

서울시 기록물을 통해 본 공원 정책의 변화 경향*

- 텍스트 마이닝 분석 기법을 활용하여 -

심지수**·이명준***

An Analysis of Changes in Park Policies by the Seoul Metropolitan Government Using the Text-Mining Method*

Jisoo Sim**·Myeong-Jun Lee***

요약: 이 연구는 서울기록원의 공원 관련 행정기록물을 토대로 서울시의 공원 정책의 변화를 살펴보았다. 서울기록원의 공공기록물은 일종의 빅데이터(big data)로, 이 연구는 최근 다양한 분야에서 부상하고 있는 빅데이터 분석의 하나인 텍스트 마이닝 기법을 사용했다. 구체적으로, 기록물의 제목을 대상으로 주제어 분석과 바이그램 분석을 실시해, 공원 관련 기록물이 생성된 시기의 주제를 파악하고 기록물 제목에 포함된 단어 사이의 관계를 분석해 공원 행정의 특성을 파악했다. 연구 결과, 서울시 공원 정책은 시간의 흐름에 따라 변화하는 양상을 나타냈다. 시기별로 공원 정책의 특성이 뚜렷하게 구별되지는 않았지만, 대체로 공원의 계획과 조성에서 시민 참여와 관리로 그 초점이 변화하는 현상을 확인할 수 있었다. 공원 관련 기록물은 빅데이터로서 과거의 공원 행정과 정책을 살펴볼 수 있는 사료이면서 동시에 앞으로의 공원 정책 제안과 수립을 위한 객관적인 참고 자료로 유용하게 활용될 가능성을 지닌다.

주제어: 공원 정책, 빅데이터 분석, 공공기록물, 서울기록원

ABSTRACT: This study examined historical changes in park policies as documented in the Seoul Metropolitan Archives. Big Data of public documentation regarding park policies were analyzed using text-mining techniques. Specifically, topic and bigram analyses of separate texts of the document titles were conducted to find out main issues regarding public park policies and their specific characteristics. The results revealed that park policies of the Seoul Metropolitan Government did not considerably change over time. However, although the policy characteristics were not distinct across specific periods, there was a shift in focus from planning and construction of public parks to their management and citizen participation. Analysis of Big Data on park policy documentation can provide important materials for historical research and serve as an objective reference for future park policy proposals, implementation, and decisions

KeyWords: Park Policies, Big Data Analytics, Public Record, Seoul Metropolitan Archives

* 이 연구는 한경대학교 2020년도 학술연구조성비의 지원으로 이루어졌습니다.

** 국토연구원 도시연구본부 부연구위원(Associate Research Fellow, Urban Research Division, Korea Research Institute for Human Settlements)

*** 한경대학교 식물자원조경학부 조경학전공 조교수(Assistant Professor, Department of Landscape Architecture, Hankyong National Univ.), 교신저자(E-Mail: june2@hknu.ac.kr, 031-670-5214)

I. 서론

1. 연구배경과 목적

공공기록물을 일반 시민이 접근할 수 있게 된 지도 20년을 넘겼다. 1996년 국민의 알 권리 확대와 국정 운영의 투명성 향상을 위해 「공공기관의 정보공개에 관한 법률(‘정보공개법’)」이 제정, 공포되고 1998년 1월 1일부터 본격적으로 시행되었다. 이에 따라 국가기관이나 지방자치단체 등을 포함하는 공공기관이 생산한 다양한 정보를 일반 시민들이 정보공개 청구를 통해 쉽게 접근할 수 있게 되었다. 이러한 정보공개는 공공기록물의 생산과 관리 체계가 마련되어야 가능하다. 1999년, 「공공기관의 기록물 관리에 관한 법률」이 지정되어 공공기관의 생산기록의 관리에 대한 법적 토대가 마련되었고, 2007년 「공공기록물 관리에 관한 법률」로 명칭이 개정되는 등 전자 기록의 포함과 기록물 관리의 표준화, 기록물 공개와 열람의 확대 등 공공기록물의 관리 시스템이 발전을 거듭하고 있다. 공공기록물이 체계적으로 관리되어 시민에게 공개되면서 시민이 정책에 참여하는 기회도 점차 늘어나는 등 긍정적인 효과가 높아지고 있다.

기록물에 대한 관심이 사회 전반적으로 증가하면서 이러한 기록물과 제공처를 지칭하는 아카이브(Archive)도 많이 생성되고 있다. 아카이브는 “장소, 기관 혹은 집단에 관한 정보를 제공하는 역사적 자료와 기록물의 컬렉션”이자, 그러한 “자료와 기록의 소장처”도 지칭한다(English Oxford Living Dictionaries). 공공기록물의 대표적인 아카이브는 국가기록원, 서울기록원 등이 있으며, 도서관과

미술관 등이 대중에게 익숙한 아카이브이다. 최근에는 전문적인 아카이브가 아닌 기관과 단체에서도 소장 자료의 관리, 대중 홍보를 위해 대안적인 주제를 가지고 아카이브를 만드는 등 다양한 규모와 특성을 지닌 아카이브가 많이 만들어지고 있는 추세이다(이명준 외, 2019). 유광흠 외(2008)가 지적하듯이 아카이브는 후세를 위해 선별된 기록물로서 당대의 사회, 정치, 문화적 상황을 보여주는 중요한 자료이다. 특히 공공기록물을 소장하는 아카이브는 국가 및 해당 지자체와 기관의 역사를 고스란히 간직한다는 점에서 역사 연구의 훌륭한 자료가 된다.

이 연구는 서울기록원의 공원 관련 행정기록물을 토대로 서울시의 공원 정책의 변화를 살펴보고자 한다. 서울기록원에 오랜 기간 동안 축적된 공원에 관한 공공기록물은 일종의 빅데이터(big data)로서 기록물이 생산된 시점의 공원 관련 정책의 경향과 변화 과정을 객관적으로 파악할 수 있는 중요한 연구 대상이 된다. 한국전쟁 이후 우리나라의 행정 기관이 공식적으로 문서를 생산하기 시작한 이래로 공원 관련 정책이 어떻게 변화했는지를 추적하고자 한다.

2. 연구범위와 방법

연구의 공간적 범위는 서울시로 한정한다. 첫째, 서울시는 우리나라의 도읍으로 지정된 지 600여년이 지난 역사 도시이자 수도로, 우리나라를 대표하는 도시이다. 둘째, 서울시는 우리나라의 공원 정책을 선도하는 지자체로, 역사적으로 공원 관련 정책 트렌드를 이끌어왔다. 셋째, 서울시의 정책 연구는 전국의 지자체로 연구 범위를 확장하기에

앞서 연구 방법을 검정하기에 적합하기 때문이다.

