

블로그 데이터 감성분석을 통한 북한산둘레길 구간별 선호도 평가

이성희* · 손용훈** ***

*서울대학교 협동과정 조경학 박사수료 · **서울대학교 환경대학원 환경조경학과 부교수 ·
***서울대학교 환경계획연구소 겸무연구원

Evaluation of Preference by Bukhansan Dulegil Course Using Sentiment Analysis of Blog Data

Lee, Sung-Hee* · Son, Yong-Hoon** ***

*Ph.D. Candidate, Interdisciplinary Program in Landscape Architecture, Seoul National University

**Associate Professor, Graduate School of Environment Studies, Seoul National University

***Adjunct Researcher, Environmental Planning Institute, Seoul National University

ABSTRACT

This study aimed to evaluate preferences of Bukhansan dulegil using sentiment analysis, a natural language processing technique, to derive preferred and non-preferred factors. Therefore, we collected blog articles written in 2019 and produced sentimental scores by the derivation of positive and negative words in the texts for 21 dulegil courses. Then, content analysis was conducted to determine which factors led visitors to prefer or dislike each course.

In blogs written about Bukhansan dulegil, positive words appeared in approximately 73% of the content, and the percentage of positive documents was significantly higher than that of negative documents for each course. Through this, it can be seen that visitors generally had positive sentiments toward Bukhansan dulegil. Nevertheless, according to the sentiment score analysis, all 21 dulegil courses belonged to both the preferred and non-preferred courses.

Among courses, visitors preferred less difficult courses, in which they could walk without a burden, and in which various landscape elements (visual, auditory, olfactory, etc.) were harmonious yet distinct. Furthermore, they preferred courses with various landscapes and landscape sequences. Additionally, visitors appreciated the presence of viewpoints, such as observation decks, as a significant factor and preferred courses with excellent accessibility and information provisions, such as information boards. Conversely, the dissatisfaction with the dulegil courses was due to noise caused by adjacent roads, excessive urban areas, and the inequality or difficulty of the course which was primarily attributed to insufficient information on the landscape or section of the course.

The results of this study can serve not only serve as a guide in national parks but also in the management of nearby forest green areas to formulate a plan to repair and improve dulegil. Further, the sentiment analysis used in this study is meaningful in that it can continuously monitor actual users' responses towards natural areas. However, since it was evaluated based on a predefined sentiment dictionary, continuous updates are needed. Additionally, since there is a tendency to share positive content rather than negative views due to the nature of social media, it is necessary to compare and

Corresponding author: Yong-Hoon Son, Associate Professor, Dept. of Environment Landscape Architecture, Graduate School of Environmental Studies, Seoul National University, Seoul 08826, Korea, Tel: +82-2-880-8107, E-mail: sonyh@snu.ac.kr

review the results of analysis, such as with on-site surveys.

Key Words: Social Media, Text Mining, Preference Analysis, National Parks, Trail Management

국문초록

본 연구는 탐방객이 자유롭게 서술한 블로그 텍스트 데이터를 자연어 처리 기술 중 하나인 감성분석을 활용하여 북한산둘레길의 선호도를 평가하고, 선호 요인과 비선호 요인을 도출하는 것을 목적으로 하였다. 이에 2019년 1년 동안 작성된 블로그를 수집하고 21개 둘레길 구간별 텍스트에 나타난 긍정 및 부정 감성 단어 도출을 통해 감성점수를 산출하였다. 이후 내용분석을 통해 탐방객이 어떤 요소로 인해 구간을 선호하거나 선호하지 않는지 파악하였다.

북한산둘레길에 대해 작성된 블로그에서는 긍정적인 단어가 평균적으로 약 73% 출현하고 있었고, 각 구간별 게시물의 감성 극성 비율에서도 긍정적인 문서의 비율이 부정적인 문서의 비율보다 높았다. 이를 통해 탐방객은 북한산둘레길에 대하여 대체로 긍정적으로 인식하고 있는 것으로 나타났다. 그럼에도 감성점수를 도출한 결과, 21개 둘레길 구간에서는 선호하는 구간과 선호하지 않는 구간이 존재하고 있었다.

선호 구간과 비선호 구간에 대해 탐방객은 난이도가 낮고 부담 없이 걸을 수 있는 구간을 선호하고 있었고, 경관에 대한 여러 요소(시각, 청각, 후각 등)가 조화롭고 계절감이 뚜렷해 다양한 경관이 연출되는 곳, 경관 시퀀스의 변화가 존재하는 구간을 선호하는 것으로 나타났다. 또한 탐방객은 전망대, 조망점 등의 뷰포인트 유무를 둘레길에서의 주요 요소로 인식하고 있었고, 접근성이 우수하고 안내판 등 정보 제공이 원활하게 이뤄지는 구간에 대해 선호도가 더 높은 것을 알 수 있다. 반면, 도로와 인접함에 따라 발생하는 주변 소음과 과도한 시가지 비율, 구간별 난이도 불균형 등으로 인한 둘레길 동선 불만족이 비선호 요인으로 크게 작용하고 있었으며, 경관 단절 및 구간에 대한 정보 부족 등이 선호도를 떨어트리는 원인으로 나타났다.

본 연구의 결과는 국립공원뿐만 아니라 근교 산림 녹지 관리에 있어서 둘레길 정비 및 개선방안 마련에 활용될 수 있으며, 연구에 활용된 감성분석은 자연지역에 대한 실제 이용자들의 반응을 지속적으로 모니터링 할 수 있다는 점에 의의가 있다. 다만 사전에 정의된 감성사전을 기반으로 평가하였기에 지속적인 사전 업데이트가 필요하다. 또한 소셜미디어 특성상 부정적인 견해보다는 긍정적인 내용을 공유하는 경향이 존재하기 때문에, 현장 설문조사 등의 분석 결과와 비교, 검토하는 작업이 필요하다.

주제어: 소셜미디어, 텍스트마이닝, 선호도 분석, 국립공원, 트레일 관리

1. 서론

도시환경에서 녹지 공간은 스트레스 감소와 건강 및 평온함을 제공함으로써 다방면으로 인간의 행복과 삶의 질을 높이는 역할을 수행하고 있다(Chiesura, 2004; Walmsley, 2006; Landers and Nahlik, 2013). 도시에 인접한 산림은 접근성이 뛰어나 일상 생활에서도 쉽게 방문 가능하고, 레크리에이션을 제공하는 공간으로 생태계서비스 중 문화서비스 기능이 높은 녹지이다(Larson *et al.*, 2016).

