한국기록관리학회지
24(3), 23-44, 2024.8
https://doi.org/10.14404/JKSARM.2024.24.3.023
pISSN 1598-1487 eISSN 2671-7247

JKSARM
Journal of Korean Society of Archives and Records Management

대통령기록포털 검색어 로그 분석 기반 이용자 정보요구 분석*

User Information Needs Analysis based on Search Terms Log of the Presidential Archives Portal*

이수현(Suhyeon Lee)[1], 오효정(Hyo-Jung Oh)[2]

E-mail: leeesuhyeon@gmail.com, ohj@jbnu.ac.kr

1 제1저자 전북대학교 일반대학원 기록관리학과 석사과정
2 교신저자 전북대학교 문헌정보학과 교수, 문화융복합아카이빙연구소 공동연구원

OPEN ACCESS

논문접수 2024.7.15
최초심사 2024.7.16
게재확정 2024.8.14

ORCID

Suhyeon Lee
https://orcid.org/0009-0002-3653-7363

Hyo-Jung Oh
https://orcid.org/0000-0001-8067-2832

© 한국기록관리학회
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0/) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

■ 본 논문은 2024년도 전북대학교 연구기반 조성비 지원에 의하여 연구되었음
■ 본 논문은 2024년도 한국연구재단 연구비 지원에 의한 결과의 일부임 (과제번호: NRF-2021R1I1A3047435)

초 록

최근 방대한 정보자원 속에서 이용자가 원하는 정보를 제공하기 위해 이용자 정보요구를 파악하여 맞춤형 서비스를 제공하는 큐레이션의 중요성이 커지고 있다. 이에 본 연구는 역사적 가치가 높은 대통령 기록물의 활용 가치 향상을 위해 대통령기록포털의 검색어 로그 분석을 통해 이용자 정보요구를 파악하고자 한다. 또한, 현 대통령기록포털에서 이용자의 정보요구에 부합하는 기록정보서비스를 제공하고 있는지 검색 품질을 파악함으로써 디지털 기록 큐레이션 서비스를 통해 개선할 요소를 제시하고자 한다. 이를 위해 최근 8년간의 검색어 로그를 기반으로 주제 분석, 워드 네트워크 분석을 수행하고, 이용자 요구가 높은 분야를 중심으로 대통령기록포털의 검색 품질을 정확도 관점에서 평가하였으며, 분석 결과를 기반으로 시사점을 도출하였다. 본 연구는 대통령 기록물 대상의 디지털 기록 큐레이션을 위한 사전적 연구로서, 이용자의 구체적인 정보요구를 파악하고 기록관 포털 사이트의 검색 품질을 정량적으로 수치화하여 이용자 만족도 향상을 위한 실효성 있는 연구를 수행하였다는 점에서 의의가 있다.

ABSTRACT

In recent years, there has been a significant increase in the importance of curation services that analyze user information requests to provide tailored information within extensive information resources. This study aims to identify user information needs by analyzing search term logs from the Presidential Archives Portal to enhance the utilization value of presidential records, which possess high historical significance. In addition, by evaluating the portal's search performance, this study seeks to determine whether the Presidential Archives Portal is providing archival information services that meet users' information needs and to suggest areas for improvement through digital record curation services. To achieve these objectives, topic analysis and word network analysis were conducted based on search term logs spanning the past eight years. The search quality of the Presidential Archives Portal was evaluated from an accuracy perspective, focusing on areas with high user demand, and recommendations were drawn based on the results of the analysis. As a preliminary study for digital record curation of presidential records, this study is significant because it identifies specific user information needs and quantifies the search quality of archival portal sites to improve user satisfaction.

Keywords: 대통령기록포털, 이용자 정보요구, 검색어 로그 분석, 검색 평가, 워드 네트워크 분석
Presidential Archives Portal, User Information Needs, Search Term Log Analysis, Retrieval Evaluation, Word Network Analysis,

1. 서론

1.1 연구배경 및 목적

현대 사회의 기술적 진보는 웹 환경의 보편화와 더불어 디지털 자원의 생산을 이끌었으며, 그 결과 기록을 특권 계층 외에도 누구나 접근할 수 있는 자원으로 자리 잡게 했다. 그러나 점차 정보의 양의 과도해지면서 이용자가 원하는 정보에 접근하기 어려운 상황이 초래됨에 따라 양질의 데이터 관리를 통해 방대한 양의 정보자원 가운데 이용자의 정보요구에 맞는 서비스를 제공하는 큐레이션의 중요성이 커지고 있다(조민지, 2020). 특히, 역사적으로 중요한 가치가 있는 기록물의 큐레이션은 가치 있는 기록물을 보존하고, 기록 활용을 통해 이용을 활성화한다는 측면에서 더욱 중요하다.

한편, 대통령의 직무수행 과정에서 생산된 대통령 기록물은 당대의 사회상을 반영할 뿐만 아니라 후대 대통령의 활동과 업적에 대한 역사적 평가를 위한 중요한 사료가 된다는 점에서 매우 가치 있는 자료이다(정광훈, 남영준, 2013). 따라서 대통령기록관은 대통령기록물을 보존하여 국민들에게 제공해야 할 책임이 있으며, 특히, 특정 시기에 큰 반향을 일으켰던 이슈나 정책에 대한 대통령 기록은 콘텐츠화를 통해 이용자가 쉽게 활용할 수 있도록 적극적으로 서비스하는 것이 필요하다(이해영, 2011).

현재 대통령기록관은 '대통령기록포털(https://www.pa.go.kr/index.jsp)'을 통해 웹 환경에서 이용자들에게 디지털 형태의 대통령 기록물 관련 서비스를 제공하고 있다. 정보공개청구[1]를 통해 입수한 2023년 대통령기록포털의 자체적인 포털 만족도 조사를 살펴본 결과, 서비스 전반에 대한 만족도는 89.8점으로 높게 나타난 반면, 해당 만족도 조사의 개선 의견으로 "검색정확도가 떨어짐", "기록컬렉션 내 필요한 콘텐츠를 사용자가 찾을 수 있는 검색 기능 필요", "기구축된 예전 콘텐츠도 주기적인 최신화 필요" 등, 검색과 관련된 이용자의 다양한 자유 의견이 수렴되어 포털 서비스에 대한 개선 필요성이 제기됨을 확인하였다. 이는 거의 90점에 가까운 높은 이용자 만족도와는 대조적으로 대통령기록포털에서 제공하는 검색 서비스가 이용자의 정보요구를 충분히 만족시키지 못하고 있음을 함의하며, 나아가 맞춤형 대통령 기록물에 대한 큐레이션 서비스 개발의 중요성을 방증한다.

이에 본 연구는 디지털 기록 큐레이션을 통한 대통령 기록물의 활용 가치 제고를 위한 사전 연구 성격으로, 이용자의 요구가 객관적이고 직접적으로 담겨있는 검색어 로그를 기반으로 다양한 분석 기법을 적용함으로써 대통령기록포털에 대한 이용자 정보요구가 무엇인지 먼저 분석하고자 한다. 나아가 현재 대통령기록포털에서 이용자의 정보요구에 부합하는 기록정보서비스를 제공하고 있는지를 검색 품질을 파악함으로써 디지털 기록 큐레이션 서비스 개발 시 고려해야 할 시사점을 도출하고자 한다.

1.2 선행연구

본 연구는 이용자의 정보요구를 파악하기 위해 검색어 로그를 활용한 연구에 주목, 특히 기록관의 웹 서비스를 대상으로 한 선행연구를 중점적으로 살펴보았다. 기록관의 포털 사이트를 대상으로 이용자 로그를 분석한 주요 연구로는 백지연(2020), 이해영(2011), 진주영(2018)이 있다. 먼저 이해영(2011)은 대통령기록관의 웹사이트 검색어 분석을 통해 이용자의 요구를 확인하고, 분석 결과를 대통령기록관 정보서비스, 특히 콘텐츠 서비스에 적용하는 방안을 모색하였다. 이를 위해 2010년 1월부터 11월까지 약 1년간 축적된 대통령기록관 웹사이트의 검색 로그 데이터를 대상으로 일반 검색어 분석과 대통령 이름별 분석, 검색어 주제 영역 분석을 수행하였다.

1) 정보공개청구, [접수번호 11766700] 2024.01.18.

https://doi.org/10.14404/JKSARM.2024.24.3.023

분석 결과, 이용자들이 많이 검색한 것으로 나타난 기록 유형이나 주제 분야를 콘텐츠화하거나, 온라인·오프라인 전시를 진행하는 등 분석 결과를 대통령기록관 콘텐츠 서비스에 활용할 수 있는 방안을 제안하였다.

진주영(2018)은 국가기록원 웹사이트의 검색 로그를 크게 검색어와 외부 유입 경로로 나누어 분석하였으며, 그중 검색어는 검색 횟수와 등장 횟수 분석, 주제별 분석, 그리고 최근 이슈별과 연계 분석으로 세분화하여 분석을 수행하였다. 이를 통해 이용자들이 가장 많이 검색한 검색어와 빈번하게 등장하는 검색어를 파악하고, 이용자들의 관심 분야를 확인하였으며, 분석 결과를 기반으로 국가기록원 웹사이트의 이용자 서비스 방안을 제안하였다.

마지막으로 백지연(2020)은 국가기록원 웹사이트의 검색 로그를 시계열 분석, 기초 분석, 정보요구 유형별, 검색 결과 유형별, 질의입력 유형별 등 다각도의 분석을 수행하였다. 이 중 정보요구 유형별은 기관명, 인명, 사건, 기타로 구분하여 이용자의 정보요구를 파악하고자 하였으며, 질의입력 유형별 분석을 통해 이용자의 검색 행태를 분석하였다. 또한 분석 결과 도출한 시사점을 바탕으로 국가기록원 웹사이트 개선방안을 제시하였다.

이상의 선행연구와 같이 대통령기록관, 국가기록원 등 다양한 기관을 대상으로 검색어 로그 분석을 통해 이용자의 요구를 파악하고자 하는 시도가 계속되고 있다. 그러나 대부분의 연구에서 이용의 관심 주제 영역 혹은 기록 유형을 유추하는 수준에 그쳤으며, 이용자의 명확한 정보요구를 파악하기에는 한계가 있었다. 특히, 본 연구의 대상 기관인 대통령기록관의 검색어 로그 분석에 관한 연구가 가장 최근 수행된 것은 2011년(이해영)으로, 현재는 이용자의 정보요구가 크게 변화하였을 것으로 예상된다.

