여행객이 COVID-19 이전 수준으로 회복되려면 2년 6개월에서 최대 4년이 소요될 것으로 예상하고 있다[4].

최근 관광객들은 디지털 생활의 일상화로 인해 사회관계망서비스(Social Network Service) 사용 비중이 급격히 증가하였으며, 이들의 SNS 이용으로 인하여 텍스트 데이터, 이미지 데이터, 로그기록 데이터, 위치 데이터 등의 비정형화된 데이터가 신속하고 방대한

규모로 축적되고 있으며, 이를 소셜미디어 빅데이터로 정의할 수 있다[3]. 소셜미디어 빅데이터 분석은 기존의 데이터에서 도출할 수 없었던 포괄적인 통찰과 다각적인 측면에서 가치를 창출하여, 많은 분야에서 활용되고 있다[4].

본 연구는 오랜 시간과 여러 굴곡 끝에 완성단계에 접어든 오시리아 관광단지에 대한 발전 동향을 살펴보고, COVID-19로 인해 고통받고 있는 어려운 현실 속에서 최근 SNS로 공유되는 키워드를 통해 사회적 거리두기 기간과 사회적 거리두기 해제 후에 관광 소비자의 인식이 어떻게 변화되고 있는지를 분석하여 향후의 관광산업의 정책방향을 제안하고자 한다.

2. 이론적 배경

2.1 COVID-19와 사회적 거리두기(Social distancing)

COVID-19는 2019년 12월 중국 우한시에서 처음 발생하여 전 세계로 확산된 새로운 바이러스에 의한 호흡기 감염질환이다. COVID-19는 감염자의 비말이 호흡기나 눈·코·입의 점막을 통해 침투할 때 전염된다. 감염되면 대략 2~14일의 잠복기를 거친 뒤 발열과 기침, 호흡곤란 등 호흡기 증상과 폐렴 증상이 주로 나타나지만 무증상 사례도 나오고 있다.

사회적 거리두기는 우리나라에서 COVID-19의 급속한 확산에 따라 2020년 3월 22일부터 2021년 10월 31일까지 시행한 정책이다. 세계보건기구에서는 사회적 거리두기라는 표현 대신 물리적으로 거리를 두기만 하면 된다는 것을 강조하기 위해 물리적 거리두기(physical distancing)라는 표현을 권장하고 있는데, 이는 ‘사회활동’이 아니라 ‘비말이 될 수 있는 거리’에 있기 때문이다. 사회적 거리두기라고만 하게 되면 직장 등 일체의 경제활동까지도 포기하라는 식으로 오해할 여지가 있으나, ‘물리적 거리두기’라 하면 현장에서의 불필요한 회의 등을 줄이고 재택 근무를 하라는 말이 된다[2].

2.2 빅데이터와 텍스트 마이닝(Big data and Text mining)

빅데이터란 기존의 데이터와 비교하여 규모가 방대하여 이전의 방법이나 도구로는 처리하기 어려운 정형, 비정형 데이터를 말한다. 빅데이터의 특징은 일반적으로 다음과 같이 세가지로 설명할 수 있다. 첫째, 데이터의 크기(Volume)로 물리적 크기의 대형화를 의미한다. 둘째, 데이터의 속도(Velocity)로 대용량의 데이터를 빠르게 실시간으로 처리할 수 있어야 함을 뜻하며 데이터의 생성, 저장과 시각화되는 과정이 빠르게 이루어져야 한다. 셋째, 다양성(Variety)은 정형 데이터뿐만 아니라 이미지, 사진, 동영상, SNS 데이터와 같은 비정형 데이터를 포함된다. 최근 위의 세가지에 더해서 정확성(Veracity), 가치(Value) 등의 새로운 속성이 등장하고 있다[1].

텍스트 마이닝이란 SNS 데이터와 같은 비정형 데이터를 자연어 처리(NLP) 기술을 이용하여 비정형과 반정형 텍스트 데이터로부터 유의미한 정보를 생산 만들어 내거나 가공하는 기술이다. 텍스트 마이닝 기술을 이용하여 비정형 데이터가 함의하고 있는 의미를 추출하고, 데이터를 이용하여 정보 간의 연관성과 범위, 위치 등을 활용하여 소비자 분석이 가능하다[1].