이러한 산림을 향유하는 다양한 행태 중 하나로 조성된 둘레길은 생태적으로 중요한 역할을 담당하고(Bryant, 2006; Aherin, 2013) 사람들에게 다양한 여가서비스를 제공하는 복합적 기능을 수행하고 있다(Shafer *et al.*, 2000; Fitzhugh *et al.*, 2010; Dallat *et al.*, 2014). 최근 여행 뉴노멀(new normal)로 '근거

리·안전·숲속·청정·생태·트레킹'이 언급되고, 코로나(COVID-19)로 인해 언택트(untact) 관광 수요가 늘어남에 따라 많은 사람을 만나기보다는 자연 속에서 가볍게 트레킹 할 수 있는 둘레길에 대한 수요가 증가하고 있다(<http://news.heraldcorp.com/view.php?ud=20200617000670>, 2020.06.17.).

일반적으로 둘레길 같은 트레일 조성은 도시계획가, 생태·조경가 및 지자체 등 전문가에 의해 계획되고 설계되어 왔으나(Plieninger *et al.*, 2013; Kabisch *et al.*, 2015), 이를 이용하는 사람은 대중, 즉 비전문가이고 그로 인해 전문가의 의도와 실제 사용자의 인식간의 잠재적 불일치가 발생하는 문제점이 나타나기도 한다(Plieninger *et al.*, 2015; Larson *et al.*, 2016; Daniels *et al.*, 2018).

지속가능한 공간 계획 및 관리방안 마련을 위해서는 실제로 공간을 이용하는 이용자의 인식과 선호도를 파악할 필요가 있

다(Bell *et al.*, 2007). 이용자의 인식을 파악하는 방법 중 하나로 소셜미디어 데이터를 활용할 수 있다. 소셜미디어 데이터는 이용자의 자발적인 참여로 만들어지며, 이는 토지이용계획 및 경관 관리에 활용 가능하다(Seeger, 2008; Martínez Pastur *et al.*, 2015). 또한 소셜미디어 데이터는 문화서비스 및 경관 가치 분석 등 여러 측면에서 대중의 인식을 파악할 수 있는 중요한 자료로서 입증되었으며, 현재까지 지오태그(geo-tagged) 데이터, 사진 및 텍스트, 해시태그 등 다양한 소셜미디어 데이터를 활용한 연구들이 수행되고 있다(Casalegno *et al.*, 2013; Dunkel, 2015; Tenerelli *et al.*, 2016; van Zanten *et al.*, 2016).

소셜미디어 데이터 중 블로그는 자유롭게 작성된 방대한 양의 텍스트를 포함하고 있어, 글 속에 담긴 이용자의 인식을 파악하는 데 효과적이다(Lee and Son, 2018). 오픈서베이가 10~50대 남녀 620명을 대상으로 조사한 '소셜미디어와 검색포털 트렌드 리포트 2020'에 따르면, 포털사이트 이용에 있어 '네이버(Naver)'가 전 연령층에서 높은 이용률을 보이고, '다음(Daum)'은 주 이용 연령층 비중이 50대에서 상대적으로 높은 것으로 나타났다(OPENSURVEY, 2020). 이는 소셜미디어 데이터의 생산 주체가 2030세대에 국한되어 있을 것이란 견해에 반하는 분석 결과로 볼 수 있다. 따라서 블로그 데이터를 활용한다면 다양한 연령층의 의견을 파악할 수 있다고 할 수 있다.

감성분석은 블로그 텍스트를 분석하는 방법의 하나이다. 감성 분석은 오피니언 마이닝(opinion mining)의 한 응용분야로, 텍스트에 존재하는 감성을 수치화하여 감성 정도를 도출하는 방법이며, 국내외 관광, 마케팅 및 심리학 등 여러 연구 분야에서 활용되고 있다(Liu, 2012; Nam, 2015; Mäntylä *et al.*, 2018). 아직 활용도가 낮지만 조경분야에서 감성분석을 활용한다면 녹지, 공원 또는 공간에 대한 방문객의 선호도 및 견해 파악 등에 활용도가 높을 것으로 기대된다.

이에 본 연구에서는 블로그 텍스트 데이터와 감성분석을 활용하여 공간에 대한 실제 수요자인 방문객의 선호도를 평가하고자 하였다. 또한, 평가된 공간에 대한 선호·비선호의 이유를 해석하고, 주요 특성을 도출하였다.

II. 연구대상 및 방법

1. 연구 대상지 선정 및 개요

본 연구는 다양한 성격과 자원을 보유하고, 탐방객의 방문수요가 높은 북한산둘레길을 대상으로, 블로그 글을 활용한 감성 분석을 통해서 선호도를 평가하였다.

북한산 국립공원은 '단위 면적당 가장 많은 탐방객이 찾는 국립공원'으로 기네스북에 등재되어 있다. 연구의 공간적 범위인 북한산둘레길은 북한산국립공원에 조성된 둘레길이다. 북한

산국립공원의 탐방객 수는 2019년 기준 약 5,574,000명이다. 이는 한려해상국립공원을 제외하고 내륙에 위치한 국립공원 중 가장 많은 탐방객이 방문하는 곳으로, 생태, 관광, 휴양자원으로서 가치가 높은 곳이다(KNPS, 2020).

북한산둘레길은 총 21개 구간으로 구성되어 있으며, 1구간-12구간은 북한산에, 13구간-20구간은 도봉산에 포함되어 있다. 마지막으로 21구간은 북한산과 도봉산 경계부인 우이령 고개에 위치하고 있다. 각 구간 별로 약간의 차이가 존재하지만 구간의 평균길이는 약 3.4km이며, 일부 구간은 서울둘레길, 은평둘레길, 고양누리길 등 지자체에서 조성한 둘레길과 중복되어 있다.

2. 연구 방법

1) 데이터 수집 및 정제

(1) 데이터 수집

본 연구에서는 북한산둘레길 방문객들의 선호도를 파악하기 위하여 블로그 게시물을 분석 대상으로 설정하였으며, 네이버(Naver)와 다음(Daum)에서 작성된 게시물을 수집하였다.

