Tourism Management and Technology Economy (2021) 4: 14-20
Clausius Scientific Press, Canada
DOI: 10.23977/tmte.2021.040203
ISSN 2616-2199

A study on the recognition of Korea National Park by using social network analysis based on big data

Wenhui Shan

Hotel & Tourism, Honam University, Gwangju, 62399, Republic of Korea
798840478@qq.com

Keywords: Big data, Korea National Park, Social network analysis, Text mining

Abstract: The purpose of this studies to generate the tourist's perception of National Park in Korea by using social network analysis based on big data. Data were gathered from Naver, Daum, and Google as analysis of data channels. The search period was limited to three years from Jan.1, 2018, to Dec.31, 2020. a total of 30,183 text data were collected by using the Textom program, and keywords analysis and frequency analysis of the data collected through text mining was performed to identify Korea National park. Density analysis, network concentration analysis, centrality analysis, and CONCOR analysis were performed using the Ucinet 6 program. As a result, first, various keywords related to Korea National Park such as 'National Park Service, Ministry of Environment, hiking, trail, travel, sunrise, picture, autumn, sunset, beautiful' were extracted. Second, as a result of CONCOR analysis, finally, 4 groups were formed. Confidently, this study discussed the implications of a marketing strategy for Korea's National Parks as a result of the semantic network analysis study. Therefore, the activation of National Park, which is a typical tourism form of sustainable tourism, is expected to be an alternative to the tourism industry in Korea.

1. Introduction

Since the 21st century, with the accelerated development of the global tourism industry, the over-exploitation of tourism resources, the destruction of the ecological environment, and global climate problems have become more and more serious. Therefore, the protection of tourism resources, the restoration of the ecological environment, and the improvement of the climate environment have promoted tourism Sustained development is an important goal of today's global tourism development. However, National park is to protect nature and to offer people opportunities to experience nature (Böhn, 2021). Yellowstone National Park is the first protected area to have been designated a national park (MacKintosh, 1985). Yellowstone was established in 1872 in the United States, and has served as a framework ultimately adopted by countries across the globe. In Korea, since Jirisan National Park was first designated in 1967, there are currently 22 national parks, occupying 4.0% of the total land area. In national parks, 45% of domestic recorded species and 65% of endangered species alone are inhabited and distributed. In addition, 41 national treasures such as scenic spots and temples and 733 designated cultural properties are also located in the national park (Korea National Park Service, 2020).

National park not only plays a key role to conserve a country's natural heritage (Gissibl, Höhler, & Kupper, 2012), but also has been spotlighted as Korea's representative tourist destinations as a place for citizens' leisure and health care, and visiting national park has been established as a general recreational culture(Lee, 2010). Therefore, National park has also became a research hotspot for scholars. However, The research methods on National park in the tourism field (Agyeman, Aboagye & Ashie, 2019; Ferretti-Gallon et al., 2021; Böhn, 2021; Souza et al., 2021) are mainly focuses on conduct through surveys, policy and statistical data review. With the development of information and communication technology, tourism information services are provided mainly through the Internet, and information related to the tourist experience is also shared in real time. In addition, smart devices and PCs facilitate information search and influence tourists' decision-making (Gursoy & McCleary, 2004). So the advent of big data, it can be said that conditions have been prepared to supplement the limitations of existing survey statistics or tourism statistics centered on industry statistics. what's more, it can be used as a cost-efficient way to explore, and for more continuous tourists' monitoring, preferences and activities in protected areas of National park(Hausmann at al., 2018).

The purpose of this research is to obtain text data by searching the keywords of Korea National park on Naver, Daum and Google which are most commonly used by Koreans. From extracting text data to final data visualization, what are the keywords related to National park, what are the factors that have a greater impact on National park, and how are the clusters of National park constituted? So as to provide basic data for National park development planning and policy direction.

2. Research method

2.1 Data collection and analysis methods

This study set the time interval from January 1, 2018 to December 31, 2020, and search for keywords related to Korea National park on the search engines Naver, Daum, and Google, which are highly used by Koreans, to collect Korea National park related Text data. A total of 30,183 text data were collected by using the Textom program, an online collection agency. On the basis of the collected data, the text is refined, and the frequency of the refined data is calculated through text mining to select keywords. In addition, the keywords text data is converted into matrix data for social network analysis. Density analysis, network concentration analysis, centrality analysis, and CONCOR analysis were performed using the Ucinet 6 program. The data analysis method and content are shown in Table 1.