2.3 네트워크 분석(Network Analysis)

네트워크 분석은 텍스트 데이터 내에서 어떤 단어를 사용했으며 어떤 순서로 배치했는가에 대한 구조적 관계를 분석하고, 구조가 가진 패턴 및 의미를 분석하여 문장의 형태 속에 내재되어 있는 의미를 도출하는 분석 기법이다. 네트워크 중심성은 연결 정도 중심성, 근접 중심성, 매개 중심성으로 구분하며 연결 정도 중심성은 하나의 단어에 직접적으로 연결된 단어의 개수를 의미하고 단어의 개수가 많을수록 중심성이 높아진다. 근접 중심성은 간접적으로 연결되어 있는 모든 단어 간의 거리를 바탕으로 측정되며, 매개 중심성은 네트워크 내의 한 단어가 다른 단어들 사이에 위치하는 정도로 네트워크 내에서 어디에 위치하느냐를 파악하여 영향력을 확인할 수 있다. 분석된 결과를 시각화하기 위해서 UCINET 6, NetDraw 등의 다양한 분석 도구를 활용할 수 있다[5].

2.4 구조적 등위성 분석(CONvergence of iterated CORrelations: CONCOR)

구조적 등위성 분석은 구조 등가 분석(structural equivalence analysis)을 진행하는 데 있어서 가장 흔하게 사용되는 통계 기법이며, 이를 통해 키워드들의 동시 출현 매트릭스의 상관관계를 분석할 수 있다. 또한 상관관계 분석을 반복하여 수행하여, 노드 간 군집을 형성하여 군집 간 관계를 증명할 수 있게 하고, 이 과정에서 유사 집단끼리 묶어 군집분석과 같은 효과를 보여준다[6]. 구조적 등위성 분석을 연결 정도 중심성 수치를 기초로 복잡한 네트워크 상의 군집화된 그룹을 찾아내는데 용이한 방법이며, 단어 간 연결 관계에서 유사성이 높을수록 노드 간의 구조적 등위성의 수치는 크다. 또한 구조적 등위성을 시각화하는 모델에서 도형이 크기가 클수록 노드의 중심도는 높고, 연결선의 굵기는 강한 연결성을 갖고 있음을 알려준다[7].

3. 연구 결과

3.1 데이터 수집

[표 1] 수집 데이터 정보

[Table 1] Collected Data Information

구분	COVID-19 사회적 거리두기 기간	COVID-19 사회적 거리두기 해제 후
검색어	‘오시리아’+ ‘관광단지’	
수집 범위	네이버(블로그, 카페, 뉴스, 웹 문서, 지식IN, 학술정보 전체) 다음(블로그, 카페, 뉴스, 웹 문서)	
수집대상기간	2020년 3월 22일~2021년 10월 31일 (589일)	2021년 11월 1일~2022년 10월 31일 (365일)
수집 도구	Textom (V6.0)	
수집 량	5,753 건	5,878 건

데이터 수집은 텍스트롬(Textom)을 이용하여 ‘오시리아’와 ‘관광단지’가 동시에 포함된 조건으로 국내 포털사이트인 네이버(웹 문서, 블로그, 카페, 뉴스, 지식IN, 학술정보 전체), 다음(웹 문서, 블로그, 카페, 뉴스)을 대상으로 COVID-19 사회적 거리두기 기간인 2020년 3월 22일~2021년 10월 31일(589일)과 COVID-19 사회적 거리두기 해제 후 2021년 11월

1일~2022년 10월 31일(365일) 두 그룹으로 수집하였고, 수집 데이터 정보는 [표 1]과 같다.