이에 본 연구는 2016년부터 2023년까지 대통령기록포털을 통해 축적된 최근 약 8년간의 검색어 로그 빅데이터를 기반으로 대통령기록포털 이용자의 정보요구를 파악하였다. 또한, 기존 검색어 로그 분석에서 주로 활용되는 시계열 분석과 주제 분석에 더해 워드 네트워크 분석을 수행하여, 이용자의 정보요구를 보다 명확하게 파악하고자 하였다. 특히 이용자 요구를 반영한 질의셋을 구축, 대통령기록포털의 검색 정확도를 측정함으로써 실제 디지털 기록 큐레이션 서비스 개발 시 고려해야 할 사항을 도출하였다는 점에서 실효성이 크다.

1.3 연구방법

본 연구는 대통령기록포털의 검색어 로그를 분석하여 파악한 이용자의 정보요구를 바탕으로 시사점을 도출하였으며, 이를 위해 다음과 같은 방법으로 연구를 진행하였다.

첫째, 대통령기록포털의 검색어 로그를 수집하고 시계열 분석을 통해 로그 분석 대상 검색어 목록을 도출하였다. 대통령기록포털의 검색어 로그 데이터는 정보공개청구[2]를 통해 입수, 수집된 로그 데이터를 기반으로 시계열 분석을 수행하여 이상치를 파악하였으며, 이를 제거하여 이용자의 정보요구를 담고 있는 분석 대상 검색어 목록을 도출하였다. 둘째, 이용자의 관심 영역을 파악하기 위해 대통령기록포털의 검색어를 9개 주제 분야로 구분하고, 이용자의 관심이 높은 5개 분야 대상 주제 분석을 수행하여 이용자의 관심이 높은 검색어 집합을 파악하였다. 셋째, 주제 분야별로 이용자의 관심이 높은 1위 검색어를 기준으로 동시 출현 키워드를 워드 네트워크로 시각화하고, 이를 분석하여 주제 분야별 검색어에 대한 이용자의 명확한 정보요구를 도출하였다. 넷째, 대통령기록포털 검색 서비스 품질 점검 테스트를 수행하였다. 이를 통해 대통령기록포털에서 이용자의 검색 의도에 부합하는 기록물을 제공하고 있는지를 정량적으로 평가하여 검색 성능을 파악하고, 개선점을 토출, 디지털 기록 큐레이션 서비스 개발 시 고려해야 할 요소를 파악하였다. 마지막으로 이전 절차에서 수행한 주제 분석, 워드 네트워크 분석, 검색 서비스 품질 점검 테스트의 분석 결과를 통해 시사점을 도출하였다.

2) 정보공개청구, [접수번호 11376667] 2023.10.19.

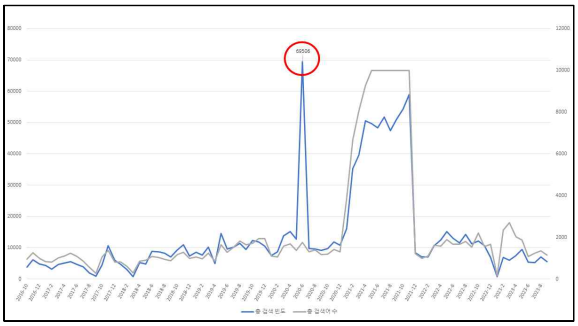
https://doi.org/10.14404/JKSARM.2024.24.3.023

2. 대통령기록포털 검색어 로그 분석

2.1 시계열 분석 기반 분석 대상 검색어 선정

<표 1> 검색어 로그 데이터 통계

연구대상 속성	개수
수집 기간	95개월(2016년 10월 ~ 2023년 9월)
전체 검색어 수	191,175개
질의어 수 (중복제거)	64,532개
총 검색 빈도	498,445회
월평균 검색 빈도	5,246회



<그림 1> 월별 시계열 분석

<표 1>은 정보공개청구를 통해 입수된 대통령기록포털의 검색어 로그 데이터에 대한 통계치로, 대상 기간은 2016년 10월부터 2023년 9월까지 약 8년(95개월)간 총 191,175개의 검색어가 수집되었다. 이 중 중복을 제거한 질의어 수는 64,532개였다. 95개월간의 총 검색 빈도는 498,455회로, 대통령기록포털의 매달 평균 검색량은 5,246회로 파악되었다.

대통령기록포털의 검색어 중 검색어 로그 분석 대상 검색어를 선정하기 위한 연도별 및 월별 시계열 분석 결과, 검색어 수와 검색 빈도 측면에서 이상치를 파악하였으며, 특히 2020년 6월의 검색 빈도가 다른 시기에 비해 큰 폭으로 증가한 것을 볼 수 있었다(<그림 1> 빨간색 동그라미 참고). 해당 시기 가장 많이 검색된 단어는 교육과 관련된 검색어로, 특히 '교육'이란 질의어는 전년도인 2019년 1년간 총 48회 검색된 것에 비해. 2020년 6월 한 달에만 무려 53,158회 등장, 검색 빈도가 기하급수적으로 증가한 것으로 파악되었다. 그러나 이 시기 이와 관련된 특별한 이슈나 기록물은 포착되지 않아 대통령기록관에 해당 현상에 대해 문의한 결과, 이러한 검색어는 소수의 이용자 혹은 기관에서 반복적이고 기계적으로 검색어를 입력한 것으로 파악되며, 이는 대통령기록관 이용자의 일반적인 검색행태로 볼 수 없으므로 통계 시 제외해도 무방하다는 답변을 받았다.

이렇듯 대통령기록포털의 일반적인 이용자의 유이미한 정보요구로 볼 수 없다고 판단된 질의어를 5개 이상치로 선별한 외에도 '1', 'a', '다', '아' 등 '의미를 파악하기 어려운 한 글자 검색어' 25개와 '&dir', '12101500000000400', '11 11 22'과 같이 '의미를 파악하기 어려운 숫자 및 문자열' 104개, '검색어가 공백인 경우'를 분석, 제외 검색어로 선정하였다. 위의 과정을 통해 도출한 분석 제외 검색어의 유형별 개수는 <표 2>와 같으며, 대통령기록포털 검색어를 대상으로 중복제거를 수행한 검색어 목록의 전체 검색어 64,532개에서 분석 제외 검색어 135개를 제외하여, 최종적으로 총 64,397개의 검색어로 구성된 주제 분석 대상 검색어 목록을 도출하였다

<표 2> 분석 제외 검색어 유형별 개수

분석 제외 검색어 유형	개수
이용자 전체의 정보요구라고 볼 수 없는 경우	5
의미를 파악하기 어려운 한 글자 검색어	25
의미를 파악하기 어려운 숫자 및 문자열	104
검색어가 공백인 경우	1
합계	135

https://doi.org/10.14404/JKSARM.2024.24.3.023

2.2 주제 분석

본 절은 이용자의 관심 영역에 대해 파악하기 위해 대통령기록포털 검색어 로그를 주제 분야별로 구분하여 분석하였다. 이를 위해 선행연구를 기반으로 주제 분야를 도출하고, 검색 빈도 상위 100개 검색어를 대상으로 주제 분야를 설정하여 이용자의 관심이 높은 주제 분야를 탐색하고자 하였다.

주제 분야 선정은 먼저 본 연구와 동일한 대통령기록포털을 대상으로 검색어 로그 분석 연구를 수행한 이해영(2011)의 연구에서 설정한 주제 분야를 기반으로 유사 기관인 국가기록원을 대상으로 검색어 로그 분석 연구를 수행한 백지연(2020)과 김대원(2021)의 연구를 참고하되, 가능한 최근 8년간의 이용자 요구가 많은 분야를 포함하고자 하였다. 현시점의 대통령기록포털 이용자 검색어를 포괄할 수 있도록 재구성한 결과, “정책 및 업무수행”을 비롯해 “인명 및 직책”, “기록관리” 등 9개 주제 분야를 도출하였다. <표 3>은 이용자의 관심이 높은 주제 분야를 파악하기 위해 분석 대상 검색어 64,397개 중 검색 빈도가 높은 상위 100개 검색어 도출된 9개 주제 분야로 구분하여 주제 분야별로 해당하는 검색어의 수와 검색 빈도, 출현 개월 수를 분석한 결과이다.

<표 3> 검색 빈도 기준 상위 100개 검색어 주제 분류

주제 분야	검색어	검색어 수	전체 검색 빈도	평균 검색 빈도	전체 출현 개월 수	평균 출현 개월 수
정책 및 업무수행	국무회의, 보고서, 국정과제, 헌법, 백서, 독도, 보고, 공약, 채용, 지방이양, 관보, 국민청원, 가스안전, 지시사항, 저도, 국무회의록, 참여정부, 경부고속도로, 국정백서, 가평, 경제, 회의록, 대통령과 함께 읽는 보고서, 제헌헌법	24	26176	1091	1217	51
인명 및 직책	노무현, 박정희, 김영삼, 이승만, 박근혜, 문재인, 김대중, 대통령, 전두환, 이명박, 노태우, 최규하, 윤보선, 프란체스카, 역대 대통령, 육영수	16	72506	4532	1154	72
기록관리	기록, 구술	2	1612	806	127	64
행정박물 및 매체	서한, 선물, 휘호, 훈장, 사진, 표창, 초상화, 문화재	8	5486	686	393	49
기관 및 단체	청와대, 정책기획위원회, 조직도, 경무대, 인수위원회, 경찰, 국민경제자문회의, 영빈관	8	7639	955	357	45
외교	남북정상회담, 정상회담, 월남, 통일, 북한, 미국, 국가안전보장회의, 남북, 대만, 유엔, apec	11	9606	873	492	45
행사	연설문, 연설, 대통령 연설문, 취임사, 취임식, 신년사, 대통령 취임사, 축사, 기자회견, 취임, 취임식 명단, 올림픽, 유엔, 대통령연설문, 노무현 연설문, 연설기록, 연설문집	16	31917	1995	953	60
사건	5.18, 세월호, 새마을, 광주, 6.25, 새마을운동, 광복절, 광주대단지, 탄핵, 4.19, 대구, imf, 한국전쟁	13	12963	997	616	47
기타	식단, 구내식당	2	1338	669	130	65

9개 주제 분야 중 검색어 수가 많은 상위 5개 주제 분야는 "정책 및 업무수행", "인명 및 직책", "외교", "행사", "사건" 분야로 해당 분야에는 상위 100개 검색어 중 80%가 속해 이용자의 관심이 집중되어 있음을 알 수 있다. 따라서 본 연구는 이용자의 수요가 높은 5개 주제 분야를 주제 분석 대상으로 선정, 각 분야에 대해 특징적으로 나타나는 검색어 집합을 통해 이용자의 관심사를 파악하고자 하였다. 또한, 대통령기록포털이 현재 이용자의 관심사에 부합하는 서비스를 제공하고 있는지 파악하기 위해 대통령기록포털의 컬렉션 및 콘텐츠를 살펴보았다.

https://doi.org/10.14404/JKSARM.2024.24.3.023

2.2.1 정책 및 업무수행

"정책 및 업무수행" 분야는 대통령이 임기 중 수행한 정책과 대통령의 업무에 대한 검색어 집합을 포괄하는 주제이다. 검색 빈도 기준 상위 100개 검색어 중 정책 및 업무수행 분야에 속하는 검색어는 총 24개이며, 각 검색어의 순위와 검색 빈도, 출현 개월 수는 <표 4>와 같다.