검색 키워드는 '북한산둘레길'로 설정하였으며, 블로그 데이터의 시간적 범위는 2019년 1월 1일부터 12월 31일까지 일 년 동안 작성된 게시물을 대상으로 하였다. 또한 북한산둘레길에 대한 단편적인 내용보다는 세세한 내용까지 분석에 포함하기 위하여 블로그 원문을 대상으로 분석을 진행하였다.

데이터 수집에는 텍스트(TEXTOM)을 활용하였다. 텍스트는 소셜 매트릭스 프로그램으로 포털사이트의 블로그, 뉴스 등의 자료뿐만 아니라 트위터까지 수집 가능하며 수집, 분석, 시각화 등이 가능해, 텍스트 데이터 분석에 유용하게 활용 가능한 프로그램으로 알려져 있다(Lee *et al.*, 2018).

(2) 데이터 정제

블로그 게시물은 본 연구에서 활용 가능한 내용뿐만 아니라 광고, 대가성 후기 글 등 많은 내용을 담고 있기 때문에, 본 연구에서 유의미한 내용을 추려내기 위한 정제작업이 필요하다.

수집된 데이터 중 '북한산둘레길' 키워드를 중심으로 작성된 맛집, 부동산 홍보 등의 내용은 제외하고, 경관 또는 방문 후기 등의 본 연구에서 유의미한 내용을 담고있는 게시물을 구분하였다. 해당 과정에서 '북한산'과 '둘레길' 키워드가 분리되어 작성된 경우와 '둘레길' 키워드만 포함하고 있는 게시물에 대해서도, 내용 검토 과정을 통해 본 연구에 유의미한 내용을 담고 있는 게시물을 포함하였다.

또한 본 연구에서는 21개의 둘레길 구간을 대상으로 선호도를 평가하고자 수집된 둘레길 전체 게시물 중 구간별 게시물을

구분하는 작업을 수행하였다.

이후 구간별 게시물 데이터에 대하여 다시 한 번 내용을 검토한 후, 단순한 나열식 둘레길 코스 소개, 향후 방문 계획, 단순 정보 전달성 게시물 등의 글은 삭제하고, 최종적으로 각 구간별 탐방 기록만 추출하였다.

2) 감성분석(Sentiment Analysis)

북한산둘레길 각 구간에 대한 탐방객의 선호도를 파악하기 위하여 구간별 블로그 게시글을 대상으로 감성분석을 실시하였다. 감성분석 결과로 구간별 감성점수를 도출하고 각 구간에 대하여 블로그 게시물의 긍정·부정·중립 정도를 살펴봄으로써 탐방객의 선호 구간과 비선호 구간을 평가하였다.

감성분석 점수 산출 및 블로그 게시물의 감성 극성 판별에는 R studio를 활용하였고, 둘레길의 선호 및 비선호 구간 시각화에는 Qgis 3.14를 사용하여 분석을 진행하였다.

(1) 감성분석 점수 도출

본 연구에서는 'KNU 한국어 감성사전¹⁾'을 활용하여 북한산둘레길에 대한 탐방객의 감성점수를 산출하였다. KNU 한국어 감성사전은 긍정 단어 4,871개, 부정 9,829개, 중립 154개로 총 14,854개의 감성 단어로 구성되어 있으며, 관용구, 축약어, 이모티콘 등을 포함하고 있다. 또한 감성 단어에 대하여 매우 긍정(+2), 긍정(+1), 중립(0), 부정(-1), 매우 부정(-2) 총 5단계의 감성 정도(degree) 값 정보가 표기되어 있다.

본 연구에서는 앞서 수집 및 구분된 구간별 블로그 게시글에 대하여 감성 사전을 기반으로 긍정 및 부정에 해당하는 단어를 판별하고 점수를 부여하였다. 이때, 감성사전에 해당되지 않는 단어에 대해서는 0점으로 계산되는 방식으로 진행하였다.

이후 구간별 감성 단어의 출현 빈도로 감성점수를 산출하였다. 산출 방식은 감성분석 관련 국내외 선행연구를 참고하여, 각 게시물의 긍정에 해당하는 단어 수의 점수 합에서 부정을 의미하는 단어 수의 점수 합을 뺀 값으로 계산하였다(Lee *et al.*, 2016; Joyce and Deng, 2017; Naldi, 2019; Hausmann *et al.*, 2020).

따라서 감성점수가 양(+)의 값이면 이는 해당 게시물에서 긍정적인 단어가 부정적인 단어보다 상대적으로 많이 출현한 것을 의미한다. 반대로 감성점수가 음(-)의 값이면 긍정적인

단어보다 부정적인 단어가 많이 등장한 것이다.

(2) 구간별 선호도 평가

북한산둘레길 전 구간에 대하여 감성 점수 결과를 도출하고, 이를 도면화하여 탐방객의 선호 구간과 비선호 구간을 구별하였다.

이후 구간별 블로그 게시물을 해석하기 위해서는 게시물 단위로 긍정, 부정, 중립 문서로 판별하고, 그 비율을 분석하였다.

각 게시물의 감성 극성은 앞서 도출된 감성 단어 점수를 기반으로 도출하였으며, 각 게시물의 감성 점수 값이 0보다 크면 '긍정적', 0보다 작으면 '부정적' 그리고 점수 값이 0이면 '중립' 성격을 가진 게시물로 구분하였다(Hausmann *et al.*, 2020).

3) 선호·비선호 구간별 특징

앞서 도출된 선호·비선호 구간에 대해 블로그 원문을 활용하여 내용분석을 실시하였다. 선호와 비선호 글을 읽고, 선호하는 구간에 대해서는 어떤 요소가 탐방객의 선호도에 영향을 미쳤는지 둘레길에서의 선호·비선호 요인을 도출하였다.

III. 연구결과

1. 데이터 수집 결과 및 분석데이터 현황

연구 초기에 2019년 1월부터 12월까지 일 년 동안 북한산둘레길에 대하여 작성된 블로그 게시물의 수집한 결과 총 17,552건이 수집되었다.

이 중 '북한산둘레길' 키워드를 포함하고 1구간부터 21구간까지의 구간명이 포함된 단어로 정제한 결과는 Table 1과 같이 7,992건이 도출되었으며, 중복 또는 무의미한 게시물 제거 등의 2차 정제과정을 거쳐 최종적으로 2,781건을 본 연구에 활용하는 분석 데이터로 구축하였다.