Table 1 Data analysis method and content

Analysis method		Date analysis method	Tool
Text mining	Keyword analysis	Extract text related to Korea National Park.	Textom
	Frequency analysis	Frequency of simultaneous occurrence of words related to Korea National Park.	
Social network analysis	Density analysis	Analysis of the connection structure and characteristics of words related to Korea National Park.	Ucine 6 & Netdraw
	Network concentration analysis	Analysis of the concentration of specific words in the network structure of Korea National Park.	
	Centrality analysis	Analysis the high influence words on Korea National Park.	
	CONCOR analysis	Analysis of similar clusters of words in the network structure of Korea National Park.	

2.2 Research question

The purpose of this research is to use big data, text mining and social network analysis to determine the key factors that affect the perception of National park in Korea, and to explore the relationship between the influencing factors. In addition, through research and analysis, it is possible to improve the awareness of Korea National park, revitalize National park, and provide countermeasures for the development of National park in Korea. Therefore, according to the purpose of this research, the research tasks are as follows.

Research question 1: What are the keywords related to Korea National park ?

Research question 2: What is the structure and characteristics of the network of relevant keywords in Korea National park ?

Research question 3: What is the significance of the connection, intermediary, proximity centrality, and clustering of keywords related to Korea National park?

3. Result

3.1 Keywords frequency of Korea National Park

Table 2 shows the frequency analysis results of the top 50 main keywords related to National park in Korean.‘National Park Service’(889) occupy the first place, followed by ‘Ministry of Environment(741), Hiking(687), Travel(682), Designation(648), Seoraksan National Park(613), Jirisan National Park(527), Event(493), Bukhansan National Park(488), Trail(468), Sunrise(465) Operate(432), Time(430), Area(422), Mudeungsan National Park(397)’ etc.

Table 2 The results of keywords frequency of Korea National Park

code	word	freq	Ratio(%)	code	word	freq	Ratio (%)
1	National Park Service	889	0.510	26	Reservation	262	0.150
2	Ministry of Environment	741	0.425	27	Change	260	0.149
3	Hiking	687	0.394	28	Gyeryongsan National Park	258	0.148
4	Travel	682	0.392	29	Campsite	256	0.147
5	Desdesignation	648	0.372	30	Juwangsan National Park	254	0.146
6	Seoraksan National Park	613	0.352	31	Nature	244	0.140
7	Jirisan National Park	527	0.303	32	Sunset	238	0.137
8	Event	493	0.283	33	Hallyeohaesang National Park	232	0.133
9	Bukhansan National Park	488	0.280	34	Location	213	0.122
10	Trail	468	0.269	35	Maple	212	0.122
11	Sunrise	465	0.267	36	Beautiful	195	0.112
12	Operate	432	0.248	37	Lodgment	192	0.110
13	Time	430	0.247	38	Gayasan National Park	189	0.109
14	Area	422	0.242	39	Deogyusan National Park	185	0.106
15	Mudeungsan National Park	397	0.228	40	Autumn	185	0.106
16	Hallasan National Park	362	0.208	41	Year-end and New Year holidays	184	0.106
17	Picture	354	0.203	42	Restrict	184	0.106
18	Naejangsan National Park	349	0.200	43	Corona	184	0.106
19	Prohibition	345	0.198	44	Car park	180	0.103
20	Control	326	0.187	45	Closing	172	0.099
21	Course	324	0.186	46	Chiaksan National Park	163	0.094
22	Visitor	315	0.181	47	Odaesan National Park	162	0.093
23	Management	291	0.167	48	Valley	157	0.090
24	Citizen	274	0.157	49	Packed lunch	155	0.089
25	Songnisan National Park	271	0.155	50	Wild animals	154	0.088

Through density analysis and overall network concentration analysis, understand the characteristics of the network structure of keywords related to Korea National park. The visualization results are shown in Figure 1. According to the research results, the network density is 0.519, and the overall network concentration is 33.08%. It can be seen that all keywords are interconnected In addition, it is mainly centered on keywords such as ‘National Park Service’, ‘Jirisan National Park’, ‘Seoraksan National Park’, ‘Hiking’ and ‘Time’ etc.