3.2 데이터 정제

수집한 데이터는 제목과 본문의 내용 기준으로 중복을 제거하고, 한글 형태소 분석기인 Espresso-K를 이용하여 명사만 한정하여 키워드 ‘오시리아’의 앞뒤로 3단어를 선택하여 분석 리스트를 작성하였다. 유의미한 정보를 추출하기 위하여 언어 모델인 N-gram과 엑셀을 이용하여 단어를 편집하였으며, 단어 묶음을 기준으로 수정하였다. 두 단어가 등장할 때 웹 데이터에 ‘롯데’와 ‘백화점’으로 각각 수집된 자료를 단어 간의 관계를 파악하여 ‘롯데백화점’이라는 한 단어로 수정하였다. 지역 명칭을 표시하는 ‘부산’, ‘부산시’, ‘부산광역시’는 ‘부산’으로, ‘기장’, ‘기장군’, ‘기장읍’은 ‘기장’으로 수정하였다[1].

3.3 빈도 및 TF-IDF 분석

‘오시리아’ + ‘관광단지’의 검색어로 추출된 단어는 아래의 [표 2]와 같이 빈도순으로 COVID-19 사회적 거리두기 기간과 해제 후를 비교하여 상위 50개를 정리하였다. 두 기간에 공통적으로 나타난 1위부터 4위까지의 검색어와 지역명을 제외하면, 사회적 거리두기 기간에는 ‘관광’, ‘동부산’, ‘테마파크’, ‘스타테라스’, ‘부산도시공사’, ‘일광신도시’ 등이 높은 빈도를 보였고, 사회적 거리두기 해제 후에는 ‘관광’, ‘해운대’, ‘동부산’, ‘테마파크’, ‘카페’, ‘롯데월드어드벤처’ 등이 높은 빈도를 보였다.

단순히 빈도가 높은 단어가 더 중요하다고 판단하는 데에는 한계가 있어 보다 정밀한 분석을 위해 TF-IDF 분석을 수행하였다. TF-IDF 분석은 TF(Term Frequency)와 IDF(Inverse Document Frequency)의 곱한 값으로, 특정 문서내에서 단어가 얼마나 중요한지를 나타내는 통계적 기법이다. 이는 여러 문서에 자주 등장하는 단어가 중요하지 않을 가능성이 높다는 가정을 할 수 있는데, 이는 단순히 단어의 출현 빈도로만 중요도를 측정하는 것이 아니라 문서에서 많이 출현하는 단어를 찾는 방법이다[4].

아래 [표 2]를 살펴보면 COVID-19 사회적 거리두기 기간의 경우 5위(관광), 7위(테마파크), 11위(소더비), 13위(시니어타운) 등이, 사회적 거리두기 해제 이후에는 5위(관광), 9위(카페), 11위(롯데월드), 12위(반얀트리) 등이 빈도순과 다르게 문서상의 중요성이 다르게 나타나고 있음이 확인되었다.

[표 2] COVID-19 사회적 거리두기 해제 전·후의 단어 빈도 순위

[Table 2] Word Frequency Ranking Before and After the COVID-19 Social Distancing Period

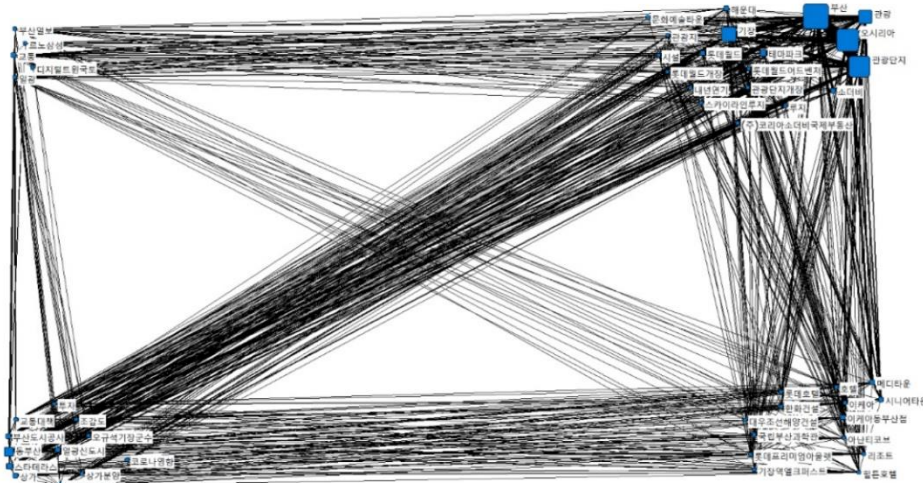
순위	COVID-19 사회적 거리두기 기간			COVID-19 사회적 거리두기 해제 후		
	단어	빈도수	TF-IDF	단어	빈도수	TF-IDF
1	부산	7,036	3,514.715	부산	8,402	3,291.022
2	오시리아	6,039	2,613.541	관광단지	6,748	2,362.055
3	관광단지	5,686	1,862.605	오시리아	5,574	2,793.550
4	기장	3,732	3,222.498	기장	4,422	3,612.153
5	관광	3,248	2,589.603	관광	3,033	2,953.260
6	동부산	1,795	2,728.755	해운대	1,641	3,246.293