<표 4> 정책 및 업무수행 분야 검색어 목록

순위	전체 순위	검색어	검색 빈도	출현 개월 수	순위	전체 순위	검색어	검색 빈도	출현 개월 수
1	15	국무회의	2888	79	13	43	가스안전	992	2
2	18	보고서	2157	79	14	47	지시사항	924	42
3	21	국정과제	1841	75	15	62	저도	759	21
4	24	헌법	1615	65	16	64	국무회의록	743	48
5	25	백서	1485	79	17	68	참여정부	709	57
6	30	독도	1274	55	18	69	경부고속도로	706	53
7	32	보고	1219	61	19	70	국정백서	698	59
8	34	공약	1121	71	20	76	가평	620	2
9	38	채용	1032	75	21	77	경제	619	51
10	39	지방이양	1012	24	22	82	회의록	602	48
11	40	관보	1007	45	23	83	대통령과 함께 읽는 보고서	598	41
12	42	국민청원	994	44	24	94	제헌헌법	561	41

"정책 및 업무수행" 분야 검색어 가운데 검색 빈도가 가장 높은 검색어는 대통령기록포털에서 2,888회 검색된 '국무회의'이다. 이외에도 '국정과제', '헌법', '지시사항' 등의 검색어가 상위 100개 검색어 내에 등장하여 국무회의에서 다루어지는 주제나, 이를 통해 생산되는 기록물에 대한 이용자의 관심이 높음을 확인할 수 있었다.

한편, 대통령기록포털은 기록컬렉션의 '정책기록' 콘텐츠를 통해 "정책 및 업무수행" 분야의 검색어를 찾는 이용자들을 위한 중요 정책 관련 기록물을 제공하고 있으나, 기제작된 콘텐츠에 대한 업데이트가 이루어지지 않아 해당 콘텐츠가 이용자들의 관심사에 부합하는 기록물을 원활하게 제공하기에 어려움이 있었다. 또한, 정책 및 업무수행 분야에서 검색 빈도가 가장 높은 '국무회의'의 경우, 출현 개월 수 또한 높아 해당 검색어에 대한 검색이 꾸준히 이루어지고 있음을 유추할 수 있으나, 이용자가 국무회의 관련 기록물을 찾기 쉽게 모아서 제공해 주는 별도의 서비스나 컬렉션을 찾을 수 없었다.

"정책 및 업무수행" 분야 검색어로 등장한 '국정과제', '공약' 등은 대통령마다 유의미한 차이가 존재하는 기록으로 대통령기록포털의 이용자가 특정 대통령 시기 생산된 기록물을 찾거나, 대통령에 따라 변화하는 정책 양상을 알기 위해 검색을 수행했을 것이라고 짐작할 수 있다. 따라서 각 대통령 임기별로 정책의 변화 또는 유의미한 차이점을 시각적으로 보여주는 콘텐츠를 제공하는 방안을 고려해 볼 수 있다.

2.2.2 인명 및 직책

"인명 및 직책" 분야는 9개 주제 분야 중 검색 빈도와 평균 출현 개월 수가 가장 높아 대통령기록포털의 검색어 가운데서도 그 중요도가 매우 큰 주제 분야이다. 본 절에서는 검색 빈도 기준 상위 100개 검색어 가운데 "인명 및 직책" 분야에 속하는 검색어 16개를 검색어의 유형별로 분류하여 분석하였으며, 그 결과는 <표 5>와 같다.

https://doi.org/10.14404/JKSARM.2024.24.3.023

<표 5> 인명 및 직책 분야 검색어 목록

순위	전체 순위	검색어	분류	검색 빈도	출현 개월 수	순위	전체 순위	검색어	분류	검색 빈도	출현 개월 수
1	2	노무현	대통령	9919	82	9	11	전두환	대통령	4145	82
2	3	박정희	대통령	8264	79	10	12	이명박	대통령	3130	80
3	4	김영삼	대통령	8087	79	11	14	노태우	대통령	3056	79
4	5	이승만	대통령	7638	83	12	22	최규하	대통령	1734	68
5	6	박근혜	대통령	6644	83	13	26	윤보선	대통령	1482	66
6	7	문재인	대통령	6117	80	14	66	프란체스카	영부인	729	24
7	9	김대중	대통령	5426	82	15	74	역대 대통령	직책	652	57
8	10	대통령	직책	4917	82	16	92	육영수	영부인	566	48

“인명 및 직책” 분야에 해당하는 16개 검색어 중 12개는 대통령의 이름이었으며, 직책과 영부인 이름에 대한 검색어가 각각 2개씩 등장하였다. 순위와 검색 빈도, 출현 개월 수 등의 수치를 종합적으로 살펴본 결과, 대통령기록포털의 검색어인 만큼 역대 대통령의 이름에 대한 검색 빈도와 출현 개월 수가 높게 나타났다. 특징적인 부분은 “인명 및 직책” 분야의 검색어로 영부인 이름인 ‘프란체스카’, ‘육영수’가 등장한 것으로, 이는 영부인 관련 기록물에 대한 이용자의 수요가 있음을 시사한다.

대통령기록포털은 상세 검색의 조건으로 대통령별 검색이 가능하며, 대통령기록포털의 ‘대통령 이야기’ 페이지를 통해 역대 대통령에 대한 약력과 취임식 자료, 대통령 기록물 등을 확인할 수 있었다. 그러나, 영부인 관련 기록물은 그와 함께 관리되고 있지 않았으며, 별도의 컬렉션이나 콘텐츠를 제공하고 있지 않았다. 영부인 기록물에 대한 이용자의 관심이 뚜렷한 만큼, 대통령 기록물과 마찬가지로 영부인의 연설문과 편지, 사진 선물 등 영부인 관련 기록물을 인물별, 기록 유형별, 연도별 등 다양한 기준에 따른 체계적인 관리 방안을 마련해야 한다.

2.2.3 외교

<표 6> 외교 분야 검색어 목록

순위	전체 순위	검색어	분류	검색 빈도	출현 개월 수	순위	전체 순위	검색어	분류	검색 빈도	출현 개월 수
1	16	남북정상회담	회의	2534	72	7	79	국가안전보장회의	회의	612	42
2	37	정상회담	회의	1068	63	8	91	남북	국가명	574	42
3	50	월남	국가명	879	40	9	96	대만	국가명	548	29
4	56	통일	안건	810	48	10	98	유엔	기구	538	32
5	60	북한	국가명	775	55	11	99	apec	기구	534	25
6	65	미국	국가명	734	44						

“외교” 분야는 대통령의 외교와 관련된 검색어의 집합으로, 검색 빈도 기준 상위 100개 검색어 중 “외교” 분야에 속하는 검색어는 11개였으며, 이에 대한 분석 결과는 <표 6>과 같았다. 분석 결과, “외교” 분야 검색어 중 검색 빈도가 가장 높은 검색어는 ‘남북정상회담’이었으며, 이외에도 ‘정상회담’이나 ‘통일’, ‘북한’, ‘남북’ 등 남북정상회담과 관련 있는 단어가 다수 등장해 대통령기록포털의 이용자들이 외교 관련 기록물 가운데 ‘남북정상회담’에 대한 관심이 높음을 확인하였다.

현재 대통령기록포털은 기록컬렉션의 ‘정책기록’ 콘텐츠 중 세부 분야인 ‘외교’ 페이지를 통해 외교 관련 기록물을 제공하고 있다. 그러나 해당 페이지에서 제공하고 있는 분류 기준만으로는 남북정상회담 관련 기록물만을 특정

https://doi.org/10.14404/JKSARM.2024.24.3.023

해서 볼 수 없어, 이용자의 관심이 높은 주제의 기록물에 대한 접근이 용이하지 않다. 또한, 해당 콘텐츠는 이승만, 박정희 대통령 시기의 외교정책만을 대상으로 하고 있어 다른 대통령 시기의 외교 기록물을 찾는 이용자들을 만족시키기에는 부족하다. 따라서 외교 관련 콘텐츠의 대상을 대통령기록포털이 소장하고 있는 외교 기록물 전체로 확장하는 서비스 확대가 필요하다.

2.2.4 행사

<표 7> 행사 분야 검색어 목록

순위	전체 순위	검색어	검색 빈도	출현 개월 수	순위	전체 순위	검색어	검색 빈도	출현 개월 수
1	1	연설문	10176	80	9	46	기자회견	925	42
2	8	취임사	5763	83	10	49	대통령연설문	885	42
3	13	연설	3098	80	11	51	취임	879	61
4	20	대통령 연설문	1860	67	12	71	노무현 연설문	693	55
5	23	취임식	1665	76	13	81	연설기록	607	59
6	27	신년사	1374	62	14	85	연설문집	591	55
7	33	대통령 취임사	1146	67	15	88	취임식 명단	582	7
8	35	축사	1111	71	16	93	올림픽	562	46

<표 7>은 검색 빈도 기준 상위 100개 검색어 가운데 “행사” 분야에 속하는 검색어 16개에 대한 분석 결과이다. “행사” 분야는 대통령이 참석하는 행사와 관련된 검색어로 구성되어 있으며, 그 중 검색 빈도가 가장 높은 검색어는 ‘연설문’이다. 이외에도 ‘취임사’, ‘신년사’, ‘축사’와 같이 ‘연설문’에 해당하는 검색어의 출현을 통해 다양한 연설기록을 찾는 이용자가 다수 존재하였는데, 이는 특정 행사 자체보다 행사에서 생산되는 대통령 연설문에 대한 이용자의 관심이 더 높음을 의미한다.