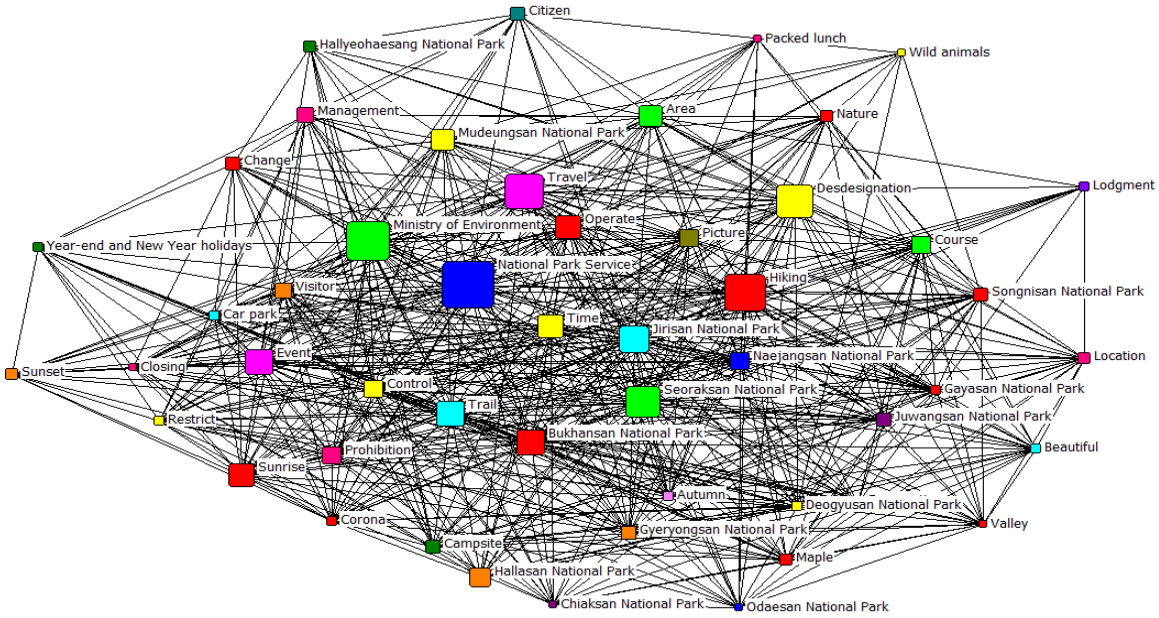

Figure 1 The network analysis result of Korea National Park

3.2 The analysis of keywords centrality with Korea National Park

In order to understand the influence of keywords related to Korea National park, this study conducted an analysis of degree centrality, closeness centrality and betweenness centrality. The analysis results are shown in Table 3.

Degree centrality refers to the number of connections between one node and another node in the structure. The higher the connectivity centrality, the greater the impact on the network structure. From the results of the degree centrality analysis, it can be seen that ‘National Park Service, Jirisan National Park, Seoraksan National Park, Hiking, Trail, Time, Ministry of Environment’ etc show high centrality. Closeness centrality means one node has the shortest path distance from all other nodes in the network. As a result of closeness centrality analysis, ‘National Park Service, Jirisan National Park, Seoraksan National Park, Hiking, Travel, Time’ etc show the highest closeness centrality. Betweenness centrality is a concept that describes the degree of controlling or mediating relationships between nodes that are not directly connected in a network, and Betweenness centrality can better capture leaders or influential members within the network structure. As a result of the analysis of betweenness centrality, ‘National Park Service, Jirisan National Park, Hiking, Time, Travel’ were found to have high betweenness centrality in the order.

Table 3 The analysis of keywords centrality with Korea National Park

word	Deg.	Clo.	Bet.	word	Deg.	Clo.	Bet.
National Park Service	0.837	0.860	3.444	Reservation	0.449	0.636	0.540
Ministry of Environment	0.735	0.790	2.248	Change	0.429	0.681	0.619
Hiking	0.755	0.803	2.764	Gyeryongsan National Park	0.531	0.671	0.650
Travel	0.653	0.742	2.253	Campsite	0.510	0.653	0.938
Desdesignation	0.592	0.831	1.386	Juwangsan National Park	0.469	0.613	0.577
Seoraksan National Park	0.796	0.860	2.189	Nature	0.367	0.598	0.300
Jirisan National Park	0.837	0.731	3.232	Sunset	0.327	0.507	0.042
Event	0.633	0.766	1.455	Hallyeohaesang National Park	0.245	0.620	0.157
Bukhansan National Park	0.694	0.803	1.473	Location	0.388	0.671	0.461
Trail	0.755	0.700	1.793	Maple	0.510	0.620	0.649
Sunrise	0.571	0.742	0.779	Beautiful	0.388	0.557	0.383
Operate	0.653	0.803	1.526	Lodgment	0.204	0.662	0.043
Time	0.755	0.700	2.368	Gayasan National Park	0.490	0.662	0.513
Area	0.571	0.671	1.808	Deogyusan National Park	0.490	0.662	0.734
Mudeungsan National Park	0.510	0.628	0.895	Autumn	0.490	0.598	0.428
Hallasan National Park	0.408	0.731	0.287	Year-end and New Year holidays	0.327	0.671	0.113
Picture	0.633	0.731	1.609	Restrict	0.510	0.690	0.505
Naejangsan National Park	0.633	0.721	1.275	Corona	0.551	0.710	0.694
Prohibition	0.612	0.778	0.912	Car park	0.592	0.662	0.817
Control	0.714	0.671	1.973	Closing	0.490	0.628	0.423
Course	0.510	0.731	1.213	Chiaksan National Park	0.408	0.613	0.356
Visitor	0.633	0.620	0.971	Odaesan National Park	0.367	0.620	0.261
Management	0.388	0.576	0.419	Valley	0.388	0.563	0.275
Citizen	0.265	0.662	0.211	Packed lunch	0.224	0.551	0.062
Songnisan National Park	0.490	0.645	0.984	Wild animals	0.184	0.636	0.076

