

빅데이터를 활용한 공원 이용행태의 시계열분석

- 올림픽공원을 대상으로 -

우경숙* · 서주환**

*경희대학교 일반대학원 환경조경학과 · **경희대학교 예술·디자인대학 환경조경디자인학과

Time Series Analysis of Park Use Behavior Utilizing Big Data - Targeting Olympic Park -

Woo, Kyung-Sook* · Suh, Joo-Hwan**

*Graduate School of Landscape Architecture, Kyung Hee University

**Dept. of Landscape Architecture, Kyung Hee University

ABSTRACT

This study suggests the necessity of behavior analysis as changes to a park environment to reflect user desires can be implemented only by grasping the needs of park users. Online data (blog) were defined as the basic data of the study. After collecting data by 5 - year units, data mining was used to derive the characteristics of the time series behavior while the significance of the online data was verified through social network analysis.

The results of the text mining analysis are as follows. First, primary results included ‘walking’, ‘photography’, ‘riding bicycles’(inline, kickboard, etc.), and ‘eating’. Second, in the early days of the collected data, active physical activity such as exercise was the main factor, but recent passive behavior such as eating, using a mobile phone, games, food and drinking coffee also appeared as a new behavior characteristic in parks. Third, the factors affecting the behavior of park users are the changes of various conditions of society such as internet development and a culture of expressing unique personalities and styles. Fourth, the special behaviors appearing at Olympic Park were derived from educational activities such as cultural activities including watching performances and history lessons.

In conclusion, it has been shown that people’s lifestyle changes and the behavior of a park are influenced by the changes of the various times rather than the original purpose that was intended during park planning and design. Therefore, it is necessary to create an environment tailored to users by considering the main behaviors and influencing factors of Olympic Park.

Text mining used as an analytical method has the merit that past data can be collected. Therefore, it is possible to form analysis from a long-term viewpoint of behavior analysis as well as to measure new behavior and value with derived keywords. In addition, the validity of online data was verified through social network analysis to increase the legitimacy of research results. Research on more comprehensive behavior analysis should be carried out by diversifying the types of data collected later, and various methods for verifying the accuracy and reliability of large-volume data will be needed.

Corresponding author: Joo-Hwan Suh, Dept. of Landscape Architecture, Kyung Hee University, Yong-in 17104, Korea. Tel.: +82-31-201-2680, E-mail: meek1126@naver.com

Key Words: SNS Analysis, Text Mining, Social Network

국문초록

본 연구는 공원 이용자의 욕구를 파악하여 이용자에게 적합한 공원 환경으로 변화되어야 할 필요성에 주목하고, 이용자의 욕구를 파악하기 위하여 행태분석의 필요성을 제기하였다. 이에 온라인 데이터(블로그)를 연구의 기초자료로 선정하고, 5년 단위로 구분하여 데이터를 수집한 후 텍스트 마이닝을 활용해 시계열적 행태의 특성을 도출하고, 사회연결망 분석을 통해 온라인 데이터의 유의성을 검증하였다.

텍스트 마이닝 분석 결과, 첫째, ‘길을 걷다’(산책), ‘사진을 찍다’, ‘자전거(인라인, 킥보드 등)를 타다’, ‘먹다’, ‘공연을 관람하다’는 올림픽공원에서 행해지는 공통적인 행태로 나타났다. 둘째, 수집된 데이터의 초기에는 운동 등 적극적인 신체활동을 행태가 주를 이루었지만, 최근에는 핸드폰, 게임, 음식을 먹고 커피를 마시는 등의 소극적인 비활동적 행태도 공원에서 나타나는 새로운 행태적 특징으로 나타났다. 셋째, 공원 이용자의 행태에 영향을 미치는 요인은 인터넷 발달, 자신의 개성과 스타일을 표현하는 문화 등 사회의 여러 가지 여건의 변화로 나타났다. 넷째, 올림픽공원에서 나타나는 특별한 행태는 공연 관람 등 문화적인 활동과 역사수업 등 교육적인 활동으로 도출되었다.

결론적으로 공원 계획·설계 시 의도하였던 목적보다는 여러 가지 시대적 변화로 사람들의 라이프 스타일이 변화하고, 공원의 행태에까지 영향을 미치는 것으로 나타났다. 이에 올림픽공원의 주요 행태와 영향을 미치는 요인을 고려하여 이용자에게 적합한 환경으로 변화되어야 할 필요성이 있다.

분석방법으로 활용한 텍스트 마이닝은 과거의 데이터도 수집이 가능하다는 장점이 있어 행태 분석 시 장기적인 관점에서 분석이 가능하고, 도출된 키워드로 새로운 행태 및 가치 측정이 가능하여 이후 행태분석 연구의 영역의 확대가 가능한 것으로 판단된다. 또한, 사회연결망 분석을 통해 온라인 데이터의 타당성을 검증하여 연구결과의 신뢰를 높일 수 있었다. 추후 수집하는 데이터의 종류를 다양하게 하여 더 포괄적인 행태분석에 대한 연구가 수행되어야 하며, 대용량 데이터의 정확성, 신뢰성을 검증할 수 있는 다양한 방법에 대한 연구가 필요할 것이다.

주제어: SNS분석, 텍스트 마이닝, 사회연결망

I. 서론

우리나라는 빠른 경제성장으로 도시가 발달하고, 물질이 앞서는 공간으로 변모하게 되면서 도시 내에서 쾌적한 생활공간을 확보하기 위해 공원의 필요성이 제기되어왔다. 과거 자연 자원의 보전 등을 위해 공원이 제공되었다면, 현재는 생태·경관·문화·사회적 등 사람들의 욕구를 포괄적으로 반영하는 계획된 공간으로, 지금 이 시대를 살아가는 사람들은 공원을 도시 내 필수적인 요소로 인식하고 있다.

기존의 연구에서 대부분의 도시공원들은 설계자나 계획가의 주관이나 직관에 의해 이루어졌고, 따라서 그 본래의 의도나 목적과는 다르게 이용되기도 하고, 실제 공원 이용자들에게 충분한 만족도 주지 못한 것이 현실이라고 지적한 바 있다 (Joo and Kim 2012). 이러한 현실은 이용자의 욕구를 파악하여 공간구성이나 디자인, 콘텐츠 등의 변화가 제공되어야 할 필요성을 제기하고 있다.