대통령기록포털은 현재 기록컬렉션 내 ‘연설기록’ 콘텐츠를 통해 대통령 기록물 중 연설기록만을 정리해 제공하고 있으며, 이는 형태별, 유형별, 대통령별, 연도별 등으로 검색 가능하다. 그러나 ‘3.1절’, ‘삼일절’과 같은 동의어에 대한 통제가 이루어지지 않고 있으며, ‘3.1절 기념사’, ‘광복절 축사’와 같이 매년 반복적으로 진행되는 행사에 대한 연설문을 연결해 제공해 주지 않고 있다. 때문에 이용자의 검색 능력에 따라 검색 결과에 차이가 있어, 일부 이용자들은 원하는 기록물을 탐색하는 데에 어려움을 겪을 수 있다는 문제가 있다. 따라서 대통령 연설문을 찾고자 하는 이용자를 만족시키기 위해 시소러스를 통해 유의어를 통제하고, 정기적인 행사에 대한 역대 대통령의 연설문을 연결해서 보여주는 등 이용자가 원하는 기록물에 접근하기 쉬운 환경을 제공해야 한다.

2.2.5 사건

주제 분야별 분석 중 마지막 관점으로 “사건” 관련 질의어를 분석하였다. “사건” 분야는 “행사” 분야에 비해 사회적 이슈나 역사적으로 중요한 일과 관련된 검색어를 분류한 주제 분야로, <표 8>은 검색 빈도 기준 상위 100개 검색어 가운데 “사건” 관련 검색어 13개에 대한 분석 결과를 정리한 것이다.

https://doi.org/10.14404/JKSARM.2024.24.3.023

<표 8> 사건 분야 검색어 목록

순위	전체 순위	검색어	분류	검색 빈도	출현 개월 수	순위	전체 순위	검색어	분류	검색 빈도	출현 개월 수
1	19	5.18	민주화운동	2120	56	8	55	광주대단지	민주화운동	831	26
2	28	세월호	참사	1335	60	9	58	탄핵	탄핵	799	57
3	29	새마을	새마을운동	1334	50	10	73	4.19	민주화운동	661	43
4	31	광주	민주화운동	1240	47	11	80	대구	참사	610	17
5	36	6.25	전쟁	1069	44	12	84	imf	IMF	591	55
6	45	새마을운동	새마을운동	974	55	13	100	한국전쟁	전쟁	531	46
7	52	광복절	일제강점기	868	60						

“사건” 분야의 검색어는 선행연구와 대통령기록포털 사이트를 참고하여 구분한 사건의 종류에 따라 민주화운동, 참사, 새마을운동, 전쟁, 일제강점기, 탄핵, IMF의 7개 분류로 나누어 분석하였다. 7개 분류 가운데 민주화운동과 관련된 단어가 총 4개로 가장 많이 등장한 것과 더불어 “사건” 분야 검색어 중 검색 빈도가 가장 높은 검색어는 ‘5.18’인 것을 고려하면, 여러 사건 가운데 ‘5.18 민주화운동’과 관련된 대통령 기록물에 대한 이용자 관심이 매우 높음을 파악할 수 있다.

대통령기록포털은 기록컬렉션 중 ‘이기록 이순간’ 콘텐츠를 통해 5.18, 4.19 등 주요 사건에 대한 콘텐츠와 연관 기록물을 함께 제공하고 있다. 그러나, ‘이기록 이순간’은 사건뿐만 아니라 기념일, 정책 등 다양한 주제를 한 번에 다루고 있고, 콘텐츠 검색이 불가능하다. 따라서 중요 사건과 관련된 기록물을 한 곳에서 찾아볼 수 있도록 주제 검색을 제공하고, 하위 주제를 통해 큰 범주의 주제 중에서도 이용자가 원하는 기록물만을 제공받을 수 있도록 하는 방안이 필요하다.

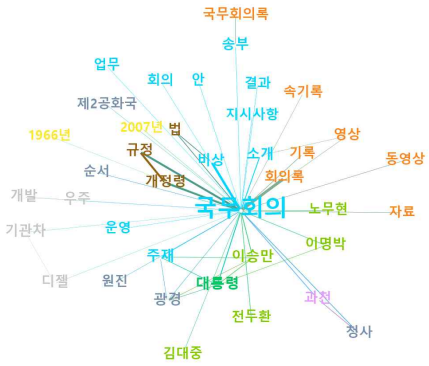
2.3 워드 네트워크 분석

상기 주제 분석 단계에서는 검색어의 검색 빈도와 출현 개월 수, 주로 등장하는 검색어 집합의 특성 등을 통해 이용자의 관심사를 파악하였다. 그러나 대통령기록포털의 검색어는 대부분 한 단어 위주의 매우 짧은 단어로 이루어져 있어 주제 분석만으로는 이용자의 명확한 정보요구를 파악하기에 어려움이 있었다. 따라서, 본 절에서는 이용자가 입력한 검색어와 함께 검색된 키워드들의 조합을 통해 이용자의 정보요구를 구체화하는 과정을 예시하고자 하였다. 이를 위해 한 검색어 내 함께 등장한 단어들을 묶은 동시 출현 단어 쌍 가운데 출현 빈도가 많은 상위 50개로 워드 네트워크를 구성하였으며, 워드 네트워크에 등장한 단어를 각 단어의 특성에 따라 군집으로 묶어 구분하였다. <그림 2>는 이용자의 관심이 높은 5개 주제 분야의 1위 검색어를 대상으로 동시 출현 키워드를 분석하여 각 주제 분야별 대표 검색어인 검색 빈도 1위 검색어에 대한 워드 네트워크를 제작한 것이다.

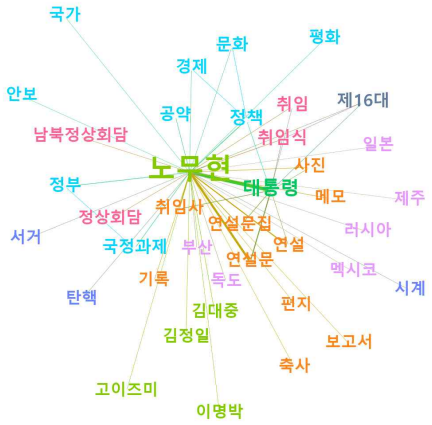
2.3.1 정책 및 업무수행

“정책 및 업무수행” 분야의 검색 빈도 1위 검색어는 “국무회의”이다. 따라서 대통령기록포털의 분석 대상 검색어 중 “국무회의”를 포함하는 검색어를 도출하여 동시 출현 키워드를 파악하고, 동시 출현 단어 쌍 50개를 도출하여 그림 <2.a>과 같이 총 39개 단어로 구성된 검색어 “국무회의” 워드 네트워크를 제작하였다.

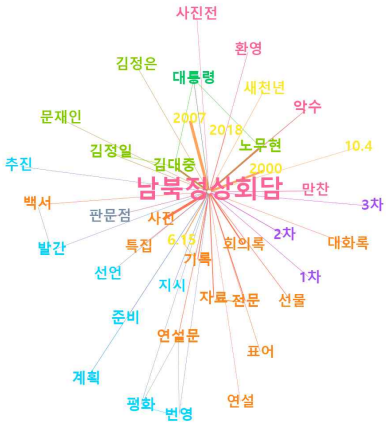
https://doi.org/10.14404/JKSARM.2024.24.3.023



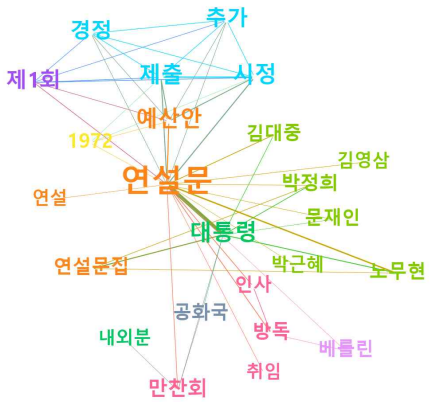
a) 정책 및 업무수행 분야 대표 검색어 "국무회의"



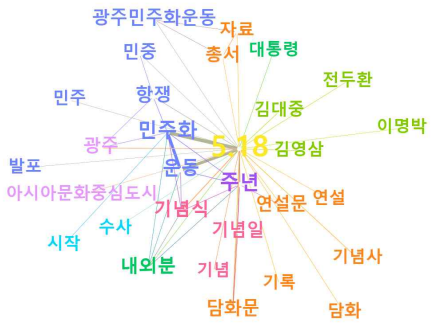
b) 인명 및 직책 분야 대표 검색어 "노무현"



c) 외교 분야 대표 검색어 "남북정상회담"



d) 행사 분야 대표 검색어 "연설문"



e) 사건 분야 대표 검색어 "5.18"

<그림 2> 분야별 대표 검색어 워드 네트워크

https://doi.org/10.14404/JKSARM.2024.24.3.023

“국무회의” 네트워크에서 ‘비상–국무회의’ 단어 쌍의 연결선이 뚜렷하게 나타났는데, 단어 쌍의 출현 빈도가 높을수록 연결선의 두께가 두꺼워지는 것을 고려하면 박정희 정부 시기 운영하였던 과도입법기구인 “비상국무회의”에 대한 관심이 높다고 추측할 수 있다. 또, ‘국무회의–개정령’, ‘국무회의–규정’, ‘개정령–규정’ 단어쌍의 연결선도 뚜렷하게 나타나 “국무회의 규정의 개정령”을 찾는 이용자가 다수 존재함을 확인할 수 있었다.

한편, “정책 및 업무수행” 분야의 대표 검색어인 “국무회의” 워드 네트워크에 출현한 단어는 각 단어의 특성에 따라 업무(■), 기록(■), 인명(■), 기타(■), 산업(■), 법률(■), 연월일(■), 지명(■), 직책(■)의 총 9가지 군집으로 구분할 수 있었다. 국무회의” 네트워크는 타 주제 분야의 워드 네트워크와 달리 유일하게 산업(■) 군집과 법률(■) 군집의 단어가 등장하는데, 이는 국무회의에서 산업 관련 법이나 규정에 대한 의결이 이루어지기 때문이다. 산업(■) 군집의 ‘우주’, ‘개발’, ‘디젤’, ‘기관차’ 등의 단어는 국무회의에서 의결한 ‘우주개발진흥법’과 ‘디젤기관차 및 엔진 도입을 위한 미국 수출입 은행 등과의 차관 협정 체결(안)’으로부터 도출된 검색어로, 산업과 법률 군집의 단어가 실질적으로 연관성이 깊음을 알 수 있었다.

2.3.2 인명 및 직책

인명 및 직책 분야의 검색 빈도 1위 검색어는 “노무현”으로, 대통령기록포털의 분석 대상 검색어 중 “노무현”을 포함하는 검색어를 도출하여 동시 출현 키워드를 파악하여 인명 및 직책 분야 대표 검색어인 “노무현” 워드 네트워크를 제작하였다(그림 <2.b> 참고). “노무현” 워드 네트워크의 단어 중 연결선이 뚜렷하게 나타나는 것은 ‘노무현–대통령’, ‘노무현–연설문’의 단어 쌍으로, 이를 통해 “노무현 대통령의 연설문을 찾고 싶다”라는 정보요구를 가지고 검색을 수행하는 이용자가 다수 있음을 추측할 수 있다.