3.3 Cluster analysis of related keywords with Korea National Park

The visualization results of the CONCOR analysis are shown in Figure 2. The analysis results show that it is mainly composed of four groups. The largest group is mainly composed of words related to the management and operation information of Korea National park such as ‘National Park Service, Restrict, Event, Closing, Operate, Control, Prohibition’ etc. Group 2 is mainly composed of some national parks such as ‘Gyeryongsan National Park, Gayasan National Park, Juwangsan National Park, Seoraksan National Park, Jirisan National Park’ etc. Group 3 is mainly formed by some words like ‘Travel, Designation, Location, Lodgment, Reservation’ etc. Group 4 is about ‘Nature, Wild animals, Area, Citizen, Picture’ etc.

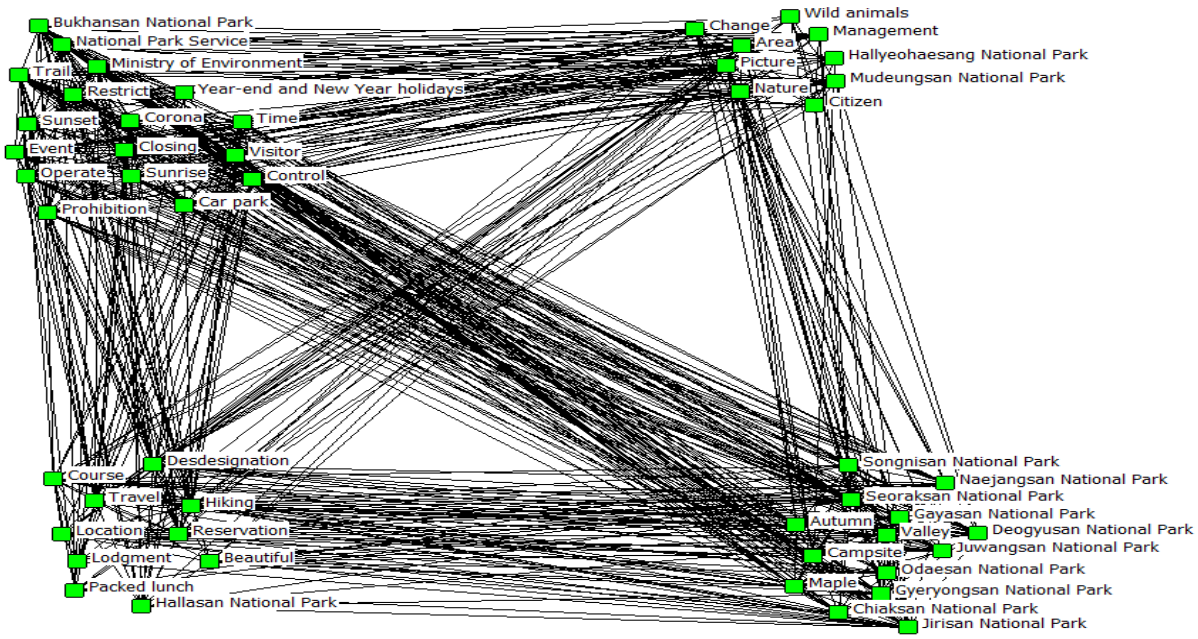

Figure 2. Cluster analysis visualization result

4. Conclusion

The results of this study are summarized as follows. First, as a result of frequency analysis, the keywords ‘National Park Service, Ministry of the Environment, Hiking, Travel, Designation, Seoraksan National Park, Jirisan National Park, Event, Bukhansan National Park, Trails, Sunrise’ have a high proportion. The awareness of Korea National park on the Internet is mainly achieved by public management agencies, designated national parks and natural landscapes. Secondly, in the central analysis results of keywords related to Korea National Park, keywords such as ‘National Park Service, Jirisan National Park, Seoraksan National Park, Hiking, Trail and Time’ showed that the centrality of degree, closeness and betweenness was confirmed to be high. In particular, when visitors thought of national parks, Jirisan National Park and Seoraksan National Park appeared a lot. In addition, the purpose of visiting the national park, it is mainly hiking. Therefore, visitors are also very concerned about the climbing time and route. Third, a total of 4 similar groups were formed through CONCOR analysis. One of the two large groups is mainly based on management and operation information related to Korea’s national park, and the other is mainly based on designated national parks.

The academic and practical significance of this research is as follows. First, from previous research, the research on national park of Korea is mainly based on environmental awareness, national park experience programs and resource development. However, this study can directly or indirectly investigate visitors' perceptions, preferences, and usage patterns of national parks through big data.

Second, by analyzing the frequency of keywords in national park of Korea, it is found that keywords such as ‘National Park Service, Designation, Seoraksan National Park, Jirisan National Park and Bukhansan National Park’ appear highly frequent. Visitors are more interested in national parks designated by the country. In addition, words such as ‘Hiking, Event, Trail and Sunrise’ appear frequently. Therefore, in order to improve the activation of national park, it is necessary to strengthen park infrastructure management, accurately and timely announce visiting information,

actively organize special activities related to national park, and develop and promote experience programs.