공원은 인간의 행태를 지원해주는 장소이고, 인간의 행태와 상호보완적이라는 맥락에서 행태 분석은 이용자의 요구를 파

악하고, 공원이라는 공간을 이해할 수 있는 기초 연구라 할 수 있다. 이에 본 연구는 공원 이용자들이 어떤 경험을 하는지 변화 과정을 고찰하고, 시계열적 행태 변화에 주목하는 것이 본 연구의 주된 목적이다. 또한, 이용자의 행태를 파악하는 과학적·합리적 접근 방법에 대한 연구의 필요성에 의하여 소셜 미디어를 기초 자료로 사용하고, 빅데이터 프로그램을 활용함으로써 행태분석 연구에서 빅데이터 연구 방법의 기대효과를 밝히고자 하는 것이 연구의 부차적인 목적이다.

II. 이론적 고찰

“도시공원 및 녹지 등에 관한 법률”이 제정되고 도시공원이 생겨나기 시작한 지 37년이 지난 지금, 사람들은 휴식을 취하고 전원적인 풍경을 보기 위해 공원을 찾는 것일까? 요즘 사람들은 공원의 한정된 공간 안에서 어떠한 경험을 하고 있을까라는 질문의 답을 찾기 위하여 이용자의 행태에 주목하였다.

도시공원의 외부공간은 활동의 장이라고 말할 수 있다. 인간의 행위가 공간과 잘 융화되면 그 공간의 질은 높은 것으로 평

가된다(Park, 2000). 이것은 인간이 심리적으로 자율적인 존재가 아니라 환경과 상호작용하여 환경을 조성하고, 그에 의해 형성되기 때문이다(Kim, 2003). 또한, 공간의 계획 상태는 사람들의 생각과 경험에 영향을 미치므로 행태분석은 사람들이 그 공간에 대해 어떻게 이해하고 있는지도 파악이 가능하다고 볼 수 있다.

행태분석 연구 시 사람들의 행태 패턴을 파악할 수 있는 기초 데이터의 수집은 중요한 과제 중 하나이다. 1990년대부터 인터넷이 급속도로 보급되면서 인터넷이 기존 미디어와 다른 것은 단순한 '전송방식'을 넘어서 '공간형식'을 갖추고 있다는 점이다(Maruta, 2011). 웹상에서 타인과 공유할 수 있는 소셜 미디어라는 공간을 만들어 자신을 표현하고, 정보를 공유하는데, 이러한 활동이 상업화되기도 한다.

소셜 미디어의 성장은 단순히 이용자 수의 증가에 머무는 것이 아니라, 사회 제 분야에 급격하게 영향력을 확장하고 있는데 의미가 있다(Kim *et al.*, 2013). 여기에서 중요한 쟁점은 소셜 미디어에서 제공되는 정보의 신뢰성과 객관성이다.

실증연구의 결과를 살펴보면 소셜 미디어가 신뢰성, 객관성, 전문성 등에 있어서 아직 전통적인 언론매체 수준에 도달하지 못하였지만, 다양성과 속보성 면에서는 오히려 전통 미디어를 압도하고 있어 기술적인 측면에서는 소셜 미디어가 우위에 있지만, 신뢰성, 객관성, 전문성 등 정보의 본질적 측면에서는 개선의 여지가 있다(Chang and Nam, 2012). 반면, 관광, 마케팅 등 많은 분야에서 사람들은 소셜 미디어를 통해 정보를 전달하고 전달받으면서 가까운 미래의 행동에 영향을 미치고 있어 한 사용자는 다른 사용자가 제공하는 정보에 신뢰를 가지고 있다고도 볼 수 있다.

국내·외에서 소셜 미디어 사용자 간의 신뢰와 객관성을 측정·예측하는 연구가 진행되고 있지만, 정확한 답을 논의하기에는 미흡한 실정이다. 그러나 공통적으로 기하급수적으로 늘어나고 있는 소셜 미디어 데이터는 중요한 정보이고, 연구의 유용한 수단이라는 태도를 취하고 있으며, 그 안에서 숨겨진 패턴을 찾아 새로운 가치를 찾아내는 방법은 계속 발전하고 있다.

III. 연구방법

1. 연구대상지 선정 및 현황

공원은 위치, 기능, 이용자의 특성 등 다양한 조건에 따라 다른 행태가 나타날 수 있으므로 여러 공원을 대상으로 행태를 분석하여 결과를 도출하더라도 공원에서 일어나는 일반적인 행태라고 판단하기는 어렵다. 이에 공원별로 이용자의 특성, 행태, 이미지 등을 분석하여 각 공원에 적합한 변화가 필요하다.

본 연구의 대상은 빅데이터 자료 기반의 특징인 시계열 분석이 가능한 이점을 활용하여 “도시공원 및 녹지 등에 관한 법률”이 제정(1980년 1월 4일)된 이후 생겨난 공원 중 규모가 100만 m² 이상, 이용자의 편의를 도모하는 복합적인 시설을 갖춘 공원, 수집되는 데이터의 양이 100만 건 이상인 점을 고려하여 ‘올림픽공원’으로 선정하였다.

올림픽공원은 서울시 송파구 올림픽로 424 일대에 위치하고, 약 43만 8천 평(1,447,122.00m²)의 면적으로 서울에서 면적이 큰 공원으로 손꼽히는 곳이다. 또한, 체육·역사·교육·휴식 등 다양한 기능을 갖고 있는 종합공원으로 이용되고 있으며, 종합운동장이 있는 도시계획 시설로 분류된다(Kim, 2013).

올림픽공원의 운영주체인 국민체육진흥공단에서는 공원 내 방객들의 편의 증진을 위하여 공원소개, 각종 공연소개, 가상현실을 이용한 공원 내 시설 찾기 등의 기능을 포함한 스마트폰 앱을 출시하였고, 사진 촬영 명소 9개를 뜻하는 ‘올림픽공원 9경’ 투어 프로그램 등 다양한 프로그램을 운영하고 있다.

본 연구의 기초 수집 자료인 블로그의 데이터양은 네이버 포털사이트에서 검색한 결과, 올림픽공원이 372,584건(2017년 12월 7일 기준)으로 양재시민의 숲(228,716건), 여의도공원(275,045건), 서울 숲(184,277건), 월드컵공원(56,195건), 선유도공원(68,187건) 등의 공원보다 데이터의 양이 월등히 많았다.

2. 연구내용 및 방법

1) 연구내용

올림픽 공원 이용자의 행태를 시간의 흐름에 따른 변화과정을 파악하기 위하여 시간적 추이로 구분하고, 텍스트 마이닝과 사회연결망 분석 방법을 실시하였다(Figure 1 참조).