2. 블로그 감성분석을 활용한 북한산둘레길 선호도 평가

1) 둘레길 전체 구간 감성분석 점수

구간별 블로그에 등장한 전체 단어는 평균 약 19,723개이고, 이중 긍정적 또는 부정적인 감성을 지닌 단어는 평균 582개로,

Table 1. Analysis data amount

Course	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	Total
Collected	956	710	536	434	423	302	379	739	347	276	190	199	201	185	217	159	154	323	259	288	715	7,992
Analysis data (Refined)	234	215	172	133	137	94	163	257	140	123	80	94	79	80	86	63	61	136	127	118	183	2,781

* The initial amount collected was 17,552, and "Collected" is the amount of blog posts including the keyword "Bukhansan Dulegil".

Table 2. Result of sentiment analysis

Course	Doc	Doc length	Sentiment words							Senti score
			Pos(+2)	Pos(+1)	Neg(-1)	Neg(-2)	Pos count(%)	Neg count(%)	Senti total	
8	257	214.29	698	673	261	167	1,371(76.21)	428(23.79)	1,799	1,474
21	183	212.82	533	435	158	114	968(78.06)	272(21.94)	1,240	985
1	234	114.95	333	325	101	65	658(79.85)	166(20.15)	824	760
9	140	160.84	325	249	136	72	574(73.40)	208(26.60)	782	619
7	163	169.05	273	290	115	101	563(72.27)	216(27.73)	779	519
10	123	170.22	236	221	80	78	457(74.31)	158(25.69)	615	457
18	136	227.68	290	287	137	138	577(67.72)	275(32.28)	852	454
3	172	104.13	207	199	81	67	406(73.29)	148(26.71)	554	398
2	215	107.78	199	178	88	80	377(69.17)	168(30.83)	545	328
15	86	212.12	169	192	84	41	361(74.28)	125(25.72)	486	364
19	127	110.60	186	135	66	54	321(72.79)	120(27.21)	441	333
5	137	122.14	183	133	79	57	316(69.91)	136(30.09)	452	306
4	133	70.89	131	92	42	23	223(77.43)	65(22.57)	288	266
6	94	122.17	137	135	51	54	272(72.15)	105(27.85)	377	250
11	80	181.46	143	141	70	59	284(68.77)	129(31.23)	413	239
14	80	164.18	157	124	54	73	281(68.87)	127(31.13)	408	238
20	118	130.80	136	128	67	48	264(69.66)	115(30.34)	379	237
12	94	134.82	101	158	34	49	259(75.73)	83(24.27)	342	228
16	63	157.65	99	82	28	40	181(72.69)	68(27.31)	249	172
13	79	79.03	80	65	23	16	145(78.80)	39(21.20)	184	170
17	61	136.61	77	54	52	38	131(59.28)	90(40.72)	221	80
Total_ave	132	147.82	223	205	86	68	428(72.6)	154(27.4)	582.4	422.7

* 'Doc Length' refers to the average number of words in documents by course.

전체 게시물에서 약 3%에 해당한다(Table 2 참조).

도출된 감성 단어 중 긍정적인 감성을 반영한 단어는 전 구간에서 평균적으로 약 73%이며, 반대로 부정적인 감성 관련 단어는 약 27% 출현하였다. 긍정 감성 단어의 출현이 부정 감성 단어보다 2배 이상 도출된 것을 통해 탐방객들은 북한산둘레길에 대하여 전반적으로 긍정적으로 인식하고 있는 것을 알 수 있다.

감성 점수는 8구간이 1,474점으로 가장 높게 나타났으며, 21구간(985), 1구간(760), 9구간(619), 7구간(519) 순으로 높았다. 반면 점수가 가장 낮게 나타난 구간은 17구간으로 80점이며, 13구간(170), 16구간(172), 12구간(228) 순이다.

2) 탐방객의 선호·비선호 구간 평가

앞서 도출된 감성 점수 결과를 구간별로 시각화한 결과는

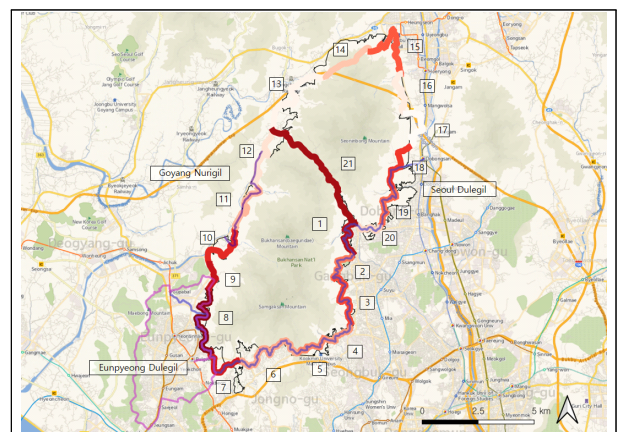


Figure 1. Preference for each course of Bukhansan Dulegil recognized by visitors

Legend: ■ Senti score(80-1,474)

Figure 1와 같다. 감성 점수에 따라서 탐방객의 선호 구간과 비 선호 구간의 분포를 살펴보았을 때, 8코스, 1코스, 9코스, 7코스와 같이 상대적으로 북한산 구역에 위치한 탐방로 구간이 도봉산 구역에 위치한 탐방로 구간보다 대체로 선호 구간으로 나타났다.

3. 구간별 선호도 특징

북한산둘레길 구간별 블로그 게시물에 대한 긍정, 부정, 중립 감성 극성 값을 띄는 문서의 비율은 Table 3과 같으며, 구간마다 정도의 차이는 있으나 대체로 긍정적인 의견의 게시물 비율이 부정적인 것보다 높게 나타났다.

중립 극성 값을 가진 문서의 비율이 의외로 높은 이유는 문서에 대하여 감성 극성 결정에 있어 블로그에 작성된 단어가 감성사전에 포함되지 않는 경우가 많은 것으로 나타났다. 또

다른 이유로는 긍정 단어와 부정 단어가 모두 등장하고 있으나, 긍정과 부정 단어의 상쇄로 감성점수가 0이 되는 경우도 존재하기 때문이다.

북한산둘레길 21개 구간에 대하여 탐방객의 선호도가 높은 구간과 낮은 구간의 특징이 무엇인지, 탐방객이 어떤 요인으로 인해 선호 또는 선호하지 않는지 구체적으로 파악하기 위하여 구간별 내용분석을 실시하였다. 내용분석은 선호도가 높은 상위 세 구간과 선호도가 낮은 하위 세 구간에 대하여 수행하였다. 상위 세 구간은 8구간, 21구간, 1구간이며, 하위 세 구간은 17구간, 13구간, 16구간이다.