인명 및 직책 분야의 대표 검색어인 “노무현” 워드 네트워크에 출현한 단어는 각 단어의 특성에 따라 기록(■), 업무(■), 지명(■), 인명(■), 행사(■), 사건(■), 직책(■), 기타(■)의 총 8가지 군집으로 구분할 수 있었으며, “노무현” 네트워크는 ‘기록(■)’ 군집에 속한 단어가 25.64%로 가장 많았다. 워드 네트워크를 군집 유형별로 나누어 살펴보면, “노무현” 검색어와 관련하여 ‘편지’, ‘보고서’, ‘메모’, ‘연설문’ 등 ‘기록(■)’ 군집에 속하는 키워드가 다수 등장하여 노무현 전 대통령과 관련된 다양한 유형의 기록물을 찾는 이용자가 많음을 유추할 수 있었다. 또한, “노무현” 네트워크에서 두 번째로 많은 키워드를 포괄하는 ‘업무(■)’ 군집은 ‘경제’, ‘정책’, ‘국정과제’ 등 대통령의 업무에 대한 기록물과 관련된 검색어가 다수 등장하였다. 이외에도 ‘지명(■)’ 군집의 ‘멕시코’, ‘러시아’, ‘일본’ 키워드나, ‘인명(■)’ 군집의 ‘김정일’, ‘고이즈미’ 키워드와 같이 외교 관련 기록에 대한 관심이 높게 나타나는 등 다양한 주제 분야의 기록물에 대한 관심이 드러났다. 따라서 이용자의 정보요구를 충족시키기 위해서는 다양한 주제 분야의 기록물을 한 곳에서 볼 수 있도록 관리하여 이용자 편의를 증진하는 방안을 모색해야 할 것이다.

2.3.3 외교

외교 분야의 1위 검색어는 “남북정상회담”으로 해당 검색어의 동시 출현 키워드로 구성된 “남북정상회담” 워드 네트워크(그림 <2.c> 참고)를 살펴보면, ‘2007–남북정상회담’ 단어 쌍의 연결선이 굵게 나타나는 것을 볼 수 있다. 마찬가지로 “남북정상회담–사진” 단어 쌍의 연결선도 뚜렷하게 보이는데, 이를 통해 대통령기록포털의 이용자들이 2007년 남북정상회담 관련 사진 자료를 찾는 이용자가 다수 있을 것임을 유추할 수 있다.

“남북정상회담” 네트워크에 등장하는 단어는 기록(■), 업무(■), 연월일(■), 인명(■), 행사(■), 회차(■), 직책(■), 기타(■)의 총 8개 군집으로 구분할 수 있으며, 그중 가장 높은 비중을 차지하는 기록(■) 군집에 ‘사진’, ‘회의록’, ‘백서’, ‘선물’, ‘표어’ 등 기록물 유형 관련 단어가 매우 다양하게 나타났다. 또, 연도나 일자 관련 단어를

https://doi.org/10.14404/JKSARM.2024.24.3.023

포괄하는 연월일(■) 군집이 14.63%로 타 주제 분야 네트워크에 비해 높게 나타났는데, 이는 “남북정상회담”이 개최된 연도와 남북공동선언을 합의한 일자에 대한 검색이 많이 이루어졌기 때문이다. 이처럼 이용자 관심사가 높은 외교 행사는 관련 기록물의 연계를 통해 이용자가 관련 기록에 쉽게 접근할 수 있도록 하는 방안이 필요하다.

2.3.4 행사

행사 분야의 검색 빈도 1위 검색어는 “연설문”으로, 그림 <2.d>는 “연설문”의 동시 출현 키워드를 시각화한 “연설문” 워드 네트워크이다. “연설문” 네트워크에 등장하는 단어는 인명(■), 기록(■), 행사(■), 업무(■), 직책(■), 지명(■), 회차(■), 연월일(■), 기타(■)의 총 9개 군집으로 구분할 수 있다. 그 중 “연설문” 네트워크에서 가장 많이 등장한 인명(■) 범주에 등장하는 단어는 ‘김대중’, ‘김영삼’, ‘박정희’, ‘문재인’, ‘노무현’, ‘박근혜’ 등으로 모두 대한민국의 전 대통령 성명이었으며, 이는 해당 인물들의 연설문에 대한 이용자의 관심이 높음을 함의한다. 또한, 두 번째로 많은 비중을 차지하는 군집인 기록(■) 군집의 ‘연설문’, ‘연설’, ‘연설문집’ 등의 단어를 통해 행사 연설문에 대한 이용자 관심이 높음을 확인할 수 있었다. 따라서 이용자가 주로 찾는 정기 행사에 대한 대통령 연설문을 행사 유형별로 묶어 컬렉션 단위로 관리하는 방안을 고려해 볼 수 있을 것이다.

2.3.5 사건

그림 <2.e>는 사건 분야의 1위 검색어인 “5.18”의 동시 출현 키워드로 제작한 워드 네트워크이다. “5.18” 워드 네트워크는 “5.18” 키워드를 포함하는 검색어를 도출하여 동시 출현 키워드를 파악해 제작한 것으로 총 31개 단어, 50개 단어 쌍으로 구성되어 있다. “5.18” 네트워크에 등장하는 단어는 기록(■), 사건(■), 인명(■), 행사(■), 지명(■), 직책(■), 업무(■), 회차(■), 연월일(■)의 총 9개 군집으로 구분할 수 있다. “5.18” 네트워크에서 가장 많이 등장한 단어 군집은 기록(■) 군집으로 전체의 26.67%를 차지했으며, ‘연설문’, ‘담화문’, ‘기념사’ 등 연설 기록에 대한 관심이 높은 것으로 나타났다. “5.18” 네트워크의 행사(■) 군집에 속하는 단어로는 ‘기념식’, ‘기념일’, ‘기념’이 있었으며, 기록(■) 군집에서 ‘기념사’를 비롯해 연설 기록 관련 단어가 다수 등장한 것을 고려하면 “5.18 기념 연설문을 찾고 싶다”라는 이용자의 정보요구가 있음을 추측할 수 있다. 이처럼 “사건” 분야는 역사적으로 중요한 사건과 관련된 대통령 기념사를 찾는 이용자의 관심사가 뚜렷한 만큼 사건 기념 연설문만을 모아 제공하는 서비스가 마련된다면 이용자의 만족도 향상에 도움이 될 것이다.

3. 대통령기록포털 검색 서비스 품질 점검 테스트

2023년 대통령기록관에서 실시한 포털 만족도 조사 결과, 검색 서비스 관련 문항만의 수치적 만족도는 평균 90.1점으로 전반적으로 높았다. 그러나 만족도 개선과 관련된 자유 서술형 답변으로는 이와는 상반되게 “검색 정확도가 떨어진다”라는 답변과 “원하는 콘텐츠를 찾을 수 있는 기능 추가해야 한다”라는 의견도 수렴되었다. 특히, 포털 이용자의 ‘이용 목적’을 살펴보면 ‘기록물 검색’을 위해 포털을 이용한다는 이용자가 35.14%로, 이용자 중 상당수가 포털 검색 서비스를 통한 대통령 기록물의 검색을 위해 대통령기록포털을 방문하는 것으로 나타났다. 이는 현재 대통령기록포털에서 제공하고 있는 검색 도구(“기록물 검색”)가 이용자의 정보요구를 충족할 수 있는지 파악하는 것이 매우 중요함을 의미한다. 따라서 본 장에서는 현재 서비스되고 있는 대통령기록포털이 이용자 정보요구를 얼마나 잘 충족시키고 있는지를 파악하여 디지털 기록 큐레이션을 통해 개선할 수 있는 요소를 찾고자 한다. 이를 위해 상기 과정을 통해 파악한 대통령기록포털 이용자의 정보요구가 반영된 질의셋에 기반하여 대통령기록포털 검색 서비스의 정량적 품질 점검 테스트를 정확도를 중심으로 수행하였다.

https://doi.org/10.14404/JKSARM.2024.24.3.023

검색 서비스의 품질 점검 테스트를 위해서는 문서집합과 질의셋, 그리고 그에 따른 정답셋으로 구성된 평가셋 구축이 선행되어야 하며, 평가셋은 이용자 요구가 반영되어야 한다(나정호 외, 2022). 본 연구는 2024년 5월 기준 대통령기록포털의 검색기를 통해 탐색 가능한 디지털 형태의 문서를 문서집합[3]으로 삼았다. 또한, 이용자의 정보요구를 반영한 질의어 선정을 위해 대통령기록포털의 검색 빈도 기준 상위 검색어를 바탕으로 이용자가 실제 자주 사용하는 질의어와 그 검색 의도가 무엇인지를 탐색하여 질의어 목록을 작성하고 10명의 동료 평가[4]를 통해 검증받았다. 동료 평가는 검색 성능 평가 결과의 신뢰성을 확보하기 위해 기록관리 분야에 대한 배경지식이 있는 5명과 일반인 5명으로 나누어 평가를 진행하였다.

한편, 질의에 대한 정답셋 구축을 위해서는 대통령기록포털의 색인 대상 기록물 전체를 추적해야 하나, 이에 대한 외부 접근이 원천적으로 불가해 완전한 정답셋을 구축하지 못하였으며 검색 결과가 이용자 정보요구를 만족하는지에 대한 정답여부만 평가할 수 있었다. 따라서 본 연구에서는 최종적으로 선정한 질의어 목록에 대한 검색 결과의 채점만을 진행하였으며, 그 결과 재현율(recall) 관련 지표는 측정이 불가, 정확도와 관련된 정확률(precision)을 비롯해 검색 효율을 측정하는 MRR(Mean Reciprocal Rank)을 평가 지표로 측정하였다.

3.1 질의어 목록

대통령기록포털의 검색어 로그 분석 결과, 이용자들이 실제 포털 검색창에 입력한 검색어는 1~2단어로 구성되어 있어 이용자의 명확한 정보요구 파악이 어려웠다. 또, 검색 빈도 기준 상위 100개 검색어 내에 유사한 검색어가 빈번하게 등장하여 동일한 검색 의도를 가진 질의어가 반복된다는 문제가 있었다. 따라서 본 절에서는 검색 빈도 기준 상위 100개 검색어 내 유사한 검색어들을 탐색하고, 이를 대표어로 통합하는 유의어 통합을 진행하여 최종적으로 품질 점검 테스트에 활용할 질의어 목록을 도출하고자 하였다.