텍스트 마이닝과 사회연결망 분석은 기존의 정성적인 예측

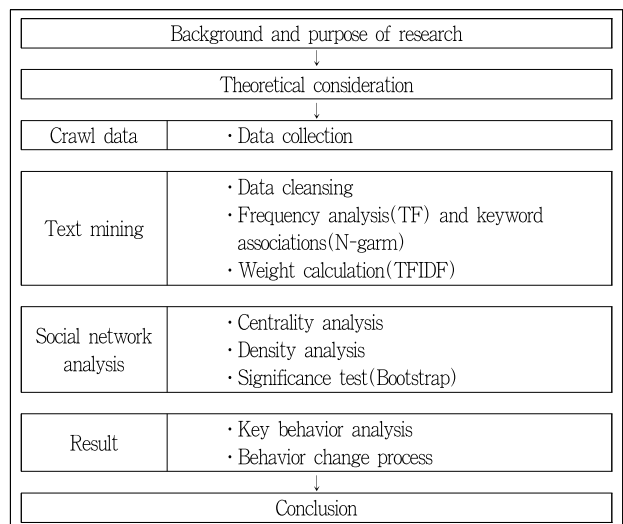


Figure 1. Study process

방법에서 보이는 전문가의 편향적인 미래예측이나 트렌드를 반영하기 어려운 기존의 방법을 보완하는 하나의 정량적인 방법으로 제시되고 있다(Jung, 2010).

텍스트 마이닝은 대용량의 데이터에서 사용자가 관심을 가지는 정보를 키워드 수준이 아니라, 배경, 정황 등의 수준의 의미를 찾아내는 과정으로(Jung, 2010) 올림픽공원에서 일어나는 주요 행태와 연관관계에 있는 배경을 파악할 수 있다. 또한, 사회연결망 분석을 통해 텍스트 마이닝의 결과에서 나타난 키워드 간의 관계를 시각적으로 파악하고, 도출된 결과의 유의성을 검증하여 이해를 도모하였다.

분석방법은 상당한 수의 블로그의 글을 대상으로 하였기 때문에 정량적 방법을 사용하지 않고서는 분석이 불가능하며, 다량의 텍스트가 있으므로 정량적 분석의 객관성이 더 확보된다는 장점 또한 존재한다(Kim, 2015).

2) 연구방법

텍스트 마이닝과 사회연결망 분석을 통해 올림픽공원의 행태를 시계열 분석하였다.

(1) 데이터 수집

텍스트 마이닝 분석과 사회연결망 분석을 위한 데이터 수집은 빅데이터 프로그램 텍스트톰(TEXTOM)을 활용하고, 검색엔진 네이버와 다음의 블로그 제목과 원문에 주제어 '올림픽공원'이 포함된 텍스트 데이터를 수집하였다. 빅데이터 프로그램인 텍스트톰은 포털 검색 사이트 네이버와 다음, 구글, 트위터, 유튜브 등의 자료를 검색하여 데이터 및 연관 키워드 순위를 제공하고, 검색 키워드의 빈도에 따른 매트릭스 정보를 제공하고 있어 네트워크분석 시 유용한 소프트웨어이다(Kim, 2015).

소셜 미디어 중 블로그(Blog)는 현재 국내에서 정보 공유의 대표적인 매체이며, 게시되는 글의 주제는 다양한데, 개인의 일상에 관한 자유로운 글과 특정 장소를 방문하여 체험한 바를 사진 이미지와 텍스트로 기술하고, 해당 장소에 관한 정보를 제공하는 글의 비중이 상당 부분을 차지하고 있어 행태분석 시 적합한 기초 자료가 될 수 있다(Lee and Chung, 2014).

수집기간은 1990년 1월 1일~2017년 12월 13일로 선정하였으나, 1990년대에는 블로그가 활성화되지 않아서 2000년대부터 수집이 가능하였다. 공원 이용자의 행태적 특성을 파악하기 위해서는 장기간에 걸친 행태분석이 이루어질 필요가 있으므로 5년 단위로 구분하여 수집하였다.

수집된 데이터는 제 I 시기(2000~2004년), 제 II 시기(2005~2009년), 제 III 시기(2010~2014년), 제 IV 시기(2015~2017년), 총 4단계로 구분하였다.

2000년부터 한해 혹은 더 세부적으로 구분하여 행태를 분석하여야 더 정확한 행태분석이 이루어질 수 있지만, '플라워마'

라는 블로그에서는 "핸드폰으로 찍은 사진들. 셀카도 보내면서 산책했어요", '쿠키'라는 블로그에서는 "예전에는 공원 같은 건 있어도 그만, 없어도 그만이었는데... 이제는 이렇게 가끔이나마 공원 산책하는 게 좋아졌어요" 등 수집된 블로그의 내용은 인용문들처럼 심도 깊은 사색이나 논쟁이 이루어지기 보다는 경험을 단편적으로 서술하는 것이 특성(Kim, 2015)으로 데이터 수집기간을 짧게 한정하면 유효한 키워드 및 의미의 결과를 도출하기 어렵다고 판단하였다.

(2) 텍스트 마이닝 분석(Text Mining Analysis)

텍스트 마이닝은 수집된 블로그 텍스트를 정제된 후 빈도분석을 통해 주요 행태를 도출하고, 행태의 의미를 부여하기 위하여 행태와 연관관계에 있는 키워드 도출하는 N-gram과 가중치를 부여하는 TF-IDF(Term Frequency - Inverse Document Frequency)를 실시하였다.

데이터 정제는 수집 데이터의 키워드의 어휘가 잘못된 데이터는 Table 1과 같이 수정하고, ~이다, ~의, ~에 등 품사, 접사 등은 불필요하여 제거하였다.

빈도분석은 수집된 전체 문서 안에서 해당 키워드가 얼마나 등장하는지 횟수를 세는 것으로 빈도분석을 통해 행태와 관련된 키워드를 구분하고, 주요 행태를 추출할 수 있다. 또한, 행태의 의미를 부여하고, 빈도분석에서는 도출되지 않은 행태의 특이점을 찾아내기 위하여 N-gram을 통하여 키워드간의 연관관계를 파악하였다.