연구의 내용적 범위로는 서울시에서 생산해 보존하고 있는 기록물 중 ‘공원’이라는 단어를 제목으로 포함한 문서로 한정한다. 공원 관련 정책을 살펴볼 때 ‘공원’이라는 특정 단어 외에 ‘녹지’, ‘나무’, ‘가로수’, ‘광장’ 등 다양한 용어가 공원 정책과 관련되어 사용되었을 가능성이 있다. 하지만 선행분석¹⁾을 수행한 결과 공원 정책과 가장 관련도가 높은 단어로 도출된 ‘공원’ 단어에 집중해 공원 정책의 흐름을 살펴보았다.

연구 방법으로는 빅데이터 분석 방법 중 하나인 텍스트 마이닝(text-mining)을 활용한다. 지금까지 기록물에 대한 연구가 미진했던 이유 중 하나로 기록물의 방대한 양을 꼽을 수 있다. 수십 년에 걸쳐 생산된 문건을 전수 조사하기 위해서는 막대한 인력과 비용이 필요하기 때문이다. 2009년부터 빠른 속도로 부상한 빅데이터 분석 방법론은 짧은 시간에 많은 양의 데이터 처리를 가능하게 했다(Kambatla et al., 2014). 특히, 텍스트 마이닝은 텍스트로 생산된 정보를 정제하고 분석하여 결과를 얻는 연구 방법으로 사회, 정치, 문화 등 다양한 연구 분야에서 사용되고 있다(Hassani et al., 2020). 이 연구에서는 텍스트 마이닝 분석 방법을 활용해 서울시의 공원 정책 변화 과정을 추적하였다.

II. 선행연구 검토

서울시의 공원 관련 정책 변화 연구는 크게 서울시장의 정책 변화를 다룬 연구(김영재 외, 2012; 임성은, 2013)와 공원 정책과 그것의 변화를 검토한 연구(최용호, 2005; 최용호, 2006; 박울진 2010; 채진해 외, 2014; 고하정, 2020a; 2020b)로 분류할 수 있다. 김영재 외(2012)는 역대 서울시장의 정책 변화를 추적하고 시장별 정책을 비교하여, 시장의 가치관에 따라 서울시 정책의 방향과 정체성이 확립된다고 주장한다. 특히 우리나라에서 역대의 정권 변화와 서울시장과의 밀접한 정치적 성향의 관계성을 지적하며 각 서울시장에 펼친 정책을 비교했다(김영재 외, 2012). 김영재 외(2012)의 연구는 초대 서울시장부터 최근의 민선시장이었든 박원순까지 각 시장의 정책과 그 성과를 정리했다는 점에서 의미가 있다. 한편, 임성은(2013)의 연구는 민선시장 임기 중 나타난 환경 정책의 변화에 초점을 맞췄다. 정책 변동을 일으키는 요인인 시장 교체로 인해 그간의 환경 정책이 어떠한 변화를 보였는지를 각 민선시장의 시정목표와 환경 분야의 계획, 편제, 조직 등을 비교해 밝혔다(임성은, 2013). 특히 환경 분야에 투입된 예산 비교를 통해 환경 정책의 변화를 분석했다는 점에 의의가 있다.

공원 정책과 변화에 관한 연구는 공원 정책 자체에 관한 연구와 서울시 공원 정책의 변화를 다룬 연구로 재분류할 수 있다. 첫째, 공원 정책에 관한 연구 중 채진해 외(2014)의 연구는 우리나라 도시공원의 시작과 도시기본계획에 대해 정리하고 서울시

1) 선행분석은 2000년부터 2001년까지 푸른도시국이 생산한 문건을 대상으로 진행하였으며, 전체 생산문서에서 ‘공원’이 가장 많이 언급되어 ‘공원’이 포함된 문서로 연구의 내용적 범위를 한정하였다.

〈표 1〉 공원 정책의 시대별 구분

최용호(2005, 2006)		고하정(2020a)	
1881~1910	개항기	해당 없음	
1910~1945	일제강점기		
1945~1961	제1, 2공화국		
1961~1979	제3, 4공화국		
1979~1993	제5, 6공화국		
1993~현재	지방자치시대	해당 없음	
1996~2002	민선 1~2기	1995~1998	민선 1기
2002~2006	민선 3기	1998~2002	민선 2기
해당 없음		2003~2006	민선 3기
		2007~2011	민선 4~5기
		2012~2019	민선 5~7기

공원녹지기본계획을 해외 도시의 것과 비교하여 시사점을 제시했다. 이 논문에 따르면 우리나라는 1967년 공원법 제정에 따라 도시공원 개념이 형성되었고, 1981년부터 도시기본계획이 법제화되면서 공원녹지계획의 법적 근거가 마련되었다(채진해 외, 2014). 둘째, 서울시의 공원 정책 변화에 관한 대표적인 연구의 하나인 최용호(2005)의 논문은 서울시 녹지정책의 변화를 개항기, 일제강점기, 공화국, 지방자치시대로 구분하여 통시적 변화를 검토했다(최용호, 2005). 이후 최용호(2006)는 서울시 공원 녹지 정책을 민선시장별로 구분해 살펴보았다. 최용호(2005; 2006)의 연구가 갖는 의의는 서울시 공원녹지 정책을 통시적으로 본 최초의 연구라는 점도 있지만, 공원녹지 정책과 관련하여 시대를 구분하고 시대별 정책을 정리하여 그에 따른 공원관리조직의 변화를 비교했다는 점에 큰 의의가 있다. 박울진(2010)과 고하정(2020a)은 정량적인 방법의 분석, 즉 관련 통계와 공원조성예산 등을 중심으로 공원녹지 정책의 변화 과정을 살펴보았다. 박울진(2010)은 서울시 도시공원녹지 관련 10년 동안의 통계를 바탕으로 공원녹지 정책의 변화를 살펴보고

시사점을 도출했다. 구체적으로, 이 연구는 공원녹지 정책지표를 공원조성면적, 1인당 생활권 면적, 가로 녹지율 등의 지표를 고려해 도출해 비교하고, 공원녹지 정책을 주요 조직부서와 구성원 수로 검토했다. 고하정(2020a)의 연구는 서울시에서 민선시장 선출 이후 25년간 서울시 공원조성예산이 어떻게 변동했는지 분석했다. 그는 민선시장의 임기기간으로 시대를 구분하고 시기별 예산편성현황과 그에 따른 공원의 유형을 구분했다(고하정, 2020a). 이러한 연구는 정량적인 지표를 활용하여 서울시의 공원녹지 정책 변화를 분석했다는 점에서 이전의 연구와 차별화된다. 또한 고하정(2020b)의 연구는 서울시의회 회의록을 서울시장의 임기별로 나누어 텍스트 마이닝 기법을 통해 분석하여 도시공원의 쟁점 변화를 살펴봤다.