7	테마파크	734	1,653.150	동부산	1,480	2,743.197
8	스타테라스	727	2,041.526	테마파크	1,384	2,320.713
9	부산도시공사	668	1,589.142	카페	1,266	4,299.042
10	일광신도시	655	1,565.630	롯데월드어드벤처	1,127	2,093.813
11	소더비	650	2,189.986	롯데월드	1,069	2,177.002
12	롯데월드	637	1,544.743	반얀트리	1,006	2,921.846
13	시니어타운	612	1,921.690	부산도시공사	658	1,748.930
14	메디타운	572	1,691.422	핵심시설	624	1,404.540
15	상가	485	1,361.953	테마파크개장	581	1,508.713
16	교통	484	1,364.757	라우어	566	1,767.133
17	스위첸마티에	408	1,218.939	대규모	528	1,377.147
18	이케아동부산점	402	1,325.066	롯데월드어드벤처개장	488	1,352.193
19	해운대	385	1,142.340	시니어	476	1,406.764
20	롯데호텔	375	1,331.824	웨딩	450	1,861.054
21	관광지	370	1,070.145	시랑리	438	1,142.406
22	관광단지개장	356	990.586	연합뉴스	402	1,318.885
23	호텔	332	1,009.514	리조트	395	1,097.734
24	르노삼성	323	1,111.940	대변리	385	1,053.426
25	아난티코브	312	989.876	부산시의회	343	1,304.694
26	상가분양	309	954.541	쌍용더플래티넘	339	1,048.094
27	교통대책	308	1,042.418	박형준부산시장	332	1,150.059
28	오규석기장군수	293	1,008.664	랜드마크	329	1,053.625
29	루지	290	924.930	이케아	327	956.932
30	시설	285	875.015	롯데월드개장	324	1,065.927
31	리조트	268	868.688	오션뷰	312	952.021
32	기장역엘크퍼스트	266	952.892	호텔	297	961.256
33	롯데월드어드벤처	258	828.370	교통소통대책	288	1,008.812
34	일광	253	830.396	부지규모	285	862.547
35	이케아	247	791.990	파티	277	1,142.651
36	롯데프리미엄아울렛	243	769.954	맛집	276	962.140
37	롯데월드개장	242	897.509	오시리아역	272	891.152
38	한화건설	237	889.271	샤브썸주머니	272	1,127.813
39	문화예술타운	237	907.511	아난티	260	868.772
40	디지털트윈국토	236	915.291	부산일보	255	956.688
41	부산일보	231	746.713	일광신도시	255	844.813
42	내년연기	229	882.441	시설	246	794.071
43	대우조선해양건설	226	851.370	숙박	246	781.717
44	투자	226	766.053	실버타운	239	802.074
45	스카이라인루지	220	747.995	부동산	234	776.337
46	조감도	215	712.755	동해선	232	784.311
47	(주)코리아소더비국제부동산	210	816.226	상가	230	843.717
48	코로나영향	208	690.547	아파트	229	803.443
49	힐튼호텔	206	696.148	관광객	228	742.923
50	국립부산과학관	205	692.769	착공	227	854.953

위해 UCINET 6와 NetDraw를 이용하여 CONCOR 분석을 수행하였다[1].