텍스트 마이닝 진행 시 정보를 추출하는 수많은 수학적 알고리즘과 방법들이 존재하는데, 그 중 간단하면서 가장 강력한 방법으로 TF-IDF 방식을 많이 사용하고 있다(Jung, 2011). TF-IDF는 주로 문서에서 핵심어를 추출하기 위한 목적으로 특정 단어가 포함된 문서들 간의 가중치를 계산하여 높은 값을 가지는 가중치를 정렬함으로써 단어의 중요도가 결정되는 방식을 사용하고 있다(An, 2017). 빈도분석에서 도출된 키워드를 대상으로 TF-IDF를 활용하여 가중치를 부여하였다.

(3) 사회연결망 분석(Social Network Analysis)

텍스트 마이닝에서 최종 도출된 데이터를 매트릭스 데이터로 변환한 후 사회연결망 분석을 실시하였다.

사회연결망 분석은 개별적 속성보다는 관계성이 중심이 되기 때문에 노드(node)들이 어떤 관계를 맺고 있으며, 어떤 특

Table 1. Data cleansing(Part of the data cleansing process)

Before data cleansing	After data cleansing
Walk, while walking, walking etc.	Walk
Scenery, scene, view, landscape etc.	Landscape
Rest, break, relaxation, rest, take a rest	Rest

성을 보이는지를 분석할 수 있다. 또한 그래프 이론을 기반으로 하는 사회연결망 분석의 그래프는 노드를 나타내는 점과 링크(link)를 나타내는 라인으로 표현되며, 노드와 링크의 관계를 표현한다(Seo, 2016).

사회연결망 분석도구들은 다양한데, 그 중 UCINET이 가장 많이 사용되고 있으며, 다양한 연결망 분석기법을 활용할 수 있는 종합적인 프로그램이다(Kim and Chang, 2010).

이에 UCINET6을 활용하여 올림픽 공원에서 일어나는 주요 행태를 구성하는 관계의 체계, 즉 연결망 구조 파악이 가능한 중심성 및 밀도 분석과 유의성 검정을 실시하고, 연동이 가능한 넷드로우(NetDraw)를 활용하여 시각화하였다.

일반적으로 네트워크의 중심성을 분석하기 위하여 중심성 분석(Centrality Analysis)을 실시하는데, 주요 행태가 '중심'에 어느 정도 접근하고 있는지를 알 수 있다(Jung, 2010).

밀도 분석(Density Analysis)은 키워드의 연결정도를 파악할 수 있다. 밀도분석 시 사용되는 데이터는 0과 1로 구분된 매트릭스 데이터로 변환하여 적용되기 때문에 0에서 1사이의 범위 내에서 나타난다. 1에 가까울수록 키워드간의 서로 연관성이 많을수록 긴밀하게 연결되고 있고 복잡한 구조를 가지고 있으며, 밀도가 높다는 것을 의미한다(Seo, 2016).

중심성 및 밀도분석 결과의 기준은 유사한 연구결과를 참고할 수 있지만, 각각의 연구가 가지고 있는 데이터의 양이나 특징이 다르므로 오류가 있을 수 있다. 이에 본 연구에서는 대상지와 관련된 데이터를 시기별로 수집하여 데이터를 비교하여 해석하였다. 예를 들어 A시기의 밀도가 B시기 밀도보다 높게 나타나면 A시기가 B시기보다 다양한 행태가 행해지고 있으며, 그 행태들이 상호밀접하게 연관되어 형성되어 있고, A시기보다 B시기의 중심성 값이 높게 나타나면 B시기에 특정한 행태에 집중되고 있다고 해석하였다.

데이터에 대한 유의성 검정은 부트스트랩 기법(Bootstrap)으로 UCINET6을 활용하여 구분된 연차별로 통계적 검정을 실시하였다. 단일 네트워크를 대상으로 하는 밀도 또는 평균 관계강도에 대한 검정은 전통적 통계분석기법의 일 표본 평균 검정과 같은 의미를 갖는다(Gwak, 2014). 구체적으로 부트스트랩 기법은 모집단 내에서 네트워크 밀도가 0이라고 가정하고, 표본으로부터 발생한 밀도가 우연인가를 파악하기 위한 방법으로 연결망의 노드들을 무작위로 재배열하는 방식으로 5,000개 네트워크들의 밀도를 계산하고, 산출된 밀도들로 구성된 표집분포를 생성한다(Seo, 2016). 생성된 표집분포의 평균과 표준오차를 이용하여 관측 값과 절대값 차이(Proportion of absolute differences as large as observed)에 대한 확률 값을 산출할 수 있는데, 유의수준 5%를 기준으로 유의성을 검정할 수 있으며, 이 확률은 일반적인 통계 검정에 사용되는 p -값에 해당된다(Gwak, 2014).

IV. 연구결과

1. 데이터 수집 결과

수집된 데이터는 Table 2와 같이 총 777,975건으로 2000년부터 2017까지 5년 단위로 구분하였기 때문에 2015년부터 2017년까지는 3년 동안의 데이터가 수집되었다.

데이터양을 살펴보면 2000~2004년에 8,173건, 2005~2009년에 120,957건, 2010~2014년에 320,395건, 2015~2017년에 328,450건으로 2010년부터 크게 증가하고 있으며, 2015~2017년은 기간이 짧지만 수집된 데이터양이 가장 많았다.

활용하는 빅데이터 프로그램 텍스트롬은 선정한 기간에 해당하는 문서 중 그림, 사진, 수학기호, 프로그래밍 언어 등 텍스트가 아니어서 읽을 수 없는 결측값을 제외한 유효한 데이터를 랜덤으로 요약수집이 가능하여 총 8,058건이 분석대상으로 선별되었다(Woo, 2017).

2. 텍스트 마이닝 분석 결과

데이터 정제 후 구분한 연차별로 빈도가 높은 키워드를 도출하고, Figure 2의 N-gram을 참고하여 행태와 관련이 있는 키워드를 선정하였다. 그 중 상위 빈도인 10개 키워드는 Table 3과 같다.

연차별로 행태를 살펴보면, 제 I 시기(2000~2004년)에서는 주로 자전거(TF(Frequency Analysis):78), 타기(TF:63), 걷기(TF:61) 등을 하며 운동을 하거나, 공원을 관람하는 것으로 나타났다. 주요 키워드는 '운동(TF:474)'으로 다이어트, 스트레스 해소, 존재감 등의 목적으로 운동을 하는 행태가 가장 많이 언급되고 있다. 또한, 올림픽공원 내 아트홀에서 공연·콘서트 관람도 이 시기의 중요한 행태로 보인다.