<표 9> 유의어 통합을 통한 대표어 선정

전체 순위	검색 빈도	검색어	대표어 검색 빈도	대표어	전체 순위	검색 빈도	검색어	대표어 검색 빈도	대표어
1	10176	연설문	17141	대통령 연설문	18	2157	보고서	3376	보고서
13	3098	연설			32	1219	보고		
20	1860	대통령 연설문			19	2120	5.18	3850	5.18
49	885	대통령연설문			31	1240	광주		
81	607	연설기록			109	490	광주사태		
102	515	대통령 연설			23	1665	취임식	2247	취임식 명단
4	8087	김영삼	8566	김영삼 대통령	88	582	취임식 명단		
115	479	김영삼대통령			29	1334	새마을	2308	새마을 운동
8	5763	취임사	7788	대통령 취임사	45	974	새마을운동		
33	1146	대통령 취임사			36	1069	6.25	1600	6.25
51	879	취임			100	531	한국전쟁		
10	4917	대통령	5569	역대 대통령	63	748	식단	1811	구내 식당
74	652	역대 대통령			86	590	구내식당		
					116	473	식당		

<표 9>는 검색 빈도 상위 100위권 검색어에서 도출한 유의어와 각 유의어 집단을 대표하는 대표어를 정리한

3) 대통령기록포털 기록물목록 검색의 총 검색 대상 문서 수는 검색 성능 평가 당시인 5월 기준 3,327,031개였으나, 논문 투고 시점인 7월 기준 3,446,934개로 증가함

4) 기록관리 전문요원 1명, 기록관리 대학원 석사과정 4명, 일반인 5명

https://doi.org/10.14404/JKSARM.2024.24.3.023

것으로, 총 27개의 유의어를 통합하여 10개의 대표어를 선정한 결과이다. 27개의 유의어 가운데 유의어 통합 과정을 통해 선정한 10개 대표어에 해당하지 않는 17개의 검색어는 대표어로 통합 후, 질의어 목록에서 제외하였으며, 그 결과 기존 순위 117위까지의 질의어로 구성된 총 100개의 질의어를 도출하였다. 도출한 100개 질의어는 검색 빈도 순위를 기준으로 'QL-nnn' 형식의 질의어 번호를 부여하였으며, 대표어의 검색 빈도는 상위 100위권 내에서 유사한 검색 의도를 가진 검색어가 많이 등장할수록 이에 대한 이용자의 관심이 높다는 점을 고려하여 유의어의 검색 빈도 합으로 설정하였다.

질의어를 품질점검 테스트에 활용하기 위해서는 질의어에 대한 검색 의도가 필수적으로 요구된다. 따라서 본 연구는 검색어 로그 분석을 통해 구체적인 이용자의 정보요구를 파악했거나, 질의어 자체만으로 이용자의 정보요구를 추측할 수 있는 경우에는 이를 그대로 검색 의도로 활용하였다. 질의어만으로 검색 의도 파악이 어려운 경우에는 대통령기록포털 검색어 로그 중 질의어를 포함하는 검색어 목록을 도출한 후, 목록의 상위 10위 내에 등장한 질의어와 동시에 출현 키워드 중 워드 네트워크 분석 과정에서 도출한 대통령기록포털의 질의어 세부 군집 11개-1) 기록, 2) 업무, 3) 지명, 4) 인명, 5) 행사, 6) 사건, 7) 직책, 8) 연월일, 9) 회차, 10) 법률, 11) 산업-에 해당하는 단어를 한 개 이상 포함하여 이용자의 정보요구를 유추함으로써 질의어의 검색 의도를 구체화하였다.

예를 들어, 질의어가 단순히 "대통령 연설문"인 경우, 앞서 행사 분야의 워드 네트워크 분석을 통해 파악한 이용자의 정보요구인 "추가경정 예산안 제출에 즈음한 대통령 시정 연설문을 찾고 싶다"를 검색 의도로 삼았다. 질의어 "김영삼 대통령"의 경우, 상위 10위 내에 등장한 동시 출현 키워드 중 검색 빈도 순위가 가장 높은 키워드가 '취임사'이므로 해당 키워드를 포함하여 "김영삼 대통령의 취임사를 찾고 싶다"라는 이용자의 검색 의도를 도출하였다. 해당 과정을 통해 유추한 검색 의도는 동료 평가 결과, 다수의 평가자에 의해 적절하다는 응답을 받았다.

품질 점검 테스트 대상 질의어 목록

질의어번호	전체 순위	검색빈도	주제분야	질의어	검색의도
QL-001	1	17141	행사	대통령 연설문	추가경정 예산안 제출에 즈음한 대통령 시정 연설문을 찾고 싶다.
QL-002	2	9919	인명 및 직책	노무현	노무현 대통령의 연설문 중 취임사를 찾고 싶다.
QL-003	3	8566	인명 및 직책	김영삼대통령	김영삼 대통령의 취임사를 찾고 싶다.
QL-096	96	488	정책 및 업무수행	정책	이명박 정부의 대북정책을 찾고 싶다.
QL-097	97	482	정책 및 업무수행	사면	대통령 특별사면 인사처리지침을 찾고 싶다.
QL-098	98	480	정책 및 업무수행	4대강	이명박 대통령의 4대강 살리기 사업 자료를 찾고 싶다.
QL-099	99	480	정책 및 업무수행	원호	군사원호법과 이에 대한 개정령을 찾고 싶다.
QL-100	100	469	정책 및 업무수행	소방	소방법 시행령 중 개정령을 찾고 싶다.

<그림 3> 품질 점검 테스트 대상 질의어 목록

<그림 3>은 최종 선정한 질의어 100개와 이에 대한 검색 의도를 구체화하여 제작한 품질 점검 테스트 대상 질의어 목록의 일부이다. 질의어 목록은 "정책 및 업무수행" 분야 질의어 32개, "인명 및 직책" 15개, "기록관리" 2개, "행정박물 및 매체" 7개, "기관 및 단체" 9개, "외교" 12개, "행사" 10개, "사건" 12개, "기타" 1개의 총 100개 질의어로 구성되어 있다. 그 중 앞서 중요하게 다루었던 5개 주제 분야("정책 및 업무수행", "인명 및 직책", "외교", "행사", "사건")에 속하는 질의어가 전체의 81%를 차지하여 여전히 이용자 관심이 5개 주제 분야에 집중되어 있음을 확인하였다.

https://doi.org/10.14404/JKSARM.2024.24.3.023

3.2 기록물목록 검색 성능 평가 결과

대통령기록포털의 품질 점검 테스트는 2024년 5월 대통령기록포털의 검색 서비스 중 가장 기본적인 검색 도구인 '기록물목록 검색'을 대상으로 진행되었으며, 상기 제작한 질의어 목록을 기준으로 검색의 품질을 평가하였다. 품질 점검 테스트는 일반적으로 이용자가 필요한 정보를 찾을 때, 검색 결과에서 상위 세 번째 페이지까지 확인하는 경향이 있다는 점을 고려하여 검색 결과 중 상위 30위 내에 검색 의도에 맞는 문서가 등장할 때까지 검색을 수행하였으며, 검색은 최대 3차까지 수행하였다(이상진, 정재승, 2011). 1차 검색은 질의어를 그대로 사용하여 검색을 진행하였으며, 1차 검색에서 30위 내에 검색 의도에 맞는 문서가 등장하지 않았을 경우, 검색 의도를 고려하여 검색어를 구체적으로 변경하거나 검색 조건을 추가하여 2차, 3차 검색을 실시하였다. 이후 검색 결과 등장한 문서의 제목을 기준으로 검색 의도와 부합하는지를 채점하였으며, 이용자 관점에서 일반적인 검색 품질을 평가하기 위하여 검색 성능 평가 과정에 대한 동료 평가를 진행하였다. 동료 평가는 10명의 평가자에게 검색의도의 적절성, 검색어 선정의 적절성, 맞는 문서 채점 기준의 적절성에 대해 묻는 방식으로 진행되었으며, 총 100개 질의어를 대상으로 각 질의어당 2명의 인원이 교차 검증을 하여 결과의 타당성을 검증하였다.

<표 10>은 총 100개 질의어에 대해 정확률, MRR 등의 성능 평가 결과를 정리한 것이다. 성능 평가 결과, 질의어 목록의 전체 질의어 100개 중 1차 검색 만에 맞는 문서를 제시한 질의어의 개수는 49개, 검색 의도를 고려하여 검색어를 변경, 검색 조건을 추가하여 3차 검색까지 진행했을 때 맞는 문서를 제시한 질의어의 개수는 97개로, 3차 검색 내에 대부분의 질의어에 대한 맞는 문서를 찾을 수 있었다.

<표 10> '기록물목록 검색' 성능 평가 결과

	전체 질의어 수	1차 맞는 문서 제시 질의 수(개)	3차까지 맞는 문서 제시 질의 수(개)	P@ 1차 (30위)	P@ 3차 (30위)	P@ 10위 (3차)	P@ 5위 (3차)	MRR (5)	평균 결과 제시 수 (건)	맞는 문서 첫 등장 평균 순위
기록물 목록 검색	100	49	97	0.123	0.403	0.408	0.436	0.500	8,509	5.24

* P@1차, P@3차: 각각 1차 검색 정확률과, 3차까지의 정확률
* P@10위, P@5위: 각각 검색 결과의 상위 10위와 5위까지의 정확률
* MRR(5):검색 결과의 상위 5위까지의 MRR

기록물목록 검색의 정확률 중 1차 검색까지의 정확률(P@1차)과 3차 검색까지의 정확률(P@3차)은 각 차시까지의 검색 결과 중 상위 30위까지의 문서를 대상으로 측정한 결과이며, 10위까지의 정확률(P@10위)과 5위까지의 정확률(P@5위)은 3차 검색까지의 검색 결과 중 각 순위까지의 결과를 대상으로 측정한 값이다. '기록물목록 검색'의 3차 검색까지의 검색 정확률은 0.403으로, 이는 기록물목록 검색을 최대 3차까지 진행했을 때, 검색 결과 상위 30위까지 등장한 문서 중 40%가 이용자의 검색 의도에 부합하는 문서임을 의미한다. 한편, 1차 검색까지의 정확률은 0.123으로 3차 검색까지의 정확률에 비해 매우 떨어지는 것을 확인할 수 있다. 이는 1차 검색어로 입력한 질의어가 검색 의도를 구체적으로 표현하지 못하여 검색 결과 상위 30위 내에 이용자의 검색 의도와 맞지 않는 문서가 다수 검색되었기 때문으로 추정된다. 앞서 대통령기록포털 만족도 조사 결과 제기된 '검색정확도가 떨어진다'는 이용자 의견 또한 질의어가 검색 의도를 충분히 반영하지 못했기 때문에 발생한 문제일 수 있다.

https://doi.org/10.14404/JKSARM.2024.24.3.023

‘기록물목록 검색’의 3차 검색까지의 검색 결과를 기준으로 측정한 상위 10위까지의 정확률은 0.408, 상위 5위까지의 정확률은 0.436으로 상위 30위까지의 정확률인 0.403과 유사하거나 약간 높은 수준으로 나타났다. 이는 하위 검색 결과의 적합도가 낮음을 의미한다. 또한 정보요구를 얼마나 쉽게 충족시킬 수 있는지에 대한 효율을 측정한 MRR은 검색 결과 중 맞는 문서가 첫 등장한 순위의 역수를 평균 낸 것이며 숫자가 높을수록 좋은 성능임을 의미하는데, 기록물목록 검색의 상위 5위까지의 MRR은 0.500 수준으로 나타났다. 이는 맞는 문서의 첫 등장 순위가 5.24인 것을 고려하면, 평균적으로 6위 내에 맞는 정답 문서가 하나 이상 있을 확률이 높음을 의미한다.