1) 선호도 높은 구간의 특징

탐방객이 8구간을 선호하는 이유 중 가장 많은 부분을 차지하는 것은 전망의 우수함으로 나타났다(Table 4 참조). 블로그 원문 내용을 살펴보면 '전망', '경치', '조망' 등의 키워드가 빈번

Table 3. Proportion of document's sentiment characteristic

Course	Doc	Senti score	Positive		Negative		Neutral
			Proportion(%)	Score_mean(SD)	Proportion(%)	Score_mean(SD)	Proportion(%)
8	257	1,474	76.26	7.81(6.77)	7.00	-1.56(0.51)	16.73
21	183	985	74.32	5.99(7.74)	9.84	-2.11(1.18)	15.85
1	234	760	59.83	6.27(7.52)	7.69	-2.06(1.26)	32.48
9	140	619	67.86	6.68(6.68)	6.43	-1.78(0.44)	25.71
7	163	519	61.96	5.74(5.47)	11.04	-3.39(2.03)	26.99
10	123	457	58.54	6.83(6.18)	17.07	-1.67(1.11)	24.39
18	136	454	64.71	5.82(6.07)	18.38	-2.32(1.63)	16.91
3	172	398	58.72	4.38(3.93)	10.47	-2.44(2.43)	30.81
2	215	328	46.51	3.98(3.53)	13.02	-2.50(1.71)	40.47
15	86	364	66.28	6.68(6.14)	6.98	-2.83(1.72)	26.74
19	127	333	61.42	4.88(4.67)	18.11	-2.09(1.35)	20.47
5	137	306	55.47	4.45(3.11)	11.68	-2.00(1.10)	32.85
4	133	266	50.38	4.15(3.87)	6.02	-1.50(0.53)	43.61
6	94	250	61.70	4.84(3.99)	12.77	-2.58(1.62)	25.53
11	80	239	56.25	5.60(7.76)	7.50	-2.17(0.41)	36.25
14	80	238	57.50	6.41(6.57)	17.50	-4.07(2.70)	25.00
20	118	237	55.93	3.94(3.47)	9.32	-2.09(1.22)	34.75
12	94	228	44.68	5.95(6.95)	14.89	-1.57(0.76)	40.43
16	63	172	50.79	5.94(5.55)	12.70	-2.25(1.28)	36.51
13	79	170	53.16	4.14(4.48)	5.06	-1.00(0)	41.77
17	61	80	32.79	5.40(5.73)	18.03	-2.55(2.30)	49.18
Total_ave	132	147.82	57.86	-	11.50	-	30.64

Table 4. Preferred and non-preferred Factors of Dulegil

Positive					
Course 08		Course 21		Course 01	
Preferred factor	Frequency(%)	Preferred factor	Frequency(%)	Preferred factor	Frequency(%)
Outstanding view(전망 우수)	109(42.4)	Outstanding view(전망 우수)	82(44.8)	Comfortable road(편한 길)	58(24.8)
Course overlay(코스 중첩)	87(33.9)	Comfortable road(편한 길)	71(38.8)	Hearing, smell(청각, 후각)	36(15.4)
Comfortable road(편한 길)	55(21.4)	A sense of season(계절감)	39(21.3)	Accessibility(접근성)	34(14.5)
Tree(수목)	36(14.0)	Ecological excellence(생태적 우수)	31(16.9)	Signpost, information(이정표, 정보)	28(12.0)
Landscape sequence change (경관 시퀀스 변화)	23(8.9)	Landscape element harmony (경관 요소의 조화)	31(16.9)	Tree(수목)	24(10.3)
Accessibility(접근성)	20(7.8)	Continuous landscape(연속 경관)	25(13.7)	A sense of season(계절감)	18(7.7)
Landscape element harmony (경관 요소의 조화)	16(6.2)	Silence(고요함)	22(12.0)	Silence(고요함)	13(5.6)
The surrounding landscape (주변 경관)	6(2.3)			Facilities, observatory (편의시설, 전망대)	10(4.3)
				Mysteriousness, anticipation (신비함, 기대감)	5(2.1)
				The Surrounding landscape (주변 경관)	2(0.9)
				Landscape change(경관 변화)	2(0.9)
Negative					
Course 17		Course 13		Course 16	
Non-preferred factor	Frequency(%)	Non-preferred factor	Frequency(%)	Non-preferred factor	Frequency(%)
Inappropriate path(부적절할 동선)	10(16.4)	Lack of guidance(안내문 부족)	7(8.9)	Hard road(힘든 길)	36(57.1)
Noise(소음)	7(11.5)	High percentage of city section (시가지 비율)	5(6.3)	Too few visitors(적은 탐방객 수)	15(23.8)
High percentage of city section (시가지 비율)	6(9.8)	Noise(소음)	5(6.3)	Lack of guidance(안내문 부족)	3(4.8)
Insincere photo points (무성의한 포토포인트)	5(8.2)	Hard road(힘든 길)	5(6.3)	Safety device required (안전장치 필요)	2(3.2)
Absence of management and landscape disconnection (관리 부재, 경관 단절)	4(6.6)	Need to improve the landscape (경관 개선 필요)	3(3.8)	Low accessibility(낮은 접근성)	2(3.2)
Absence of course information and need improvement (구간 정보 부족, 개선 필요)	3(4.9)	Loneliness(쓸쓸함)	2(2.5)	Lack of view(조망 부족)	2(3.2)
Lack of view(조망 부족)	2(3.3)			Need improvement(정비 필요)	1(1.6)

* Frequency(%) is the number of content that appears in positive or negative documents, so the sum of the frequency may be over or below 100%.

* If factors appear more than twice, they are expressed in bold.

하게 등장하고 있으며, 탁 트인 하늘과 풍경을 볼 수 있어 기분이 상쾌해진다는 내용이 주를 이루고 있다. 또한 코스를 따라 걸으면서 마주하는 경관 시퀀스가 극적으로 변화하고, 빼곡한 소나무, 오래된 느티나무 등 수목이 주는 청량한 분위기와 물길-흙길-숲길 변화 등 다양한 경관요소의 조화로움이 존재하고

있어 이용자들이 이와 같은 요소를 선호하는 것을 확인하였다.