제 II 시기(2005~2009년)의 주요 키워드는 '아이스링크(TF:470)'로 2006년에 올림픽공원 내 평화의 광장에 조성되었다. 당시 서울 시내 야외 스케이트장은 올림픽공원을 포함하여 총 3곳으로 스케이트를 탈 수 있으며, 다양한 행사가 개최되어 높은 빈도를 보인 것으로 추측된다. 이 시기에도 자전거를 타고 공원을 관람하는 행태가 언급되지만, 제 I 시기보다 빈도가 적고, 키워드 간의 연관성이 적어 주된 행태를 도출하기 어렵다. N-gram을 살펴보면 연관된 키워드는 몬트리올의 올림픽

Table 2. BLOG analysis data amount

Division	Year	Data amount	Analytical data
I	1th 2000~12th 2004	8,173	2,075
II	1th 2005~12th 2009	120,957	1,984
III	1th 2010~12th 2014	320,395	1,955
IV	1th 2015~12th 2017	328,450	2,044
Total		777,975	8,058

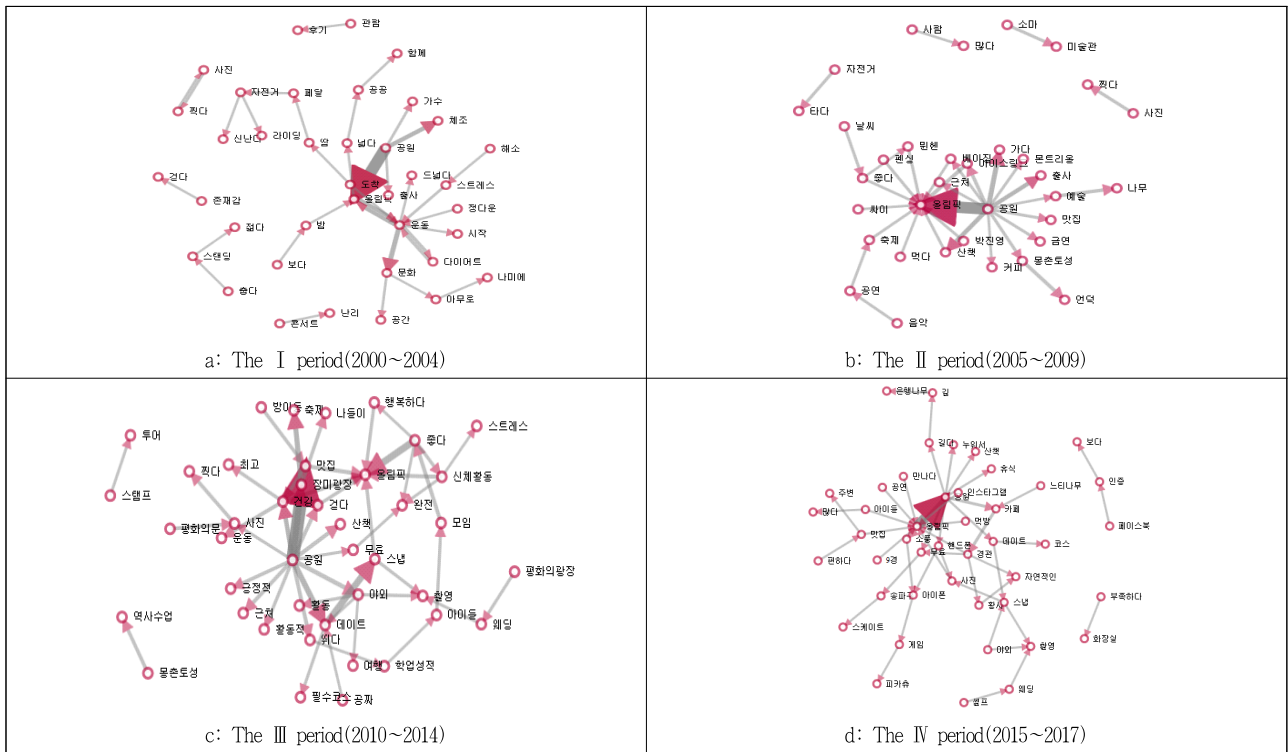


Figure 2. Keyword associations(N-gram)

Table 3. Frequency of behavior-related keywords

The I period(2000~2004)		The II period(2005~2009)		The III period(2010~2014)		The IV period(2015~2017)	
Keyword	Frequency	Keyword	Frequency	Keyword	Frequency	Keyword	Frequency
	Percentage		Percentage		Percentage		Percentage
Exercise	474	Ice rink	470	Famous and delicious tasty food are sold	840	Picnic	1178
	1,756%		1,45%		2,06%		2,58%
Take a picture	226	Picture	313	Date	453	Cafe	831
	0,84%		0,97%		1,11%		1,82%
See	130	Take a walk	219	Physical activity	415	Landscape	708
	0,48%		0,68%		1,02%		1,55%
Bicycle	78	Eat	161	Walk	389	Picture	475
	0,29%		0,50%		0,95%		1,04%
Roller skating	73	Bicycle	138	Picture	364	Show	457
	0,27%		0,43%		0,89%		1,00%
Ride	63	Ride	117	Snapshot	357	Date	370
	0,23%		0,36%		0,88%		0,81%
Walk	61	See	100	Shooting	224	Famous and delicious tasty food are sold	354
	0,23%		0,31%		0,55%		0,78%
Take a walk	60	Show	91	Take a walk	206	Cell phone	330
	0,22%		0,28%		0,51%		0,72%
Eat	60	Landscape	90	Festival	199	Snapshot	325
	0,22%		0,28%		0,49%		0,71%
Preview	56	Art gallery	87	Run	185	Instagram(SNS)	324
	0,21%		0,27%		0,45%		0,71%

공원, 베이징의 올림픽공원, 핀헨의 올림픽공원 등 해외에 조성되어 있는 같은 명칭의 공원이 언급되어 있고, 올림픽공원 내 아트홀에서 공연하는 가수와 관련된 키워드가 등장하여 이 시기만의 특별한 행태는 도출되지 않았다.

제Ⅲ시기(2010~2014년)의 주요 키워드는 '맛집(TF:840)'이다. 이전에도 '먹다'와 관련된 키워드들이 언급되었지만, 이 시기에 크게 급증하여 주요 키워드로 도출되었다. 공원 이용객이 공원 안에서 혹은 주변에서 음식을 먹은 경험을 SNS에 게시하고, 게시한 경험을 다시 새로운 이용객이 경험하여 SNS에 전파하는 과정이 반복되면서 공원에서 먹는 행태가 하나의 경향이 된 것으로 볼 수 있다.