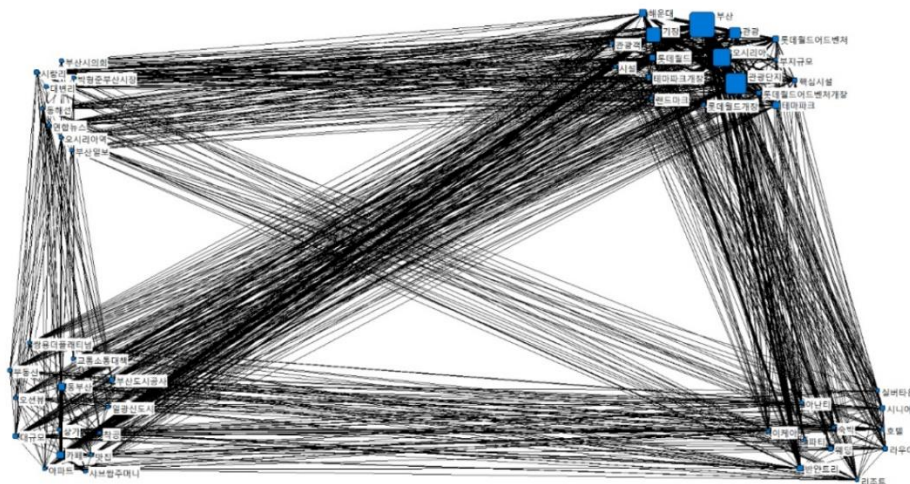
CONCOR 분석은 변수 간의 상관관계를 반복적으로 수행하여 적절한 수준의 유사한 그룹을 찾는 방법으로 본 연구에서는 의미연결망 분석 결과를 바탕으로 수행하였고, 또한 클러스터의 분석은 관련성에 따라 동종 그룹으로 구분하였다.

CONCOR 분석 결과, [그림 3], [그림 4]와 같이 4개의 클러스터가 형성되었음을 볼 수 있으며, 각각의 클러스터를 구성하는 노드는 [표 3]과 같다.



[그림 3] CONCOR 분석 결과: 사회적 거리두기 기간

[Fig. 3] CONCOR Analysis Results: Social Distancing Period



[그림 4] CONCOR 분석 결과: 사회적 거리두기 해제 후

[Fig. 4] CONCOR Analysis Results: After Social Distancing Is Lifted

[표 3] 클러스터별 구성 노드

[Table 3] Configuration Nodes per Cluster

클러스터	COVID-19 사회적 거리두기 기간	COVID-19 사회적 거리두기 해제 후
1	부산, 오시리아, 관광단지, 기장, 관광, 테마파크, 소더비, 롯데월드, 해운대, 관광지,	부산, 관광단지, 오시리아, 기장, 관광, 해운대, 테마파크, 롯데월드어드벤처, 롯데월드, 핵심시설,

	루지, 시설, 롯데월드어드벤처, 롯데월드개장, 문화예술타운, 내년연기, 스카이라인루지, (주)코리아소더비국제부동산, 관광단지개장	테마파크개장, 롯데월드어드벤처개장, 랜드마크, 롯데월드개장, 부지규모, 시설, 관광객
2	동부산, 스타테라스, 부산도시공사, 일광신도시, 상가, 스위첸마티에, 이케아동부산점, 상가분양, 교통대책, 오규석기장군수, 코로나영향, 투자, 조감도	동부산, 카페, 부산도시공사, 대규모, 쌍용더플래티넘, 오션뷰, 교통소통대책, 맛집, 샤브썬주머니, 일광신도시, 부동산, 상가, 아파트, 착공
3	교통, 일광, 디지털트윈국토, 부산일보, 르노삼성	시랑리, 연합뉴스, 대변리, 부산시의회, 박형준부산시장, 오시리아역, 부산일보, 동해선,
4	호텔, 아난티코브, 리조트, 기장엘크퍼스트, 이케아, 롯데프리미엄아울렛, 이케아동부산점, 대우조선해양건설, 힐튼호텔, 국립부산과학관, 메디타운, 시니어타운, 롯데호텔, 한화건설	반얀트리, 라우어, 시니어, 웨딩, 리조트, 이케아, 호텔, 파티, 아난티, 숙박, 실버타운