이처럼, 선행연구 대부분은 시장과 정책 간의 밀접한 관계성에 초점을 맞추며 시장 임기별로 시대를 구분해 정책의 특성을 살펴보는 경향이 있다. 또한, 정성적인 분석뿐만 아니라 공원 정책 관련 조직의 변화, 구성원의 수, 예산변화, 통계 등 정량적인 지표를 통한 연구가 점차 증가하는 추세에 있다는 사실을 확인할 수 있었다. 이 연구는 앞선 연구들과 유사하게 공원 정책의 변화 경향을 추적한다. 다만, 그러한 공원 정책의 변화를 가늠하는 방법으로 공원 관련 공공기록물의 빅데이터 분석을 시도한다는 점에서 선행연구와 차별성을 지닌다. 고하정의 연구(2020b)가 서울시장 임기에 따라 관련 정책이 변화한다고 전제하여 1995년부터 서울시의회 회의록을 시장 임기별로 나누어 분석했다면, 이 연구는 공원 관련 행정기록물이 생산된 1950년대부터의 자료를 바탕으로 하였으며, 임기별 분류가 아니라 관련 기록물의 양적인 증감에 따라 객관적으로 시기를 분류하여 분석을 진행했다.

III. 연구 과정

1. 연구대상과 기간

이 연구는 서울기록원이 소장 중인 시정기록물을 연구 대상으로 한다. 서울기록원은 2007년 서울특별시 지방기록관리기관 건립 기본계획을 수립한 이래, 2012년 서울 아카이브 건립이 재논의됨에 따라 2013년부터 2015년까지 건립이 본격적으로 추진되어 2019년 정식 개원한 서울의 기록관리 전문기관이다. 서울기록원은 “서울시 행정에 대한 설명책임을 구현하는 기관”, “기록 문화유산 기관”이라는 취지에 걸맞게 시정에 관한 공공기록물과 함께 시민기록물을 함께 아카이빙하고 있다([https:// archives.seoul.go.kr/](https://archives.seoul.go.kr/)). 이 연구는 서울기록원이 보유한 기록물 중 ‘공원’이라는 키워드를 포함한 문서 제목의 리스트를 정보공개청구(www.open.go.kr)를 거쳐 서울기록원 측에게 전달받았다. ‘공원’ 단어를 포함하는 기록물(건)은 1926년부터 2017년까지 생산된 총 424,738건이었다(〈표 2〉 참조). 2017년 이후의 데이터는 아직 서울기록원 측으로 이관되지 않은 상태라 제외했다. 또한 우리나라의 행정 기관이 공식 행정기록을 생산하기 이전이라 ‘공원’ 관련 문건이 거의 없는 1954년을 전후로 하는 한국전쟁 이전 시기의 문건은 연구 기간의 연속성을 이유로 연구에서 제외했다. 구체적으로, 1926년 10건, 1939년 17건, 발생연도 미상이나 1900년대에 작성된 1건의 기록물은 분석에서 제외되었다. 또한, 2013년의 경우 2012년 53,838건 대비 2,122건으로 매우 적은 양을 보이고 2014년은 56건으로 기록물 목록화 과정에서

오류가 발생한 것으로 판단하여 2012년을 연구의 종점으로 정했다. 이에 따라 연구 기간은 1954년부터 2012년으로 한정하였고, 이 기간 생산된 기록물이자 분석 대상은 총 422,296건이다.

〈표 2〉 공원 관련 공공기록물 생산연도별 문건 수(건)

연도	문건 수	연도	문건 수	연도	문건 수
1926	10	1972	228	1992	490
1939	17	1973	184	1993	405
1954	4	1974	117	1994	348
1955	4	1975	170	1995	335
1956	4	1976	200	1996	386
1957	3	1977	111	1997	688
1958	38	1978	182	1998	443
1959	34	1979	188	1999	385
1960	35	1980	228	2001	4,495
1961	58	1981	186	2002	9,853
1962	55	1982	339	2003	13,137
1963	29	1983	368	2004	19,174
1964	32	1984	457	2005	23,284
1965	22	1985	493	2006	26,066
1966	56	1986	445	2007	31,911
1967	85	1987	452	2008	46,531
1968	124	1988	370	2009	59,477
1969	95	1989	664	2010	55,143
1970	61	1990	401	2011	59,022
1971	142	1991	421	2012	63,838

2. 텍스트 마이닝

이 연구는 서울시의 공원 기록물을 활용해 공원 정책의 변화를 살펴보기 위해 빅데이터 분석의 하나인 텍스트 마이닝을 사용한다. 서울기록원 공원기록물 리스트는 문서의 제목으로 제공되어 있는데, 이러한 문서 제목에서의 특정 경향을 읽어내는 데 텍스트 마이닝은 적합한 도구이다. 기록물 제목을

구성하는 많은 양의 텍스트 사이에서 어떤 질서를 발견하고 그 질서를 해석해 서울시 공원 정책의 변화를 살펴보았다.

연구자료인 공공기록물은 서울기록원이 보유한 기록물 목록을 엑셀 파일의 형태로 전달받았고, Rstudio 환경에서 오픈소스인 프로그래밍 언어 R을 사용하여 연구를 진행하였다. R은 오픈소스로서 무료로 활용할 수 있으며 데이터프레임 사용을 통해 통계를 비롯한 다양한 분석이 가능하다.

텍스트 마이닝은 텍스트 자료를 분석하는 방법의 하나로, 자연어를 처리해서 데이터 마이닝, 통계 등의 방법으로 유의미한 정보를 얻어내는 방법론이다(He et al., 2013). 텍스트 마이닝은 일반적으로 빅데이터 분석 방법인 자연 언어 처리(Natural Language Processing), 기계 학습(Machine Learning), 정보 검색(Information Retrieval), 데이터 마이닝(Data Mining) 기법들을 사용한다. Hu and Liu(2012)는 텍스트 마이닝 방법을 텍스트 전처리(text pre-processing), 단어 표현(text representation), 텍스트 분석(text analytics)의 세 가지 단계로 구분한다. 첫째, 텍스트 전처리 과정은 텍스트 자료 중 불필요한 텍스트 정보를 삭제 및 변경하여 텍스트 데이터를 정제하는 단계이다. 예를 들어, 과거형 언어를 현재형으로 바꾼다거나(‘갔었다’→‘가다’), 오타자를 교정하거나(‘왔다’→‘왔다’) 중복된 단어를 삭제하는(‘공공원’→‘공원’) 등의 작업을 포함한다. 또한, 분석에 불필요한 불용어(stopwords)를 제외하는 과정도 여기에 포함된다(Hu and Liu, 2012).