둘레길 외적 요소로서 진입로의 편의성, 지하철역과 버스 정류장 등의 우수한 접근성 또한 탐방객이 8구간을 선호하는 요인으로 도출되었다.

8구간은 특히 서울둘레길, 은평둘레길 등 타 둘레길과 중첩

되는 일부 구간이 존재하기 때문에 다른 구간들보다는 탐방객의 방문 빈도가 높고, 그로 인해 작성된 블로그 게시물의 수가 많았다. 이러한 요인들도 8구간의 선호도가 더 높게 나타난 이유임을 알 수 있었다.

21구간은 다른 구간에 비해 계절감과 관련된 내용이 많이 등장하며, '단풍' 키워드가 높은 빈도로 등장하며, 단풍 명소를 인식되는 구간으로 나타났다.

또한 북한산둘레길 중 유일하게 예약제로 운영되고 있어 탐방객 수가 제한적이라 이용자들에게 한적한 느낌을 제공하고 있었고, 그로 인해 탐방객은 적막감과 편안함을 느낄 수 있다고 하였다. 또한 개통된 지 오래되지 않아 생태적으로 우수한 환경이 제공되는 점이 선호 요인으로 나타났다.

등산로와 일부 중복되는 몇몇 구간의 난이도가 높은 것에 비해 21구간은 경사가 완만한 숲길이라 편안해서 좋다는 의견이 다수 등장하기도 한다. 게다가 조망대상인 '오봉'을 계속 바라보면서 구간을 걸을 수 있는 연속되는 경관에 대해 탐방객들은 우수한 경관을 지속적으로 감상 가능해서 좋다는 의견을 보이고 있다.

1구간에서의 선호요인은 8구간과 21구간의 요인과 비슷한 맥락을 보이고 있으나, 후각 및 청각에 대한 내용은 1구간의 특징으로 등장하였다. 탐방객들은 1구간의 소나무 군락지에 진입했을 때 느껴지는 솔향이 시원한 느낌을 받게 하면서도 다른 구간에서 경험하지 못했던 점이라 선호한다고 하였다. 또한 유일하게 계곡을 따라 걸으며 물소리를 들을 수 있는 구간인 점을 통해 시각, 후각, 청각 모두 만족할 수 있는 구간으로 인식되고 있다.

또한 1구간에서는 이정표나 안내판 등의 정보 제공의 부족으로 인해 겪는 불편함에 대하여 작성된 문서가 없는 것으로 나타났으며, 이에 대한 만족도가 높은 것을 알 수 있었다. 방향 설정에 있어 다양한 선택지가 존재하고 등산로와 중복되는 구간이 많은 둘레길에서 이러한 점은 탐방객에게 선호하는 요소로 작용하고 있는 것을 알 수 있었다.

선호도 높은 구간에 대한 내용분석 결과, 세 구간에서 탐방객들은 공통적으로 어려운 코스보다는 부담 없이 편하게 걸을 수 있는 구간을 선호하는 것으로 나타났다.

등산에 요구되는 기본 체력, 장비 등이 없어도 가볍게 산책하듯이 트레킹 할 수 있는 구간에 대하여 선호도가 높았고, 이는 '가족', '아이', '엄마' 등의 키워드 출현을 통해 동반인과 함께 가벼운 트레킹, 산책 등을 하고자 방문하는 경우가 많은 것을 알 수 있었다.

다음으로 연속되는 경관과 시퀀스가 극적으로 변화하는 등 다양한 경관을 마주했을 때와 '단풍', '설경' 등 경관에 계절감이 부여될 때 선호도가 높은 것으로 나타났다. 또한 원경을 감상할 수 있는 '조망점', '전망대' 등의 뷰포인트(view point)를

둘레길의 주요 요소로 인식하고 있었고, 우수한 접근성과 정보 안내 및 공유 정도에 따라서도 구간에 대한 선호도가 높아지는 것을 확인하였다.

2) 선호도 낮은 구간의 특징

17구간의 선호도가 낮은 요인으로 둘레길 동선에 대한 불만족이 큰 부분을 차지하는 것으로 나타났다(Table 4 참조). 동선 불만족에 대한 세부 내용으로는 고속도로와 차도가 인접한 부분에서의 소음 및 매연 발생, 과도한 시가지 비율, 아스팔트 도로 포장 구간, 사격장 등 기타 주변 소음으로 인해 발생하는 불편함 및 불안감 조성 등이 언급되었다.

국립공원 측에서 지정한 포토포인트에 대해서는 관리되지 않고 있으며 다소 무성의 혹은 무의미하게 지정된 것으로 느껴지고, 구간을 대표하는 포토존에 대한 의문을 제기하는 내용이 등장하기도 하였다. 또한 구간에 대한 정보가 부족하고 이정표를 좀 더 명료하게 표시할 필요가 있다고 언급되기도 하였다. 이외에도 17구간을 걷다보면 관리가 필요한 부분이 자주 보인다고 하였으며, 이러한 요소들이 탐방객들에게 비선호를 유발하고 있어서 둘레길 전체 구간에서 선호도가 가장 낮게 나타난 것으로 해석된다.

13구간과 16구간에서도 비선호 요인에 대하여 17구간과 비슷한 항목이 언급되고 있다. 13구간에서는 진입로와 그 주변에 대한 이정표 및 안내에 대해 탐방객들이 헷갈리는 경우가 발생해 둘레길이 아닌 다른 길로 진입하는 문제점이 비선호를 유발한 가장 큰 요인으로 나타났다. 그 외에도 시가지 및 포장도로를 지나는 점이 불만족스럽다고 말하고 있고, 공사로 인해 발생하는 경관이 개선될 필요가 있다고 하였다. 또한 해당 구간을 지나는 탐방객이 적어서 다소 쓸쓸하고 스산한 분위기가 연출되는 것에 대해서도 비선호를 유발하는 것으로 나타났다.

16구간은 다른 둘레길 구간에 비해 난이도가 높는데, 이에 대하여 약 57%가 부정적으로 언급하고 있었다. 이는 선호도가 높은 구간이 편한 길이기 때문에 선호한다는 앞의 내용과 상응하는 결과로 볼 수 있다. 또한 난이도가 높은 구간에서는 탐방객 수가 적고, 이로 인해 쓸쓸함과 동시에 불안감, 무서움을 느끼고 있었다.