사진(TF:364), 스냅(TF:357), 촬영(TF:224) 키워드도 높은 빈도를 나타내며, 관심 있는 키워드로 나타났다. 웨딩촬영 등 전문적인 스튜디오에서 이루어지던 행태가 야외에서 자신만의 개성과 스타일을 추구하는 형태로 변화되면서 웨딩촬영 장소에 대한 변화가 공원에까지 영향을 미치고 있다고 볼 수 있다. 또한, 공원이용객에 해당하는 아이들과 올림픽공원 내에 소재한 백제시대 토성인 몽촌토성과 역사, 역사수업이 함께 언급되고 있어 교육의 공간으로 활용되고 있음을 파악할 수 있다. 이는 올림픽공원에서 나타나는 특별한 행태로 볼 수 있다.

제Ⅳ시기(2015~2017년)의 주요 키워드는 '카페(TF:831)'이다. 제Ⅲ시기부터 '커피', '카페' 키워드의 빈도가 급증하는데, 현대 사회에서 급속도로 확산된 커피문화가 영향을 미치는 것으로 볼 수 있다. 커피를 마시는 문화가 확산되면서 카페라는 새로운 공간이 생산되고, 사람들은 또 다른 분위기의 편안함을 주는 공간에 대한 요구를 점차 많이 가지고 있는데, 이러한 사회·문화적 변화가 공원의 행태에도 영향을 미치는 것으로 볼 수 있다(Lee, 2011). 또한, SNS(인스타그램, 페이스북 등)관련 키워드와 휴식, 소풍 등의 연관관계로 보아, 제Ⅲ시기와 같이 공원에서 핸드폰을 사용하여 사진을 찍고, SNS에 경험을 게시하는 행태의 비중이 날로 커져가고 있음을 알 수 있다. 또한, 온라인상에서 혹은 현실을 배경으로 사용자가 적극적인 활동으로 경험할 수 있는 다양한 모바일 게임을 하는 것이 공원에서 이루어지는 새로운 여가의 행태로 형성되고 있음을 알 수 있다.

전체 데이터에서 행태와 관련된 키워드를 살펴보면 '길을 걷다(산책)', '사진을 찍다', '자전거(인라인, 킥보드 등)를 타다', '먹다', '공연을 관람하다' 등이 올림픽공원의 공통적인 행태로

나타났다.

빈도분석 결과, 크게 두드러지는 키워드의 특징은 제Ⅰ시기의 주요 키워드인 '운동'은 시간이 지날수록 급감하고, 제Ⅲ시기부터 '맛집', '카페' 등 먹고 마시는 행태와 관련된 키워드가 급증하고 있다. 2000년대 초반에는 공원에서 주로 신체활동, 운동을 하는 사람이 많았지만, 2010년부터 음식을 먹고 커피를 마시기 위해 공원을 찾는 문화가 형성된 것으로 볼 수 있다. 또한, 제Ⅰ시기에는 운동을 하는 자신의 모습과 길에 관심을 갖고 있어 경관에 대한 언급이 되지 않았지만 제Ⅲ시기, 제Ⅳ시기에 웨딩촬영, 스냅촬영, 야외촬영 등을 공원에서 하는 경우가 많아 배경이 되는 경관에 대한 관심이 높아지는 것을 알 수 있다. 즉, 올림픽공원의 경관이 새로운 행태를 형성하는 요인 중 하나로 적용될 수 있다.

특이한 점은 제Ⅳ시기에 SNS(인스타그램, 페이스북 등)관련 키워드와 핸드폰, 게임 등의 키워드가 도출된 것이다. 이러한 현상은 과거 공원에서 활발한 신체활동을 필요로 하는 행태가 주로 이루어졌다면 현재는 비활동적인 행태도 공원에서 이루어지는 중요한 행태적 특성으로 볼 수 있다.

또한, 제Ⅲ시기, 제Ⅳ시기에 '공짜', '무료'의 키워드가 도출되었으며, 연관관계가 높게 나타난 것이다. 제Ⅲ시기부터 데이트, 나들이, 소풍 등의 키워드가 높은 빈도를 보였는데, 공원 이용이 무료라는 요인이 영향을 미치는 것으로 볼 수 있다.

구분된 연차별로 한 문서 안에서 가장 많이 등장하여 가중치가 부여되는 주요 키워드를 검출하기 위하여 TF-IDF를 활용하였는데, Table 4는 도출된 주요 키워드 중 상위 3개이다.

각 연차별로 도출된 주요 키워드를 살펴보면 대부분 빈도가 높은 키워드가 주요 키워드로 도출되는 일치율을 보였는데, 제Ⅳ시기는 약간의 차이를 보여주고 있다. 제Ⅳ시기의 전체 데이터에서는 '소풍'이 가장 높은 빈도를 보였지만, '카페'라는 키워드가 등장한 문서의 건수가 더 많고, 문서 안에서 중요하게 언급되는 키워드로 가중치를 부여하여 이 시기의 주요 키워드로 볼 수 있다.

3. 사회연결망 분석 결과

텍스트 마이닝 결과, 도출된 데이터의 키워드간의 연결망 구조를 파악하기 위하여 밀도 및 중심성 분석과 유의성 검정을 실시하였다. Figure 3은 구분된 연차별로 도출된 키워드의 연

Table 4. Weight calculation(TFIDF)

The I period(2000-2004)		The II period(2005-2009)		The III period(2010~2014)		The IV period(2015~2017)	
Keyword	TF-IDF	Keyword	TF-IDF	Keyword	TF-IDF	Keyword	TF-IDF
Exercise	1107.65	Ice rink	651.56	Famous and delicious tasty food are sold	1,351.93	Cafe	1,337.44
Plaza	609.28	Picture	609.07	Outdoor	1,012.34	Picnic	1,294.17
Take a picture	561.59	PSY(Singer)	548.97	Date	995.34	Landscape	1,139.48