4. 결론

본 연구는 COVID-19 사회적 거리두기 해제 전·후로 구분하여 국내 포털사이트인 네이버, 다음에 게시된 자료를 대상으로 ‘오시리아’ + ‘관광단지’ 관련 텍스트 데이터를 분석하여 관광 소비자들이 이를 어떻게 인식하는지 비교한 연구이다. 자료의 수집과 정제를 통해 자주 등장하는 단어들을 추출하였고, 추출한 단어 간의 상관관계를 이용하여 네트워크 분석과 CONCOR 분석을 수행하여 다음과 같은 결론을 도출하였다.

첫째, 빈도분석 결과를 통해 ‘오시리아’ + ‘관광단지’에 대한 COVID-19 사회적 거리두기 기간에는 관광단지의 핵심 시설인 롯데월드의 개장 연기와 관련된 단어(테마파크, 롯데월드어드벤처, 롯데월드개장연기, 내년연기, 코로나영향), 건설이 진행 중인 관광 시설(소더비, 롯데월드, 시니어타운, 메디타운, 롯데호텔, 문화예술타운)에 관심이 많았고, 상가, 오피스텔, 아파트 등 부동산 관련 단어(스타테라스, 일광신도시, 상가, 스위첸마티에, 상가분양, 기장역엘크퍼스트, 한화건설, 대우조선해양건설, 투자, 조감도), 개장되어 운영 중인 시설에 관련된 단어(이케아동부산점, 아난티코브, 루지, 이케아, 힐튼호텔, 국립부산과학관)들이 등장하였다. 경제 활동의 위축과 사람의 이동마저 불편했던 시기인 점을 감안하면 관광의 관점에서 볼 때 개장하여 운영 중인 시설과 상가 시설에 입점한 식당, 카페 등의 단어들이 상위 50위 안에서는 볼 수 없었다. 이에 반해 COVID-19 사회적 거리두기 해제 후에는 롯데월드의 개장이라는 이벤트도 있었고, 카페, 웨딩, 오션뷰, 파티, 맛집, 샤브썬주머니, 숙박 등의 새로운 단어들이 등장하여 관광이 활성화된 것으로 보였다.

둘째, COVID-19 사회적 거리두기 해제 후의 TF-IDF 분석 결과를 통해 빈도분석 에서 9위(카페)가 TF-IDF 분석에서는 1위에 올랐다. 이외에도 4위(기장), 5위(관광), 6위(동부산), 12위(반얀트리), 15위(테마파크개장), 16위(라우어), 19위(시니어), 20위(웨딩), 33위(교통소통대책), 35위(파티), 38위(샤브썬주머니) 등이 빈도 분석 순위 보다 문서상의 중요성은 높은 것으로 나타났다. 이는 COVID-19 사회적 거리두기 해제 후에는 관광과 관련된 키워드가 검색된 모습을 볼 수 있다.

셋째, CONCOR 분석을 통해 유사도가 높은 그룹을 분류한 결과 4개의 클러스터로 분류하였다. 각 클러스터를 살펴보면, 1번 클러스터는 관광단지, 테마파크 시설과 관련된 내용으로 구성되었고, 2번 클러스터는 카페, 맛집, 상업 시설, 주거 시설 등과 주말에 증가하는 교통난에 대한 내용이, 3번 클러스터는 관광단지의 인프라에 관한 것으로 철도,

도시철도, 부산시의 교통 정책 등이, 4번 클러스터는 호텔, 리조트, 쇼핑 시설 등이 포함되어 있다. 2022년 3월에 개장한 롯데월드로 인하여 압도적으로 많은 데이터를 볼 수 있었다. 특히, 2번과 3번 클러스터는 관광단지가 숙박·레저·쇼핑 시설 등이 포함된 사계절 체류형으로 개발된 이유로 관광단지 내부와 인근에 많은 주거 시설들이 조성되어 주거 인구와 주말에 폭발적으로 증가하는 유동 인구에 의해 발생하는 교통 정체를 해결해야 하는 과제를 안고 있다.