둘째, 단어 표현 과정은 문장이나 긴 표현을 단어 단위로 구분하여 분석하는 단계이다. 예를 들어, ‘나는 공원에 갔다’라는 문장은 우선 텍스트 전

처리 과정을 거치면서, ‘나 공원 가다’라는 문장으로 변환되는데, 이 문장을 단어 단위인 ‘나’, ‘공원’, ‘가다’로 구분하는 것이 단어 표현 과정에서 수행하는 작업이다. 이때 사용되는 방법은 엔그램(N-gram) 언어 모델 중 일부인 바이그램(bigram) 분석인데, 이를 통해 단어 단위로 구분된 텍스트에서 연속된 두 단어의 관계를 도출할 수 있다(Broder et al., 1997). 예컨대, 위의 문장을 다시 보면 ‘나’-‘공원’, ‘공원’-‘가다’를 바이그램 분석을 통해 도출할 수 있다. 이러한 관계가 누적되면 단어 사이의 관계성을 분석할 수 있다. 바이그램 분석은 사람들의 행동(Myslin et al., 2013), 감정(Tang et al., 2014), 사회적 관계(Compton et al., 2013) 등 다양한 사회현상을 분석하는 데 근래에 활발히 활용되고 있다.

셋째, 텍스트 분석 과정은 분류(classification), 군집 분석(clustering), 주요사건 확인(event detection), 감정 분석(sentiment analysis), 주제어 분석(topic analysis) 등을 포함한다(Hu and Liu 2012). 이 중 감정 분석은 각 단어가 포함한 긍정적 의미와 부정적 의미에 각각 수치를 부여해 문장이나 글에 적용된 각 수치의 합을 도출해 그것이 긍정표현인지 부정표현인지를 가늠하는 분석 기법이다. Hutto and Gilbert(2014)의 베이더 모델(VADER)이 가장 널리 쓰이는 대표적 감정 분석의 도구이다. 베이더 모델은 짧은 문장이나 글귀의 감정을 분석하는 데 최적화되어 있다(Hutto and Gilbert, 2014). 이외에, 주제 분석은 단어를 정제하고 단어 단위로 구분한 후 각 단어의 출현 빈도를 통해 글귀의 주제를 파악하는 분석을 의미한다(Hu and Liu, 2012).

지금까지 텍스트 마이닝의 가장 기본적인 세

단계와 단계별 분석 방법에 대해 간략히 설명했다. 이 연구에서는 그러한 텍스트 마이닝 방법 중 주로 주제어 분석과 바이그램 분석을 활용해 1) 공원 관련 기록물이 생성된 시기의 주제를 파악하고, 2) 기록물 제목에 포함된 단어 사이의 관계를 분석하여 공원 정책의 특성을 파악한다.

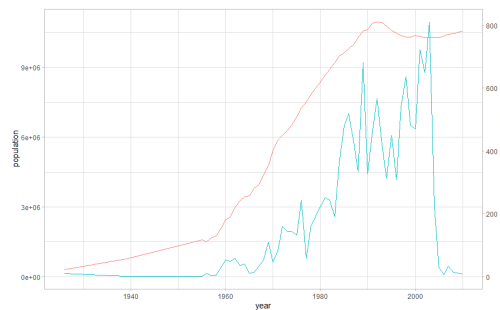
구체적으로, 우선 주제 분석을 통해 각 단어가 사용된 빈도를 계산하고 이를 바탕으로 워드클라우드 방법을 이용해 시각화한다. 이 과정을 통해 시대별 서울시의 공원 정책 방향을 파악하고 빈번하게 언급된 공원을 찾을 수 있다. 이어, 바이그램 분석을 활용해 공원 관련 기록물의 제목에 포함된 연속된 단어들 사이의 관계를 살펴보면서 공원에 관한 기록물이 어떤 이유와 목적으로 생산되는지를 파악해 본다.

IV. 연구 결과

시대별 분석에 앞서 시대구분을 위해 다양한 분석을 진행하였다. 첫째, 서울시 인구의 증가와 기록물의 증가를 비교했다. 그 결과, <그림 1>에서 나타나듯이 서울의 인구수에 비례해 기록물도 양적으로 증가하는 것을 확인할 수 있었다. 2000년대부터 전자문서의 정착에 따른 문서의 증가를 감안하더라도, 인구 증가와 공원 관련 기록물 생산의 증가는 정(+)의 관계가 있다는 사실을 발견할 수 있다. 둘째, 서울시가 생산한 공원 관련 기록물의 주제를 파악하기 위해 전체 기록물을 대상으로 주제어 분석을 실행했다. <그림 2>에서 보이듯, 전체

적으로 ‘조성계획’, ‘조성’, ‘실시계획’, ‘허가’, ‘위원회’와 같이 공원의 정책과 연관된 단어가 주를 이루고, ‘근린공원’, ‘도시’, ‘어린이’와 같이 공원의 유형과 특징과 연관된 단어, 그리고 ‘서울대공원’, ‘한강시민공원’과 같이 특정 공원의 이름이 빈번하게 언급되었다는 사실을 확인할 수 있었다. 셋째, 생산부서별 기록물의 수를 연도별로 시각화한 결과, 한강시민공원사업소, 공원녹지과, 도시계획국이 가장 많은 문서를 생산한 사실을 알 수 있었다(<표 3> 참조).

<그림 1> 서울시 인구변화와 문서 수의 변화



<그림 2> 연구 대상의 워드클라우드 결과



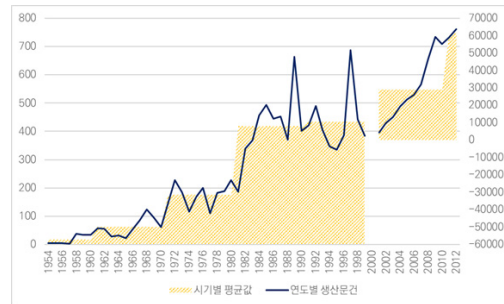
〈표 3〉 공원관련 공공기록물 생산부서별 문건 수(건)

부서명	문건 수	%
한강사업본부	58,727	13.91
푸른도시국	50,062	11.85
녹지사업소	21,836	5.17
한강시민공원사업소	18,950	4.49
동부푸른도시사업소	18,228	4.32
서부푸른도시사업소	17,328	4.10
환경국	17,136	4.06
상수도사업본부	16,590	3.93
월드컵공원관리사업소	13,930	3.30
체육시설관리사업소	12,031	2.85
공원녹지관리사업소	11,463	2.71
남산공원관리사업소	10,609	2.51
서울대공원관리사업소	10,215	2.42
도시계획국	10,047	2.38
서울대공원	8,942	2.12
동부공원녹지사업소	8,849	2.10
중부푸른도시사업소	8,789	2.08
서부공원녹지사업소	8,358	1.98
재무국	5,966	1.41
중부공원녹지사업소	5,904	1.40
공원녹지국	5,702	1.35
도시기반시설본부	5,241	1.24
행정국	3,629	0.86
도시교통본부	3,227	0.76
건설안전본부	3,214	0.76
보존환경연구원	3,103	0.73
경영기획실	2,872	0.68
문화국	2,600	0.62
감사관	2,095	0.50
주택국	2,061	0.49

이 연구에서는 연구의 기간을 십 년 단위로 구분해 분석을 진행했다. 선행연구는 특정 정치적 요인에 따라 민선시장(고하정, 2020a)이나 공화국(최용호, 2005) 등으로 구분한 바 있지만, 이 연구는 그러한 정치적 요인에 따른 정책의 변화를 살펴보는 것이 아니라, 최대한 객관적인 데이터 분석으로 공원의 정책의 흐름의 변화를 파악하기 위해 서울

시장의 변화보다는 기록물 양의 변화를 기준으로 하여 약 십 년 단위로 기간을 구분했다. 왜냐하면, 대략 십 년 단위로 기록물의 수가 크게 증가하는 경향이 확인되었기 때문이다(〈그림 3〉 참조).