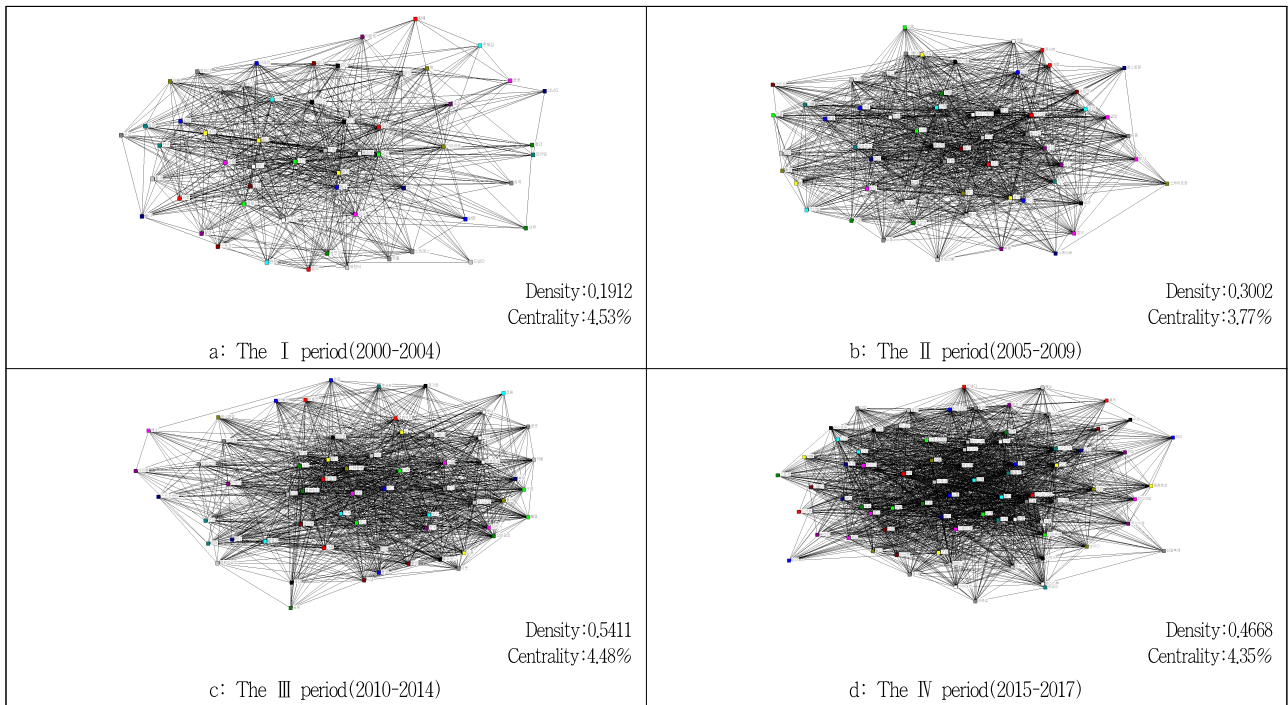


Figure 3. Yearly network

결망 분석을 시각화한 것이다.

구분된 연차별로 분석 결과는 제 I 시기의 밀도는 0.1912, 중심성은 4.53%이고, 제 II 시기의 밀도는 0.3002, 중심성 3.77%, 제 III 시기의 밀도는 0.5411, 중심성 4.48%이며, 제 IV 시기의 밀도는 0.4668, 중심성은 4.35%로 나타났다.

각 키워드간의 연결된 정도를 나타내는 밀도분석 결과, 제 I 시기, 제 II 시기보다 제 III 시기, 제 IV 시기의 연결망구조의 밀도가 높은 것으로 나타났다. 또한, 구분된 연차별로 집중되는 행태가 있는지를 파악할 수 있는 중심성 분석 결과, 제 II 시기를 제외하고 모두 4% 이상의 중심성 값이 나타났다. 즉, 제 I 시기에는 다양한 행태가 도출되지 않고 특정한 행태가 형성된 반면, 제 II 시기, 제 III 시기, 제 IV 시기에는 다양한 행태가 도출되었지만, 제 II 시기의 중심성이 높지 않아 행태를 형성하는 연결 관계가 밀접하지 않은 것으로 볼 수 있다. 제 III 시기, 제 IV 시기에는 밀도 및 중심성 분석 값이 높게 나타나, 다른 시기보다 다양한 행태가 형성되고 있으며, 행태 간 연결 관계도 밀접하여 올림픽공원의 행태 형성에 큰 영향을 미치는 것을 알 수 있다.

부트스트랩 기법을 사용하여 분석에 사용된 데이터의 유의성 검정 결과는 Table 5와 같다. 표집분포 평균과 표준오차 값을 이용한 z-score를 산출한 결과, 구분된 연차별로 z-score보다 절대값이 큰 값이 나올 확률(Proportion of absolute differences as large as observed)은 유의수준 5%에서 모두 유의한 것으로 나타났다.

V. 결론

공원은 인간의 행태를 지원해주는 장소이며, 인간의 행태와 상호보완적이라는 맥락에서 공원 이용자의 행태를 파악하는 연구는 중요하며, 행태를 파악하는 과학적·합리적 접근방법에 대한 연구의 필요성에 주목하였다.

이에 본 연구의 대상을 올림픽 공원으로 선정하고, 텍스트 마이닝과 사회연결망 분석을 활용하여 시간적 추이로 행태를 분석하였다. 도출된 행태의 특징은 다음과 같다.

첫째, 구분된 연차별로 꾸준히 제기되는 키워드는 '길을 걷

Table 5. Significance test

Year	Average bootstrap density	Estimated standard error for density	z-score	Proportion of absolute differences as large as observed
1	0.1977	0.0787	2.4550	0.0188
2	0.3077	0.1440	2.0843	0.0356
3	0.5617	0.2069	2.6159	0.0136
4	0.4783	0.1642	2.8435	0.0092

다(산책)', '사진을 찍다', '자전거(인라인, 키포드 등)를 타다', '먹다', '공연을 관람하다'로 시간적 추이와 무관하게 올림픽공원에서 행해지는 공통적인 행태로 나타났다.

둘째, 구분된 연차별로 도출된 행태의 변화는 2010년부터 공원에서 새로운 행태가 나타났고, 다양성이 증가한 것으로 나타났다. 2000~2004년에는 운동과 같은 적극적인 신체활동을 하는 행태가 주를 이루었지만, 2010년부터는 음식을 먹고 커피를 마시는 행태가 집중되고, 사진, 스냅, 촬영 등의 행태와 핸드폰, SNS, 게임을 하는 행태가 나타났다. 즉, 이전에는 적극적이고 활동적인 행태를 중심으로 이루어졌다면 현재는 소극적인 비활동적 행태도 공원에서 나타나는 새로운 행태로 나타났다.