선호도 낮은 구간에서의 요인은 공통적으로 동선에 대한 불만족이 가장 큰 부분을 차지하고 있었으며, 구간 정보 부재 및 개선 필요, 무성의한 포토포인트 지정, 관리부재 및 경관 단절, 둘레길 주변 경관 개선 필요 등 다양한 요인이 나타났다.

특히 동선에 대한 불만족이 세 구간 모두 선호도가 낮은 원인으로 크게 작용하고 있는 것을 확인하였고, 이는 세부적으로 주변 소음, 구간 내 시가지 통과 비율, 구간별 난이도 불균형 등이 그 원인으로 도출되었다. 또한 구간 정보 부재 및 개선 필요 항목에 대해서는 앞서 언급된 바와 같이 이정표가 기존보다

더 명료하게 표시될 필요가 있으며, 둘레길 구간과 등산로가 겹치는 곳, 타 둘레길과 중첩되는 지점에서 추가적인 정보를 제공함으로써 탐방객에게 혼란이 생기지 않도록 할 필요가 있는 것으로 나타났다.

이와 같이 탐방객이 인식하는 북한산둘레길의 문제점 및 비선호 요인에 대하여 관리주체가 인지하고 향후 개선방안 수립에 적용할 필요가 있다.

IV. 결론

본 연구는 실제 이용자가 작성한 블로그 데이터를 자연어 처리 기술 중 하나인 감성분석을 활용하여 북한산둘레길 21개 구간에 대한 선호도를 평가하고 특성을 분석하였다. 각 구간별 블로그 텍스트에 나타난 긍정·부정 감성 단어의 출현 빈도로 감성점수를 도출하고 내용분석을 통해 선호·비선호 요인을 정리하였다. 본 연구의 결과는 다음과 같다.

첫째, 북한산둘레길에 대하여 탐방객이 작성한 블로그 게시물에서는 전반적으로 긍정적인 견해를 보이고 있었다. 작성된 글에는 긍정에 해당되는 감성 단어가 부정에 해당되는 단어보다 월등히 많이 등장하고 있었다. 21개 전체 구간에서 긍정적인 단어는 평균적으로 약 73%가 출현하였으며, 이는 약 27%의 부정적 단어보다 2배가 넘게 작성되었다. 따라서 탐방객은 북한산둘레길에서 전반적으로 만족하며 긍정적으로 인지하고 있는 것으로 나타났다.

둘째, 감성 단어에 정도(degree) 값을 적용해 감성 점수를 계산한 결과, 21개 구간 중 8구간이 가장 선호도가 높고 17구간이 선호도가 가장 낮은 것으로 나타났다. 이러한 결과는 구간별 긍정 및 부정 단어의 출현 빈도 결과에서도 이와 비슷한 경향을 보이고 있었다. 이를 통해 감성 단어 중 긍정 감성 단어의 출현 빈도가 전체 점수에 영향을 미치고 있는 것을 확인하였다.

셋째, 도출된 선호·비선호 구간에 대하여 내용분석을 통해 선호하거나 비선호하는 요인을 도출한 결과, 탐방객은 난이도가 낮고 아이 또는 가족과 함께 부담 없이 걸을 수 있는 구간을 선호하고 있었고, 여러 경관 요소가 조화된 곳, 계절감이 뚜렷해 다양한 경관이 연출되는 곳, 경관 시퀀스의 변화가 존재하는 곳을 선호하는 것으로 나타났다. 또한 탐방객은 전망대, 조망점 등의 뷰포인트 유무를 중요한 요소로 인식하고 있었고 접근성의 편리함, 정보 제공이 원활하게 이뤄지는 곳에 대해 선호도가 더 높은 것을 알 수 있다.

반면, 도로와 인접한 구간에서의 소음 발생, 시가지 비율, 구간별 난이도 불균형 등으로 인한 둘레길 동선 불만족이 비선호 요인으로 크게 작용하고 있었으며, 구간에 대한 정보 제공 부족과 경관관리 부족 및 경관 단절 등 또한 선호도를 떨어트리는 원인으로 나타났다.

기존 소셜미디어 데이터를 활용한 텍스트마이닝 연구에서는 형태소 분석을 통해 형용사 출현 빈도에 기반하여 선호 정도를 파악하였으나, 본 연구에서 적용한 감성분석은 실제 이용자가 경험한 내용을 토대로 작성된 주관적 견해에 대하여 감성 정도를 수치화하여 평가하였다. 이와 함께 내용분석을 수행하여 선호 구간과 비선호 구간에 대하여 세부적인 내용을 파악하고, 요인을 도출하여 평가 대상지의 특성을 정리했다는 점에서도 의의가 있다.

본 연구의 결과는 국립공원뿐만 아니라 근교 산림 녹지 관리에 있어 둘레길 정비 및 개선방안 마련에 활용될 수 있을 것으로 기대할 수 있다. 또한 연구에 활용된 연구 방법은 자연지역에 대한 실제 이용자들의 반응을 지속적으로 모니터링 할 수 있다는 점에 의의가 있다.

단 소셜미디어에서 작성되는 텍스트는 여러 단어를 결합한 약어, 속어 등이 존재하며, 새롭게 만들어지는 텍스트의 변화 속도는 매우 빠르다.

본 연구는 사전에 정의된 감성사전을 기반으로 평가를 진행함에 따라 이와 같은 내용을 세세하게 반영하여 평가를 진행하기에 한계가 있었다. 따라서 정확한 감성분석을 위해서는 지속적인 사전 업데이트가 요구된다고 하겠다.

또한 국립공원 탐방과 같은 소재는 소셜미디어 특성상 부정적인 견해보다는 긍정적인 내용을 공유하고자 작성되는 경우가 빈번하기 때문에 상대적으로 부정적인 내용의 게시글이 적었다고 할 수 있다(Waterloo et al., 2018). 따라서 설문조사 등의 추가적인 평가 방법을 적용하여 결과를 비교하고 검토해 볼 필요가 있을 것이다.

주 1. KNU 한국어 감성사전은 군산대학교 소프트웨어융합공학과에서 구축한 사전으로, 오픈소스로 접근 가능하고 R 소프트웨어에서 활용 가능하다.