〈그림 3〉 연도별 생산문건(좌: 2000년 이전, 우: 2000년 이후)



1. 주제어 분석

1) 1950년대

〈그림 4〉 1950년대 워드클라우드



1950년대부터 1960년까지 중 가장 주목되는 연도는 1958년이다. 1958년 기록물은 도시계획국이 ‘서울 도시계획 공원’에 관한 문건을 다수 생산했다. 같은 해 생산된 총 38건의 문건 중 16건이 ‘서울 도시계획 공원’ 관련 기록물로 당시 도시계획 차원에서 공원이 인식되기 시작하여 조성계획이 이루어졌다고 짐작할 수 있다. 그 외 탑골공원,

남산공원, 사직공원, 풍치공원보안림이 주요 공원 키워드로 언급되었다. 또한 사유지(공원예정지) 사용 허가에 관한 문건이 3건 작성되었다.

1950년대 생산된 기록물 전반의 워드클라우드 분석 결과(〈그림 4〉 참조), ‘효창’, ‘사직’, ‘남산공원’, ‘삼청공원’의 키워드가 빈번하게 언급된 것으로 보아 당시 정책적으로 중요한 공원이었다고 추측해 볼 수 있고, ‘보안림’이라는 키워드는 한국전쟁 이후 황폐해진 산림녹화 및 사방공사를 위해 보안림이 증가했던 당시의 시대상을 고스란히 반영한다. ‘도로공사’, ‘보수공사’ 등의 단어가 공원과 관련되어 언급된 것도 공원이 전후 복구를 위해 하나의 도시기반시설로 인식되어 조성되었다는 당대의 인식을 반영하고 있다.

2) 1960년대

〈그림 5〉 1960년대 워드클라우드



1960년대에도 1950년대와 같이 ‘서울 도시계획 공원’이 가장 많이 언급된 단어로 나타났다(그림 5 참조). 가장 많은 기록물이 생산된 1966년의 문서부터 살펴보면, 당시 문건은 총 56건으로 ‘서울 도시계획 공원’ 관련 문건 22건, 그 외 공원녹지와에서 생산한 효창공원 시설공사 관련 문건 6건, 환경국에서 작성한 ‘공공용지(공원) 이양’ 관련 문건 5건, 국유재산(공원용지) 관리에 관한 문건이 공원

녹지과에서 총 4건 생산되었다. 이 시기 공원용지를 확보하고 국유재산을 공원으로 활용하는 사례가 증가했다는 사실을 알 수 있다. 전반적으로, 시간의 흐름에 따라 ‘서울 도시계획 공원’ 지정 관련 단어의 빈도 수가 증가하고, ‘남산공원’, ‘탐골공원’, ‘야외음악당’, ‘북악공원’, ‘장충단’ 등이 주요 공원으로 언급되었다. ‘설계’라는 단어도 1950년대 이어 1960년대에도 등장했다는 점이 특징이다.

3) 1970년대

1971년 생산 기록물은 총 142건으로 ‘도시계획’ 25건, ‘허가’ 19건으로 다수의 문건이 도시계획상 공원과 공원용지 허가와 관련된 것을 알 수 있다(〈그림 6〉 참조). ‘어린이대공원’ 28건으로, 1971년 개원한 어린이대공원 관련 문서가 다수 생산된 것을 확인할 수 있다. 특정 공원명으로는 ‘효창공원’과 ‘장충공원’이 각각 4건과 7건으로 가장 많이 나타났다. 특히, 효창공원은 1950년대부터 꾸준히 설계 및 보수와 관련해 언급되는 것으로 보아 우리나라 공원의 역사에서 차지하는 비중이 크다는 사실을 짐작할 수 있다. 주목할 점은 1970년대 처음으로 공원 관련 기록물에 ‘관리’라는 단어가 등장한다는 사실이다. 1950~1960년대에 조성된 공원의 관리 문제가 행정상 인식되고 있는 것으로 추측된다.

〈그림 6〉 1970년대 워드클라우드



1982년 기록물은 총 339건으로, 이 중 ‘도시계획’ 90건과 각 공원에 대한 지정 관련 문건이 78건으로 도시계획에 근거해 공원을 지정하는 행정 관련 기록물이 절반 정도를 차지하고 있다(그림 7) 참조). ‘조례’ 13건으로 1980년대 공원 관련 조례가 제정되고 있음을 알 수 있으며, ‘공원용지’ 14건으로 공원으로 지정되는 용지에 대한 언급은 1970년대에 비해 감소했음을 알 수 있다. 공원별 문건으로는 ‘잠실호수’가 11건으로 당시 잠실호수를 지정하고 조성하는 데 많은 행정적 관심이 있었음을 추측할 수 있다.

[illegible]

1989년 문건은 총 664건으로 주택국에서 생산한 ‘퐁치지구 해제’ 14건으로 ‘퐁치지구’가 해제되는 시기이다. 또한, ‘도시계획’ 197건으로 공원에 대한 도시계획 변경이 많은 시기임을 알 수 있다. 주목할 점은 ‘한강’이 언급된 기록물이 300건으로 해당 기간 기록물 중 절반 정도를 차지하고, 또한 전체 문서 중 ‘한강관리사업소’가 생산한 문서가 290건으로 절반 정도를 차지하는 것으로 보아 1989년에는 한강에 대한 공원 조성과 관리가 중요한 관심사였음을 알 수 있다. 한강에 홍수방제시설이 건설되면서 공원으로 조성되었던 1980년대의

1990년대는 이전 시기와 비교했을 때 특별한 차이점을 발견할 수는 없었다(그림 8) 참조). ‘도시계획시설’, ‘조성계획’, ‘변경’, ‘결정’ 등이 이전 시기와 유사하게 빈번하게 언급되고 있다는 점에서 공원 조성의 행정적 절차가 표준화되는 시기라고 추측할 수 있다. 주목할 점은 ‘심의’, ‘위원회’와

같은 단어가 용어의 등장이다. 공원 조성과 관리의 행정에서 전문가와 일반 시민들의 의견을 반영하기 위한 위원회의 공식적인 활동이 기록물에 본격적으로 등장하고 있다.