분석 결과를 종합해 보면, 대통령기록포털의 ‘기록물목록 검색’은 3차 검색까지 수행했을 때 맞는 문서를 하나 이상 찾을 확률이 97%로 매우 높지만, 정확률은 0.403 수준으로 3차까지 검색 결과 상위 30위까지의 문서 중 40%만 이용자의 검색 의도에 부합하는 기록물인 것으로 나타났다. 특히 1차 검색까지의 정확률이 0.123으로 3차 검색까지의 정확률에 비해 매우 저조하게 나타났는데, 이는 대통령기록포털의 이용자들이 실제 입력하는 질의어가 검색 의도를 충분히 반영하지 못하고 있기 때문이다. 따라서 정확률 상승을 위해서 검색어 자동 추천 기능의 개선과 같이 이용자가 검색 의도에 적합한 구체적인 검색어를 입력할 수 있도록 하는 방안을 마련할 필요가 있다.

3.3 주제 분야별 기록물목록 검색 성능 평가 결과

기록물목록 검색의 검색 성능 평가 결과를 보다 상세히 분석하기 위해 이용자의 관심도가 높은 주제 분야인 "정책 및 업무수행", "인명 및 직책", "외교", "행사", "사건"의 5개 주제 분야별로 나누어 살펴보았다. <표 11>은 5개 주제 분야별 기록물목록 검색의 성능 평가 결과로 맞는 문서 제시 수, 정확률, MRR 등을 각 주제 분야별로 정리한 것이다.

<표 11> 주제 분야별 성능 평가 결과

	전체 질의어 수	1차 맞는 문서 제시 질의 수 (개)	3차까지 맞는 문서 제시 질의 수 (개)	P@ 1차	P@ 3차	P@ 10위	P@ 5위	MRR (5)	평균 결과 제시 수 (건)	맞는 문서 첫 등장 평균 순위
정책 및 업무수행	32	16	32	0.154	0.497	0.496	0.525	0.594	2,126	4.44
인명 및 직책	15	4	14	0.049	0.293	0.313	0.347	0.471	45,272	5.57
외교	12	6	12	0.089	0.344	0.392	0.417	0.600	2,110	4.58
행사	10	7	10	0.277	0.433	0.390	0.360	0.470	1,532	6.10
사건	12	6	12	0.103	0.532	0.508	0.500	0.638	479	4.50

3.3.1 정책 및 업무수행

“정책 및 업무수행” 분야는 질의어 목록에서 가장 많은 질의어가 속한 주제 분야로 총 32개의 질의어가 이에 해당한다. 그 중 16개의 질의어는 1차 검색만으로 검색 의도에 부합하는 문서를 찾을 수 있었고, 3차 검색까지 진행했을 때 32개 질의어 전부 하나 이상의 맞는 문서를 찾는 데에 성공했다.

“정책 및 업무수행” 분야 질의어는 3차 검색까지의 정확률이 높은 편에 속하며, 5위까지의 정확률 값은 0.525 수준으로 5개 주제 분야 가운데 가장 높게 나타났다. “정책 및 업무수행” 분야의 질의어 중 ‘보고서’, ‘헌법’, ‘백서’, ‘관보’, ‘회의록’ 등 기록의 유형을 나타내는 질의어가 다수 존재하는 것을 감안하면 해당 키워드를 포함하여 검색하였을 때, 포털 검색 시스템 내에서 검색 결과로 제시하는 기록물의 유형이 한정되면서 정확률이 상당히 올라간

https://doi.org/10.14404/JKSARM.2024.24.3.023

것으로 추정된다. 이는 대통령기록포털에서 '기록물목록 검색' 시, 이용자가 기록 유형을 포함하여 검색할 수 있도록 유도한다면 검색의 정확률을 높일 수 있음을 시사한다.

3.3.2 인명 및 직책

"인명 및 직책" 분야의 질의어는 총 15개로 1차 검색까지 진행했을 때, 맞는 문서 수가 4개, 1차 검색까지의 정확률이 0.049로 5개 주제 분야 중 가장 낮게 나타났으며, 3차 검색까지의 정확률도 0.293으로 매우 낮은 성능을 보였다. "인명 및 직책" 분야의 평균 결과 제시 수가 45,272건으로 다른 4개 분야와 비교해 압도적으로 높은 것을 고려하면, 검색어가 인명일 때, 해당 인물과 관련된 모든 기록물이 검색되면서 검색 결과 제시된 문서 수가 증가하여 자연스럽게 정확률이 떨어진 것으로 추정된다. 따라서 검색 정확률의 향상을 위해서는 이용자가 질의를 보다 구체적으로 작성할 수 있도록 도와 검색 결과 제시 문서의 범위를 축소시키는 방안을 고려해야 할 것이다.

또, "인명 및 직책" 분야는 5개 주제 분야 가운데 유일하게 3차까지 맞는 문서를 제시하지 못한 질의어 1개가 포함되어 있다. 해당 질의어는 '윤보선'으로 '윤보선 대통령의 취임사를 찾고 싶다'라는 검색 의도를 가지고 검색을 수행하였으나 3차 검색 내에 맞는 문서를 찾을 수 없었다. 이는 해당 질의어에 대한 맞는 문서가 '기록물목록 검색' 내에 존재하지 않기 때문으로, 3차 검색어를 이용해 '웹사이트 검색'을 수행하였을 때, 기록컬렉션 중 연설기록 페이지에서 제공 중인 윤보선 대통령의 취임사를 찾을 수 있었다. 이처럼 이용자가 자신이 찾는 기록물의 위치를 정확히 알지 못하는 상황에서 검색을 수행할 경우, 검색 의도에 부합하는 문서를 찾기까지의 과정이 복잡해져 검색에 어려움을 겪을 수 있다. 따라서 기록물의 위치에 관계없이 포털 내 모든 기록물을 대상으로 검색할 수 있는 검색 도구를 추가하여 이용자의 정보 접근을 용이하게 해야 한다.

3.3.3 외교

"외교" 분야 질의어에 대한 성능 평가 결과, 맞는 문서 첫 등장 평균 순위가 4.58위, 5위까지의 MRR이 0.600으로 평균보다 높은 수치를 보였다. 이는 기록물목록 검색에서 "외교" 분야 질의어를 검색했을 때, 상위 5위 내에 맞는 문서가 하나 이상 나올 확률이 상당히 높음을 의미한다. 그러나 "외교" 분야 질의어에 대한 성능은 3차 검색까지의 정확률이 0.344 수준으로 나타났으며, 10위까지의 정확률, 5위까지의 정확률이 각각 0.392, 0.417로 '기록물목록 검색'의 전체 성능에 비해 다소 저조하게 나타났다. 외교 분야의 질의어가 모두 한 단어로 구성된 짧은 질의어이며, 특히, '월남', '북한', '미국' 등 국가명과 관련된 질의어가 그 자체만으로는 이용자의 검색 의도를 파악하기 어려운 일반적인 단어인 것을 고려하면, 검색 의도를 충분히 반영하지 못하는 짧은 질의어가 검색 성능에 부정적인 영향을 미친 것으로 추정된다. 따라서 이용자가 검색어를 보다 구체적으로 입력할 수 있도록 보조하는 방안을 마련하여 이용자가 적은 횟수의 검색으로 검색 의도에 부합하는 검색 결과를 제공받을 수 있도록 해야 할 것이다.

3.3.4 행사

"행사" 분야 질의어는 1차 검색만으로 맞는 문서를 제시한 질의어의 수와 1차 검색까지의 정확률이 5개 주제 분야 중 가장 높은 것이 특징이다. 1차 검색까지의 정확률을 통해 "행사" 분야에서 1차 검색어로 입력한 질의어가 다른 4개 주제 분야에 비해 이용자의 검색 의도를 충분히 반영하고 있음을 유추할 수 있다. "행사" 분야의 질의어를 살펴보았을 때, 1차 검색에 성공한 "행사" 분야 질의어는 '대통령 연설문', '대통령 취임사', '취임식 명단', '신년사', '축사' 등과 같이 기록 유형을 포함한 검색어였다. 분석 결과, "정책 및 업무수행" 분야의 성능 평가 분석 결과와 마찬가지로 기록 유형을 포함해 검색 시, 검색 정확률이 상승함을 유추할 수 있다.

"행사" 분야에서 맞는 문서가 처음으로 등장한 순위의 평균값이 6.10위로 5개 주제 분야 중 가장 낮게 나타난

https://doi.org/10.14404/JKSARM.2024.24.3.023

것은 다른 주제 분야 질의어에 비해 검색어를 구체화하기 전인 1차 검색에서 맞는 문서를 하나 이상 찾은 질의어 수가 많았기 때문이다. 1차 검색은 가장 포괄적인 범주의 질의어로 검색을 수행하므로 맞는 문서가 등장하더라도 검색 결과 상위에 제시되지 않을 확률이 높다. 따라서 이용자 만족도 향상을 위해서는 외교 분야와 마찬가지로 이용자가 처음부터 구체적인 검색어를 입력할 수 있도록 유도하는 방안을 마련할 필요가 있다.