Third, according to the centrality analysis results, keywords such as ‘Jirisan National Park, Seoraksan National Park, Hiking, Trail and Time’ occupy high centrality. Jirisan National Park and Seoraksan National Park are the most representative national park in Korea. In particular, Jirisan National Park, which is the first designated national park in Korea, has a leading and exemplary role. Therefore, the management of public institutions should pay more attention to such national park, such as the management of park opening time, climbing route planning, and general facilities.

Fourth, according to the results of CONCOR analysis, the largest group is mainly based on the National Park Service, the Ministry of Environment, events and national park management information. It can be seen that local government and national park activities and information disclosure are recognized and discussed by visitors online. Therefore, in order to revitalize national park in the future, development of various events or programs for each period should not only promote through major media, but also establish and execute a viral marketing plan through advertisements on major portal sites or SNS.

The limitations of this research and future tasks are as follows. First, the data about Korea National Park collection channels for this study were limited to Naver, Daum and Google. Therefore, the future research uses multiple channels to collect data, such as travel websites and SNS data. Secondly, this research only collected text data, so future research can use such as video or audio data. Third, this study selects the keywords through frequency analysis, so it is difficult to determine whether the meaning of the word is a positive or negative influence.

References

[1] Agyeman, Y. B., Aboagye, O. K., & Ashie, E. (2019). Visitor satisfaction at kakum national park in Ghana. Tourism Recreation Research, 44(2), 178-189.

[2] Böhn, D. (2021). National Parks in Germany: Let nature be nature–But which nature?. International Journal of Geoheritage and Parks, 9(1), 30-35.

[3] Ferretti-Gallon, K., Griggs, E., Shrestha, A., & Wang, G. (2021). National parks best practices: Lessons from a century's worth of national parks management. International Journal of Geoheritage and Parks.

[4] Gissibl, B.,Höhler, S., & Kupper, P. (2012). Civilizing nature: National parks in global historical perspective. New York, NY: Berghahn Books.

[5] Gursoy, D. & McCleary, K. W. (2004). “An integrative model of tourists' information search behavior”, Annals of tourism research, 31(2), 353-373.

[6] Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., & Di Minin, E.(2018). Social media data can be used to understand tourists’ preferences for nature-based experiences in protected areas. Conservation Letters, 11(1), e12343.

[7] Korea National Park Service. (2020). 2020 National Park Basic Statistics.

[8] Lee S. W. (2010). A Study on the Relation between Environmental Awareness and Behavior of National Park Tourists: With the Emphasis on Chiaksan National Park. Published masteral dissertation. Sangji University. Republic of Korea.

[9] MacKintosh, B. (1985). The National Parks: Shaping the System. Washington, D.C.: The National Parks Service.

[10] Souza, C. N., Rodrigues, A. C., Correia, R. A., Normande, I. C., Costa, H. C., Guedes-Santos, J., & Ladle, R. J. (2021). No visit, no interest: How COVID-19 has affected public interest in world's national parks. Biological Conservation, 256, 109015.