6) 2000년대 이후

〈그림 9〉 2000년대 워드클라우드



2000년대 들어 특정 공원의 언급 비율이 증가하였다. 2001년 문건은 총 4495건으로 ‘선유도(148회)’ ‘노을(86)’ ‘하늘(24)’ ‘난지(142)’ ‘추모공원(185)’ ‘탐골(206)’ ‘여의도공원(66)’ ‘남산(54)’ 등 공원별 문건이 양적으로 증가했다. ‘용산공원’과 ‘서울숲’도 언급되기 시작했다. 또한, ‘생태(85)’ ‘근린(350)’ ‘어린이(295)’ 등 공원의 기능이 중요하게 언급되는 양상을 보인다. 실제로, 공원 디자인의 역사에서, 1990년대에는 다양한 기능과 테마 등 독특한 정체성을 지닌 공원이 등장한 시기라고 평가된다. 이와 함께, ‘도시계획’이 730회 빈도수를 보여, 공원의 조성 및 관리가 법률, 행정적으로 도시계획의 일부로 다뤄지는 경향은 계속 관찰되었다. ‘시유’라는 단어를 담은 문건이 총 74건으로 시유지의 공원화 작업이 진행되었으며, ‘위원회’ 287건, ‘협의’ 255건, ‘회의’ 378건, ‘회의’ 191건으로 공원의 조성 관리에서 타당한 방식이 아니라 전문

가와 일반인의 의견 수렴과 피드백이 중요해졌다는 사실을 알 수 있다(〈그림 9〉 참조). 또한, 전자 문건이 보편화되면서 지형도를 포함한 시각 자료도 아카이빙이 본격화되는 시기라고 판단할 수 있다. ‘생태’와 더불어 ‘환경영향평가’ 단어의 등장은 환경과 생태에 관한 관심이 높아졌다는 사실을 보여준다.

주제어 분석과 워드클라우드를 이용한 시각화 결과, 50년대는 전후 복구를 위한 산림녹화로서의 공원, 1960년대부터 1980년대는 공원 조성이 본격화되어 공원이 팽창하는 시기, 1990년부터 2000년대는 공원 기능의 다양화와 생태에 대한 관심, 위원회 활동 등 전문가와 시민의 참여, 공원의 관리가 점차 증가하는 경향을 확인할 수 있었다.

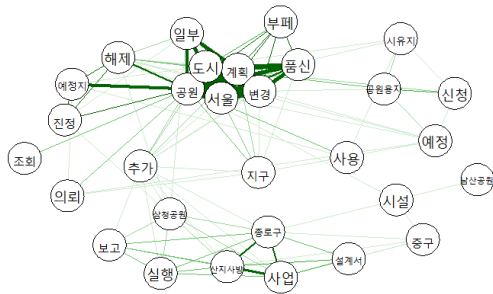
2. 바이그램 분석

기록물 제목을 대상으로 1950년대부터 2021년까지를 십년 단위로 구분하여 바이그램 분석을 진행하였다. 바이그램 분석은 단어와 단어의 연결 빈도를 시각화하는 것으로 이 분석을 통해 각 시기별 공원 행정의 특성을 추정하였다. 각 다이어그램은 해당 시기의 문건 제목을 바이그램으로 분석하여 개별 단어 사이의 관계를 도출한 것이다. 단어와 단어 사이의 거리가 짧을수록 밀접한 관계를 지니며, 단어와 단어를 잇는 선이 굵을수록 단어와 단어가 함께 언급된 횟수가 많다는 것을 보여준다.

〈그림 10〉은 1950년대 공원과 관련된 문건 제목을 단어 간 연결 빈도에 따라 시각화한 것이다. 이에 따르면 서울-공원-도시-계획-일부-변경-예정지가 서로 연관이 높은 것으로 시각화되었다. 이외, 종로구-산지사방-사업 간의 관계도 높았다.

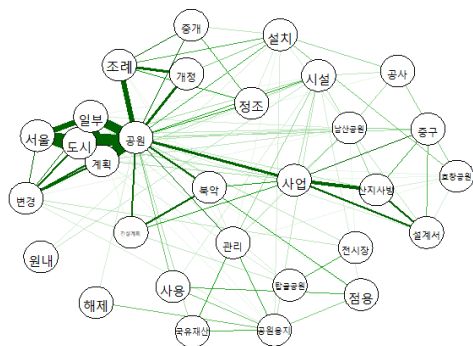
이는 해당 시기 서울의 도시공원계획이 예정되거나 일부 변경되는 것에 대한 문건과 종로구 산지사방 사업이 주로 진행되었음을 보여준다.

〈그림 10〉 1950년대 바이그램



1960년대 공원과 관련된 문건의 바이그림 분석 결과에 따르면, 해당 시기에는 서울 도시공원의 계획 지정과 일부 변경이 지속되었으며 특히 조례가 신설되었음을 알 수 있다(그림 11) 참조). 조례의 제정과 개정이 주된 문건으로 나타났다. 그 외 국유재산과 공원용지, 국유재산 사용 등도 이전 시기와 다른 점이다.

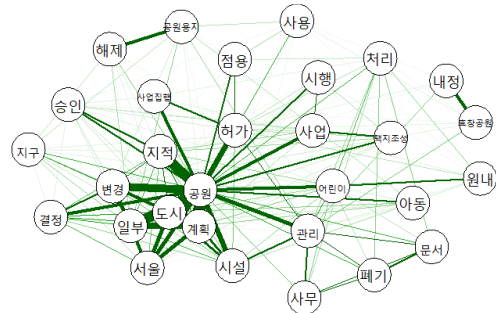
〈그림 11〉 1960년대 바이그램



〈그림 12〉는 1970년대의 바이그램 분석 결과를 시각화한 것으로 이 시기에도 서울-도시-공원-계획은 꾸준히 많이 언급되는 단어의 관계로 나타

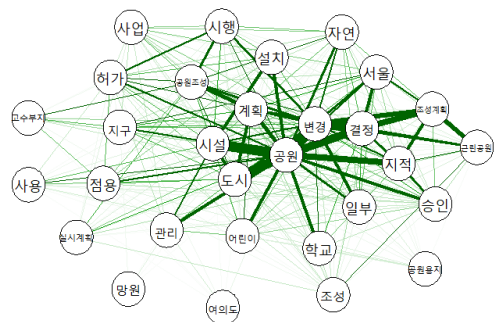
났다. 그 외 다른 시기와 구분되는 점으로는 사업집행-허가-점용-사용 간의 관계가 돋보이며 효창공원-내정-관리의 관계도 주로 언급된 단어 관계로 나타났다. 또한 공원용지-해제도 연관도가 높은 것으로 나타난 것으로 보아, 해당 시기에 공원용지로 지정된 용지가 다수 해제되었음을 추측할 수 있다.