셋째, 주요 행태의 변화과정을 살펴보았을 때 여러 가지 시대 여건의 변화는 공원 이용자의 행태에 영향을 미치는 것을 알 수 있다. 인터넷 발달 등의 변화는 공원에서 핸드폰을 사용하고, 게임을 하는 등의 비활동적인 행태를 증가시키고, 자신의 개성과 스타일을 추구하는 문화는 웨딩, 사진촬영의 행태를 지속적으로 증가시키며, 이에 경관에 대한 관심도 높아지는 것으로 나타났다. 또한, '맛집', '커피', '카페'가 여가, 휴식의 활동으로 이어지면서 공원에서 먹고 마시는 행태가 현재 공원에서 집중되는 중요한 행태로 나타났다. 이러한 행태의 변화는 공원이 무료로 이용이 가능한 점도 영향을 미치는 것으로 나타났다.

넷째, 올림픽공원에서 나타나는 특별한 행태는 공원 내 공연장과 아이스링크의 활성화로 공연관람과 스케이트를 타는 행태와 몽촌토성이 위치하고 있어 교육의 공간으로 활용되는 행태의 특성을 보이고 있다.

공원의 행태는 시간의 흐름에 따라 여가, 휴식활동에 대한 인식이 적극적인 활동에서 소극적인 활동으로 변화하고 있으며, 이는 인터넷 발달 등의 현대적 사회변화 현상을 고려할 때 당연한 행태적 변화라고 볼 수 있다.

결론적으로 공원은 본래의 의도나 목적에 의해 행태가 이루어지기도 하지만, 사회·문화적 등 시대의 변화에 영향을 많이 받으며, 사람들의 라이프 스타일에 가치를 부여하는 공간으로 볼 수 있을 것이다.

올림픽공원의 행태에 영향을 미치는 요인과 기회요소를 파악하여 이용자에 적합한 공간을 제공하는 공원의 환경에 대한 고민이 필요하다. 예를 들어 먹고 마시는 행태가 집중되고 있으므로 벤치 등 앉을 수 있는 시설 등의 환경을 조성해야 한다. 결혼문화에 대한 의식이 변화되면서 새로운 웨딩촬영 시설의 방안을 제시하고, 몽촌토성은 공원의 다른 공간과 연계해 디테일한 프로그램 등을 제공하여 사람들이 새로운 행태를 만들어 내는 공간으로 변화 등에 대한 고려를 해볼 수 있다.

앞으로도 사회의 여러 가지 여건의 변화는 공원에서 사람들의 행태에 큰 영향을 미칠 것이다. 그러므로 도시공원에서의 행태적 변화 추이를 지속적으로 파악하여 향후 도시공원 재생

등을 통한 공원 환경의 재조성에 반영되어야 할 것이다.

본 연구의 기초자료로 활용한 빅데이터 기반의 자료는 사생활보호 등의 문제로 데이터를 제공하는 사람의 특성 등 다양한 요인에 대한 파악이 어려워 이용자에 대한 정확한 측정이 불가능하다는 한계점을 지닌다. 또한, 도출된 키워드를 활용하여 결과를 해석하는 텍스트 마이닝 기법은 주관적인 해석이 될 수 있다. 이와 같은 기술적인 한계점을 보완할 수 있는 방법에 대한 모색이 필요하다. 그러나 빅데이터를 활용한 연구방법은 과거의 데이터를 수집할 수 있는 장점이 있고, 장기적인 관점에서 행태를 파악할 수 있는 장점이 있어, 행태분석 연구에서 하나의 수단으로 수행된다면 새로운 행태 및 가치 측정이 가능할 것이다.

References

1. An, H. S.(2017) A Study on the Sentiment Analysis in Fashion Design using Big Data: Focused on Text Mining and Semantic Network Analysis, Ph.D. Dissertation, University of Ewha Womans, Korea.
2. Chang, B. H. and S. H. Nam(2012) Social media and network journalism: Focusing on audiences' evaluations on journalism functions of social media, Korea Regional Communication Research Association 12(4):457-496.
3. Gwak, K. Y(2014) Social Network Analysis, Chungram.
4. Joo, S. H. and Y. H. Kim(2010) The cognition of design concepts for urban parks: The cases of Seoul Forest, Yeouido Park, and Seonyudo Park, Journal of the Korean Institute of Landscape Architecture 38(5): 53-63.
5. Jung, G. H.(2011) A study of foresight method based on textmining And complexity network analysis, KISTEP.
6. Kim, B. S., M. H. Shin, D. K. Lim and I. H. Lee(2013) Motivation and types of social media users in their social network activities, Social Science Research Institute Kyungsoong University 29(2): 105-134.
7. Kim, J. A.(2003) Case Study of an Environmental Design Model for an Urban Park based on the Features to afford Human Behaviors, Master's Thesis, Ewha Womans University, Korea.
8. Kim, S. B.(2015) Blog text analysis about visitors' experience change of Seochon, The Architectural Institute of Korea 31(6): 93-102.
9. Kim, S. E.(2013) A Comparative Study on Park Marketing Strategies by Operating Agencies: The Case of Seoul Forest, World Cup Park and Olympic Park, Master's Thesis, University of Seoul, Korea.
10. Kim, S. H. and R. S. Chang(2010) The study on the research trend of social network analysis and the its applicability to information science, Korea Society for Information Management 27(4): 71-87.
11. Lee, J. H. and H. S. Chung(2014) Producing and consuming landscape images via social media blogs the case of Seochon, Seoul-. Journal of the Korean Urban Geographical Society 17(3): 123-140.
12. Lee, J. H. and Y. S. Kim(1993) A study on the use-type in national park, Journal of Korean Institute of Traditional Landscape Architecture 11(1): 51-68.
13. Lee, S. S.(2014) Network Analysis Methods, Nonhyung.
14. Maruta, H.(2011) Theory of Location in the Information Society, Simsan.
15. Seo, J. A.(2016) Analyzing the Destination Image of Daegu from

Online Content through Social Network Analysis, Ph.D. Dissertation, University of Keimyung, Korea.

16. Woo, K. S. and J. H. Suh(2017) Urban landscape image study by text mining and factor analysis: Focused on Lotte World Tower.

Journal of the Korean Institute of Landscape Architecture 45(4): 104-117.

17. <https://blog.naver.com/me0816/10189925806>
18. <https://blog.naver.com/phantasma86/220145264711>

Received : 29 January, 2018

Revised : 22 February, 2018 (1st)

13 March, 2018 (2nd)

Accepted : 13 March, 2018

3인익명 심사필