본 연구는 COVID-19 사회적 거리두기 해제 전·후로 '오시리아' + '관광단지'의 핵심 키워드들을 추출하여 이를 빅데이터 단어들로 정리한 것이다. 그 결과, COVID-19 사회적 거리두기 해제 후에는 롯데월드의 개장, 카페, 웨딩, 오션뷰, 파티, 맛집, 샵스잡주머니, 숙박 등의 새로운 단어들도 등장하여 거리두기 해제 전보다 관광 활동이 더욱 활성화된 것으로 확인되어 관광 소비자들의 인식변화를 시사하고 있다.

본 연구의 한계점으로 '오시리아' + '관광단지'의 키워드 분석에서 빈도가 높았던 상위 50위 안에 드는 단어들에 대해 관련성을 설명할 수 있었지만 이외의 단어들에 대한 관계를 설명할 수 없었다. 또한, SNS 데이터를 바탕으로 분석하였기 때문에 참여도가 낮은 청소년, 노령층의 인식을 파악할 수 없어서 모든 관광 소비자의 인식이라고 할 수 없었고, 텍스트만으로 수집한 원본 자료를 검토한 결과 채널별로 수집할 수 있는 최대 수가 제한되어 있어서 대상 기간의 앞쪽 자료가 많이 반영 되었음을 알 수 있었다. 따라서 후속 연구에서는 데이터를 분기별 또는 월별로 세분하여 분석하는 것이 필요하고, 연령별로도 비교하여 분석한다면 향후, 관광산업의 정책방향 결정에 더 많은 기여할 수 있을 것으로 생각된다.

References

- [1] Y. H. Hong, S. M. Shin, N. W. Kim, J. H. Kim, H. K. Kim, A Study on the Perception Based on Big Data Analysis Related to Osiria Tourism Complex in Busan, *Asia-pacific Journal or Convergent Research Interchange*, (2022), Vol.8, No.3, pp.59-68.
DOI: <http://dx.doi.org/10.47116/apjcri.2022.03.06>
- [2] <https://namu.wiki/w/%EC%82%AC%ED%9A%8C%EC%A0%81%20EA%B1%B0%EB%A6%AC%EB%91%90%EA%B8%B0>, Nov 2 (2022)
- [3] D. H. Park, A comparative analysis of public perception and tourist needs of Andong before and after of COVID-19 outbreak: Text mining and semantic network analysis using big data on social media, *Korean Journal of Hospitality and Tourism*, (2021), Vol.30, No.5, pp.231-246.
DOI: <https://doi.org/10.24992/KJHT.2021.7.30.05.231>
- [4] H. G. Song, A study on food tourism perception using big data: comparison before and after the outbreak of COVID-19, *Journal of Foodservice Management*, (2021), Vol.24, No.5, pp.177-200.
DOI: <https://doi.org/10.47584/jfm.2021.24.5.177>
- [5] J. H. Kim, H. R. Lee, Comparative Study on the Changes in Perceptions of Camping in Tourists before and after the COVID-19 Outbreak Using Social media Big Data: Focused on Semantic Network Analysis, *Korean Journal of Hospitality and Tourism*, (2021), Vol.30, No.5, pp.117-134.
DOI: <https://doi.org/10.24992/KJHT.2021.7.30.05.117>
- [6] K. S. Woo, Time Series Analysis of Urban Parks Usage Patterns Using Big Data, Kyunghee University, Doctoral Dissertation, (2020)
- [7] M. H. Ko, J. E. Lee, Text mining related to Gimhae Airport based on big data: Comparing before and after COVID-19, *International Journal of Tourism and Hospitality Research*, (2022), Vol.36, No.6, pp.139-153.
DOI: <https://doi.org/10.21298/IJTHR.2022.6.36.6.139>