〈그림 12〉 1970년대 바이그램

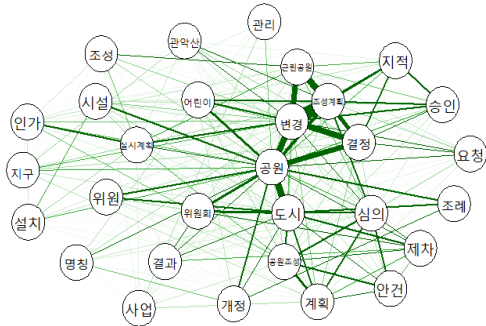


〈그림 13〉은 1980년대의 바이그램 분석 결과로 공원-시설-설치-변경-공원조성-결정 등이 짚은 빈도로 함께 언급되었음을 보여준다. 특히 근린공원-조성계획-지적-승인 간의 관계가 다른 시기와 달리 해당 시기에 주로 언급된 것으로 보아 1980년대는 근린공원의 조성계획이 주로 승인된 것을 알 수 있다.

〈그림 13〉 1980년대 바이그램

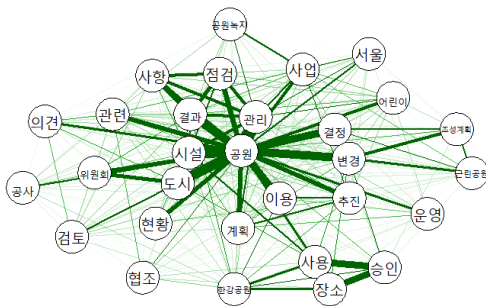


〈그림 14〉 1990년대 바이그램



1990년대는 공원-변경-근린공원-조성계획-결정-승인 등이 함께 언급되었다(〈그림 14〉 참조). 또한 심의-안전-위원회-위원-결과 등이 다수 함께 언급된 것으로 보아 1990년대부터 심의위원회의 활동이 시작되었으며 위원회의 심의가 중요한 역할을 한 것을 추측할 수 있다.

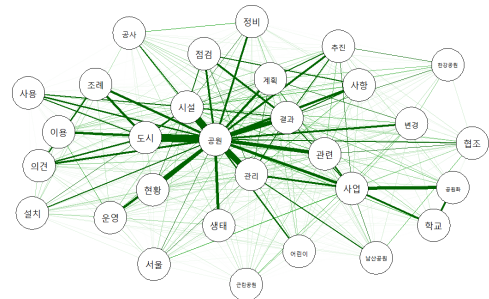
〈그림 15〉 2000년대 초반(2000~2004년) 바이그램



최근에 해당하는 2000년대는 더 자세한 분석을 위해 5년 단위로 구분하여 분석하였다. 2000년부터 2004년은 다른 시기와 마찬가지로 공원-계획-결정-변경 등 공원 계획의 승인과 변경과 관련된 문건이 많은 것을 알 수 있었고, 이전 1990년대에 이어서 위원회-검토-의견 등이 주요한 문건임을 추측할 수 있었다(그림 15) 참조). 또한 한강공원의 사용과 승인에 관련한 문건이 다수 등장함을

통해 해당 시기에 진행된 한강공원 관련 정책을 파악할 수 있었다.

〈그림 16〉 2000년대 후반(2005~2010년) 바이그램



〈그림 16〉은 2005년부터 2010년까지의 문건에 대한 바이그램 시각화 결과로 이전의 시기와 마찬가지로 공원-계획-시설과 같은 단어가 가장 빈번하게 언급되었다. 다른 시기와 구분할 수 있는 점은 남산공원이 새로 등장했고, 남산공원은 특히 관리, 사업과 연관성이 높았다. 또한 생태라는 단어가 공원과 직접적인 연관을 가진 단어로 부상했고, 특히 생태-공원-설치 간의 관계가 돋보였다.

바이그램 분석으로 단어 간 연결 관계와 연결 빈도를 시각화하였고, 이를 통해 시기별 주요 문건의 내용과 해당 시기의 공원 관련 정책의 주요점을 파악하였다. 그 결과, 공원 정책에 관한 기록물은 1950년대부터 2010년까지 주로 도시공원계획의 승인, 변경 등에 관련된 내용의 문건이 가장 많이 생산된 것으로 보아 공원 관련 기록물에서 가장 중요한 생산물은 공원 계획임을 알 수 있었다. 시기별로 공원 정책의 특징을 정리하면, 1950년대에는 산지사방 사업이 주로 언급되었고, 1960년대에는 국유재산 해제를 통한 공원용지 확보가 진행되었음을 알 수 있었다. 1970년대에는 공원용지가 해제되고 어린이에 대한 중요성이 부상했고, 1980년

대에는 자연-공원, 근린공원이 처음으로 등장했다. 1990년대에는 공원 심의와 관련된 위원회의 활동이 시작되었고, 2000년대 초반은 한강공원이, 2000년대 후반은 남산공원이 공원 행정 문건에서 중요한 장소로 언급되었다.

V. 결론

지금까지 서울시 공원 정책의 변화를 서울시가 생산해 보관하고 있는 행정기록물을 중심으로 살펴보았다. 이 연구는 방대한 양의 공원 관련 기록물의 제목의 텍스트를 대상으로 빅데이터 분석을 활용해 그간의 서울시 공원 정책의 경향을 통시적으로 살펴보았다는 점에서 의의가 있다. 또한, 기존에 진행된 서울시 공원 정책의 변화에 관한 연구에서 언급했던 일반적인 공원 정책의 특징을 이 연구의 결과에서도 어느 정도 확인할 수 있었다. 선행연구에서 도출된 공원 정책의 변화를 공공기록물 빅데이터 분석을 통해서도 유사하게 발견할 수 있었다는 점은 빅데이터 분석이 추후 공공기록물을 활용한 정책 경향을 살펴보는 데 유효한 방법론이 될 가능성을 지닌다.

서울시 공원 정책은 시간의 흐름에 따라 변화하는 양상을 나타냈다. 1950년대는 도시공원이 우리나라의 행정력으로 조성되기 시작하는 태동기로서, 이때 공원은 한국전쟁의 상흔으로 황폐화된 국토를 녹화하기 위한 도시기반시설 사업으로 다뤄지는 경향이 있었다. 이어지는 1960년대와 1970년대 역시 이전 시기와 비슷하게 공원이 양적으로 증가하는

시기로 사유지와 국유지의 공원화 작업이 추진되었고, 어린이공원이 주요하게 언급되었다. 1980년대에는 한강에 홍수방제시설이 건설되면서 공원화되고, 서울올림픽을 비롯한 국가적 행사에 대한 대응으로 공원이 조성되기도 했다. 근린공원 등 생활권 공원에 대한 관심이 높아졌으며, 올림픽공원, 서울대공원 등 설계가 디자인한 공원이 조성되었다. 1990년대에 들어 공원 조성의 행정적 절차가 표준화되는 경향을 보이며 많은 수의 공원 관련 문건이 작성되었고, 공원 행정에서 '심의', '위원회'와 같이 전문가와 일반 시민들의 의견을 반영하기 위한 노력이 시작되었다. 2000년대부터는 공원의 기능이 보다 특화되어 독특한 특성을 지닌 공원이 많이 조성되었으며, 특히 생태와 자연에 대한 관심이 생겨났다는 사실을 알 수 있었다. 또한, 공원의 조성과 관리에서 전문가와 일반인의 의견 수렴과 피드백이 보다 중요해지고, 전자문건의 보편화에 따른 도면 등의 시각 자료가 아카이빙되는 경향을 보였다. 시기별로 공원 관련 정책의 특성이 뚜렷하게 구별되지는 않았지만, 전반적으로 공원의 계획과 조성에서 참여와 관리로 변화하는 현상은 확인할 수 있었다.