References

- Ahern, J.(2013) Urban landscape sustainability and resilience: the promise and challenges of integrating ecology with urban planning and design. *Landscape Ecology* 28(6): 1203-1212.
- Bell, S., L. Tyrväinen, T. Sievänen and U. Pröbstl(2007) Outdoor recreation and nature tourism: A European perspective. *Living Reviews in Landscape Research* 1: 1-46.
- Bryant, M. M.(2006) Urban landscape conservation and the role of ecological greenways at local and metropolitan scales. *Landscape and Urban Planning* 76(1-4): 23-44.
- Casalegno, S., R. Inger, C. Desilvey and K. J. Gaston(2013) Spatial covariance between aesthetic value & other ecosystem services. *PLoS One* 8(6): e68437.
- Chiesura, A.(2004) The role of urban parks for the sustainable city. *Landscape and Urban Planning* 68(1): 129-138.
- Dallat, M. A. T., I. Soerjomataram, R. F. Hunter, M. A. Tully, K. J.

- Cairns and F. Kee(2014) Urban greenways have the potential to increase physical activity levels cost-effectively. *European Journal of Public Health* 24(2): pp. 190-195.
7. Daniels, B., B. S. Zaunbrecher, B. Paas, R. Ottermanns, M. Ziefle and M. Roß-Nickoll(2018) Assessment of urban green space structures and their quality from a multidimensional perspective. *Science of The Total Environment* 615: 1364-1378.
 8. Dunkel, A.(2015) Visualizing the perceived environment using crowdsourced photo geodata. *Landscape and Urban Planning* 142: 173-186.
 9. Fitzhugh, E. C., D. R. Bassett and M. F. Evans(2010) Urban trails and physical activity: a natural experiment. *American Journal of Preventive Medicine* 39(3): 259-262.
 10. Hausmann, A., T. Toivonen, C. Fink, V. Heikinheimo, R. Kulkarni, H. Tenkanen and E. D. Minin(2020) Understanding sentiment of national park visitors from social media data. *People and Nature* 2(3): 750-760.
 11. Joyce, B. and J. Deng(2017) Sentiment analysis of tweets for the 2016 US presidential election. *IEEE MIT Undergraduate Research Technology Conference(URTC)*, Cambridge, MA, USA. pp. 1-4.
 12. Kabisch, N., S. Qureshi and D. Haase(2015) Human-environment interactions in urban green spaces - A systematic review of contemporary issues and prospects for future research. *Environmental Impact Assessment Review* 50: 25-34.
 13. Landers, D. H. and A. M. Nahlik(2013) Final ecosystem goods and services classification system. (FECS-CS) (Report No. EPA/600/R-13/ORD-004914), Environmental Protection Agency, Washington, DC.
 14. Larson, L. R., V. Jennings and S.A. Cloutier(2016) Public parks and wellbeing in urban areas of the United States. *PLoS ONE*, 11(4), p. e0153211. 10.1371/journal.pone.0153211.
 15. Lee, S. H., D. H. Park and M. S. Kim(2018) A study of perceptions of Jeju-Island tourism using social media big data analysis: Before and after the outbreak of ban on Korean entertainer's. *Journal of Tourism and Leisure Research* 30(2): 55-75.
 16. Lee, S. H., J. Cui and J. W. Kim(2016) Sentiment analysis on movie review through building modified sentiment dictionary by movie genre. *Korea Intelligent Information System Society* 22(2): 97-113.
 17. Lee, S. H. and Y. H. Son(2018) Identifying landscape perceptions of visitors' to the taean coast national park using social media data : focused on Kkotji Beach, Sinduri coastal sand dune, and Manlipo beach. *Journal of the Korean Institute of Landscape Architecture* 46(5): 10-21.
 18. Liu, B.(2012) Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*.
 19. Mäntylä, M. V., D. Graziotin and M. Kuutila(2018) The evolution of sentiment analysis - A review of research topics, venues, and top cited papers. *Computer Science Review* 27: 16-32.
 20. Martínez Pastur, G., P. L. Peri, M. V. Lencinas, M. García-Llorente and B. Martín-López(2015) Spatial patterns of cultural ecosystem services provision in Southern Patagonia. *Landscape Ecology*.
 21. Naldi, M.(2019) A review of sentiment computation methods with R packages. *arXiv:1901.08319*.
 22. Nam, M. J.(2015) A Study on Classification Method of User Sentiment Using SNS Hashtags, Master's Thesis, Chosun University, Korea.
 23. OPENSURVEY.(2020) Reports on social media and search portal 2020, OPENSURVEY.
 24. Plieninger, T., C. Bieling, N. Fagerholm, A. Byg, T. Hartel, P. Hurley, C. A. López-Santiago, N. Nagabhatla, E. Oteros-Rozas, C. M. Raymond, D. van der Horst and L. Huntsinger(2015) The role of cultural ecosystem services in landscape management and planning. *Current Opinion in Environmental Sustainability* 14: 28-33.
 25. Plieninger, T., S. Dijks, E. Oteros-Rozas and C. Bieling(2013) Assessing, mapping, and quantifying cultural ecosystem services at community level. *Land Use Policy* 33: 118-129.
 26. Seeger, C. J.(2008) The role of facilitated volunteered geographic information in the landscape planning and site design process. *Geo-Journal* 72: 199-213.
 27. Shafer, C. S., B. K. Lee and S. Turner(2000) A tale of three greenway trails: User perceptions related to quality of life. *Landscape and Urban Planning*, 49(3): 163-178.
 28. Tenerelli, P., U. Demšar and S. Luque(2016) Crowdsourcing indicators for cultural ecosystem services: A geographically weighted approach for mountain landscapes. *Ecological Indicators*, 64: 237-248.
 29. van Zanten, B. T., D. B. van Berkel, R. K. Meentemeyer, J. W. Smith, K. F. Tieskens and P. H. Verburg(2016) Continental scale quantification of landscape values using social media data. *Proceedings of the National Academy of Sciences of the United States of America* 113(46): 12974-12979.
 30. Walmsley, A.(2006) Greenways: Multiplying and diversifying in the 21st century *Landscape and Urban Planning* 76(1): 252-290.
 31. Waterloo, S. F., S. E. Baumgartner, J. Peter and P. M. Valkenburg(2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. *New Media & Society*, 20, 1813 - 1831.
 32. Heraldcorp, <http://news.heraldcorp.com/view.php?ud=20200617000670>, 2020.06.17
 33. KNPS, <http://www.knps.or.kr>
 34. KNU Korean Sentiment Dictionary, <http://dilab.kunsan.ac.kr/knusl.html>

Received : 12 April 2021

Revised : 29 April 2021

Accepted : 29 April 2021

(1st)

3인익명 심사필