이 연구가 서울시의 공원 관련 정책에 시사하는 바는 다음과 같다. 첫째, 공원 기록물의 생산과 보존을 체계화할 필요가 있다. 특정 공원, 특정 행정업무, 부서와 관련하여 키워드로 관계 업무자와 일반 시민이 검색과 열람이 용이하도록 종이문건과 전자문건을 포괄하는 공원의 조성과 특성을 면밀히 고려한 분류 체계 마련이 필요하다. 둘째, 공원 관련 기록물은 방대한 빅데이터로서 과거의 공원 행정과 정책을 살펴볼 수 있는 사료의 가치를 지니는 동시에 앞으로의 공원 정책 수립을 위한 객관적인 참고 자료로 활용할 수 있다. 따라서, 빅데이터

분석 전문 인력을 확보하여 서울시의 다양한 정책 모니터링과 방향 제시에 도움을 줄 필요가 있다. 실제로, 이 연구에서 살펴보았듯이 공원을 조성하고 관리하는 데는 어느 한 기관과 부서가 일임한다기보다 다양한 분야의 행정전문가가 서로 복잡하게 관여하고 있으므로 이러한 기록물의 빅데이터 분석은 공원 정책 제안과 수립의 협의에 아주 유용하게 활용될 가능성을 지닌다.

물론, 연구의 한계도 존재한다. 이 연구는 방대한 분량의 공원 기록물의 텍스트 분석으로 그간 공원 정책의 전반적인 변화를 추적해보는 데는 분명 유효했다. 다만, 1950년대부터 2000년대까지 짧은 기간의 기록물의 내용 전체가 아니라 제목만을 다뤘다는 점에서 공원 관련 정책의 큰 변화 경향만을 파악했다는 한계가 있다. 향후 공원 기록물 내용까지 확장해 분석하면 관련 정책의 변화를 보다 면밀하게 파악할 수 있을 것이다. 한편으로 공공기록물의 제목은 신문과 소설 미디어를 비롯한 여타의 미디어 출판물의 제목과 비교했을 때 그 내용에 있어서 편차가 크지 않아 정책 내용의 큰 변화를 확인하는 데 어렵다는 한계가 있다. 다만, 공공기록물의 제목은 기록물의 내용을 간명히 축약하는 대표성을 지닌다고 가정해 연구를 진행하였으며, 기록물의 내용을 포함해 분석하면서 이러한 한계를 보완하는 다양한 후속 작업이 필요하다. 또한, 이 연구는 기록물의 양에 근거해 십 년 단위로 구분해 살펴보았지만, 더욱 다양하게 시기를 분류해 분석해보면서 공원 정책의 여러 변곡점을 살펴보고 대안적 시기 구분을 제안하는 후속 연구를 요청한다.

참고문헌

- 고하정, 2020a, “서울시 도시공원조성예산 변동 추세-예산서를 중심으로”, 『한국조경학회지』, 48(3).
- 고하정, 2020b, “텍스트마이닝을 활용한 도시공원에 대한 쟁점변화: 서울시의회 회의록을 중심으로”, 『서울도시연구』, 21(4).
- 김영재 외 1인, 2012, “서울시 정책의 변천에 관한 연구: 역대 서울시장을 중심으로”, 『한국행정사학지』, 30.
- 박울진, 2010, “최근 통계현황으로 본 도시공원녹지 변천 특성: 서울시를 중심으로”, 『한국산림휴양학회지』, 14(1).
- 이명준 외 2인, 2019, “미국 조경 아카이브 구축 동향과 특성 연구”, 『한국조경학회지』, 47(6).
- 유광흠 외 3인, 2008, “건축·도시공간 아카이브구축 전략 수립을 위한 연구”, 건축도시공간연구소.
- 임성은, 2013, “서울시장 교체에 따른 환경정책 변동유형 분석: 박원순 시장의 취임 전후 정책변동을 중심으로”, 『한국행정연구』, 22(3).
- 채진해 외 4인, 2014, “공원녹지기본계획의 운영체계 및 계획내용에 관한 연구: 런던, 뉴욕, 베를린, 시드니, 서울 사례를 중심으로”, 『한국조경학회지』, 42(2).
- 최용호, 2005, “공원녹지정책의 분석 및 방향설정 연구: 서울시를 중심으로”, 고려대학교.
- 최용호, 2006, “서울시 공원녹지정책”, 『한국환경생태학회 학술발표논문집』, 11~22.
- Broder, A.Z., et. al., 1997, “Syntactic clustering of the web”, *Computer Networks and ISDN Systems*, 29: 8~13.
- Compton, R., et. al., 2013, “Detecting future social unrest in unprocessed twitter data: emerging phenomena and big data”, *2013 IEEE International Conference on Intelligence and Security Informatics*, 56~60.
- Hassani, H., et. al., 2020, “Text mining in big data

- analytics", *Big Data and Cognitive Computing*, 4(1).
- He, W., et. al., 2013, "Social media competitive analysis and text mining: A case study in the pizza industry", *International Journal of Information Management*, 33(3).
- Hu, X., et. al., 2012, "Text analytics in social media", *Mining Text Data*. Boston: Springer, 385~414.
- Hutto, C., et. al., 2014, "Vader: A parsimonious rule-based model for sentiment analysis of social media text", *Proceedings of the International AAAI Conference on Web and Social Media*, 8(1).
- Kambatla, K., et. al., 2014, "Trends in big data analytics", *Journal of Parallel and Distributed Computing*, 74(7).
- Myslin, M., et. al., 2013, "Using twitter to examine smoking behavior and perceptions of emerging tobacco products", *Journal of Medical Internet Research*, 15(8).
- Tang, D., et. al., 2014, "CooooIII: A deep learning system for twitter sentiment classification", *Proceedings of the 8th International Workshop on Semantic Evaluation*, 208~212.
- <https://archives.seoul.go.kr>.(서울기록원, 2021 접속)
- <https://www.oed.com>.(English Oxford Living Dictionaries, 2021 접속)

원 고 접 수 일 : 2021년 6월 29일

1 차 심 사 완 료 일 : 2021년 9월 27일

최 종 원 고 채 택 일 : 2021년 9월 30일