3.3.5 사건

"사건" 분야 질의어에 대한 성능 평가 결과를 살펴보면, 3차 검색까지의 정확률이 0.532로 5개 주제 분야 가운데 가장 가장 높게 나타났다. "사건" 분야의 평균 결과 제시 수가 479건으로 5개 주제 분야 가운데 가장 적은 것을 감안하면, "사건" 분야 질의어의 정확률이 높게 나타난 것은 대통령기록포털의 "기록물 검색"에서 질의어와 관련된 사건 기록물이 적게 검색되기 때문임을 짐작할 수 있다. 이는 검색 결과로 제시되는 기록물의 범위가 협소할수록 정확률이 올라간다는 의미이다. 따라서 이용자의 검색 의도와 관련이 있는 범위의 기록물만을 모아 이용자에게 제공하기 위한 방안을 고려해 보아야 한다.

정확률 향상을 위해서는 이용자가 질의어를 구체적으로 입력하는 방법과 더불어 세분화된 주제검색을 도입하는 방안도 고려해 볼 수 있다. 주제검색을 도입할 경우, 이용자는 검색어 입력에 대한 고민 없이 원하는 주제 분야 또는 기록 유형에 대한 기록물만을 제공받고, 그 안에서 맞는 문서를 탐색할 수 있으므로 이용자 편의와 검색 정확률을 모두 향상시킬 수 있을 것이다.

4. 시사점

상기한 3가지 관점의 검색어 로그 분석 결과를 통해 도출된 대통령기록포털의 디지털 기록 큐레이션 서비스 개발 시 우선 고려해야 할 시사점은 다음과 같다.

4.1 주제 분석

주제 분석 결과, 각 주제 분야별로 주로 등장하는 키워드들의 집합과, 대통령기록포털에서 해당 주제 분야에 대해 제공하고 있는 서비스 중 컬렉션 및 콘텐츠의 문제점, 이를 통해 파악할 수 있는 분야별 시사점을 정리하였다.

먼저, "정책 및 업무수행" 분야의 경우, 분석 결과를 종합해 보면 '국무회의'를 비롯하여 '국정백서', '국정과제', '공약' 등 정책 관련 기록물을 찾는 이용자가 다수 존재하나, 현재 대통령기록포털은 해당 기록물에 대한 별도의 서비스를 제공하고 있지 않은 것으로 나타났다. 따라서, 특정 주제와 관련하여 이용자의 수요가 높은 여러 기록물을 연계하여 제공하거나, 정책변화를 시각화하여 보여주는 서비스를 제공하는 등 이용자 관심사에 부합하는 서비스를 제공할 수 있는 방안을 모색해야 할 것이다.

"인명 및 직책" 분야의 경우에는 대통령 이름과 함께 영부인의 이름이 등장한 점이 특징적이었으나, 대통령기록포털에서 영부인 관련 기록물을 적극적으로 서비스하고 있지 않았다. 영부인 기록물에 대한 이용자의 수요가 높은 만큼, 영부인의 연설문과 편지, 사진, 선물 등 영부인 관련 기록물을 인물별, 기록 유형별, 연도별 등 다양한 기준에 따라 체계적으로 관리하여 영부인 관련 기록물을 찾는 이용자의 관심사에 맞는 서비스를 제공할 필요가 있다.

"외교" 분야의 경우, 남북정상회담에 대한 이용자의 관심사가 높게 나타났다. 해당 분야와 관련하여 대통령기록포털에서 '정책기록' 콘텐츠 중 '외교' 페이지를 통해 외교 관련 기록물을 제공하고 있으나, 이승만, 박정희 대통령

https://doi.org/10.14404/JKSARM.2024.24.3.023

시기의 기록물만을 한정적으로 제공하고 있었다. 따라서 대통령기록포털 소장 외교 기록물 전체로 대상 서비스를 확대하고, 이용자가 자주 찾는 주제에 대해 보다 쉽게 접근할 수 있도록 패싯 분류체계 개발과 같은 방법을 통해 이용자의 편의를 개선할 필요가 있다.

“행사” 분야 검색어로는 대통령 연설 기록물을 찾는 이용자 검색어가 다수 등장하였으나, 대통령기록포털에서 3.1절 기념사, 광복절 축사와 같이 매년 반복적으로 진행되는 행사에 대한 연설문을 연결해 제공해 주지 않고 있었다. 따라서 이용자가 특정 행사 관련 연설문을 찾을 때, 현재까지 축적된 해당 행사의 연설문을 한 곳에서 모두 볼 수 있도록 정기적인 행사에 대한 역대 대통령의 연설문을 연계하여 관리하는 방안을 고려해야 할 것이다.

“사건” 분야의 경우, 민주화운동 관련 기록물에 대한 관심이 높게 나타났으나, 현재 대통령기록포털은 사건 관련 기록만을 모아 제공하는 서비스가 마련되어 있지 않았다. 사건 기록물은 역사적으로 가치 있는 기록물인 만큼 기록의 홍보와 이용 활성화를 위해 사건별로 기록물을 분류하여 관련 기록을 쉽게 찾을 수 있도록 해야 한다.

4.2 워드 네트워크 분석

워드 네트워크 분석을 통해 파악한 주제 분야별 이용자의 정보요구와 시사점은 다음과 같이 정리할 수 있다. “정책 및 업무수행” 분야의 경우 규정의 개정령이나 정책 관련 회의록을 찾는 대통령기록포털 이용자의 다양한 정보요구를 파악할 수 있었으며, 법을 의결하는 국무회의의 특성상 법률과 산업 군집의 키워드가 등장했다. 해당 분야는 이용자의 정보요구가 다양하게 나타나는 한편, 이용자들이 주로 검색하는 검색어가 뚜렷한 만큼, 주요 검색어와 연관된 기록물들을 연계해 이용자에게 제공하는 방안이 필요할 것이다.

“인명” 분야는 “대통령의 연설문”을 찾고 싶다는 이용자의 정보요구가 가장 뚜렷했으며, 이외에도 대통령의 업무나 외교와 같이 폭 넓은 주제 분야에 대한 기록물을 찾는 이용자가 다수 나타났다. “인명”에 대한 이용자의 정보요구가 매우 광범위하게 나타나는 만큼, 대통령과 관련한 다양한 주제 분야의 기록물을 한 곳에서 볼 수 있도록 조직하여 이용자에게 서비스할 수 있는 방안을 고려해야 한다.

“외교” 분야의 검색어를 검색한 이용자들의 정보요구 중 가장 대표적인 것은 특정 시기 발생한 외교 행사와 관련된 기록물을 찾고 싶다는 것이었으며, 외교 행사와 관련하여 사진, 연설문, 회의록, 백서 등 다양한 유형의 기록물에 대한 관심이 높은 것을 확인할 수 있었다. 따라서 이용자들이 주로 찾는 외교 행사와 관련된 기록물을 연계해서 제공하되, 이용자가 원하는 기록물을 쉽게 찾을 수 있도록 접근점을 확대하는 방안이 필요하다.

“행사” 분야는 특정 행사 자체보다는 반복적으로 발생하는 행사의 대통령 연설문에 대한 이용자의 관심이 높음을 확인하였다. 따라서 이용자의 정보요구를 충족시키기 위해서는 정기 행사에 대한 대통령 연설문을 연계해 관리할 수 있는 방안의 모색이 필요하다.

마지막으로 “사건” 분야는 역사적 사건에 대한 대통령 기념 연설문을 찾는 이용자가 다수 존재하는 만큼, 대통령 연설문 중 사건 기념 연설문을 모아 이용자에게 제공해 주는 서비스가 마련되어야 할 것이다.

4.3 검색 서비스 품질 점검 테스트

대통령기록포털 품질을 평가한 결과, 가장 시급한 문제점으로는 이용자 질의 그대로를 검색어로 사용한 1차 검색 결과의 정확률이 매우 낮다는 점이다. 이는 질의 대부분이 매우 짧고 광범위한 점에서 기인한 것으로, 애초 이용자 질의가 실제 이용자가 원하는 정보요구를 잘 반영하지 못하고 있기 때문이다. 따라서 정확률 상승을 위해서는 검색어 로그 분석을 통해 이용자가 입력한 검색어의 동시출현 키워드, 또는 검색어와 관련 있는 주제의 군집

https://doi.org/10.14404/JKSARM.2024.24.3.023

내 등장하는 키워드를 함께 추천해 주는 방식의 검색어 자동 추천 기능을 활용하여 이용자가 검색의도에 적합한 구체적인 검색어를 입력할 수 있도록 유도해야 할 것이다.

주제 분야별 성능 평가 결과와 이를 통해 파악한 주제 분야별 시사점은 다음과 같다. 먼저 “정책 및 업무수행” 분야의 경우, 기록 유형과 관련된 키워드를 포함하여 검색할 경우, 검색 결과로 제시되는 기록물의 유형이 한정되어 검색 정확률이 높게 나타나는 것을 파악하였다. 따라서 이용자가 검색어 입력 시, 기록 유형을 포함한 검색어를 입력할 수 있도록 유도하여 검색 정확률을 높이는 방안을 고려해 보아야 한다.

“인명 및 직책” 분야 검색어에 대해 대통령 또는 영부인의 이름을 그대로 검색어로 사용하였을 때, 해당 인물과 관련된 모든 기록물이 검색되면서 검색 결과 제시된 문서 수가 증가하여 정확률이 떨어지는 문제가 발생했다. 따라서 검색 정확률 향상을 위해서는 이용자가 질의를 보다 구체적으로 작성할 수 있도록 도와 검색 결과 제시 문서의 범위를 축소시켜야 한다. 또한, 이용자가 본인의 검색 능력에 관계없이 원하는 검색 결과를 얻을 수 있도록 하는 방안 마련이 필요하다.

“외교” 분야의 질의어는 대부분 모두 한 단어로 구성된 짧은 질의어로, 그 자체만으로는 검색 의도를 충분히 반영하고 있지 못해 검색 정확률이 저조하게 나타났다. 따라서 검색어 입력 단계에서 이용자가 검색어를 보다 구체적으로 입력할 수 있도록 보조하는 방안을 마련하여, 적은 횟수의 검색으로 검색 의도에 부합하는 검색 결과를 제공받을 수 있도록 해야 한다.

“행사” 분야는 검색어를 구체화하기 전인 1차 검색에서 맞는 문서를 찾은 질의어 수가 많았으며, 이 경우 질의어가 매우 큰 범주를 포괄하므로 맞는 문서가 등장하더라도 검색 결과 상위에 제시되지 않았다. 때문에 외교 분야와 마찬가지로 이용자가 처음부터 구체적인 검색어를 입력할 수 있도록 유도하는 방안을 마련하여 포털 검색 성능의 향상과 이용자 만족도 상승을 꾀해야 한다.

“사건” 분야는 질의어와 관련된 사건 기록물이 다른 분야 질의어의 검색 결과에 비해 적게 검색되면서 검색 정확률이 상대적으로 높게 나타남을 확인하였다. 따라서 이용자의 정보요구에 부합하는 기록물만을 모아 이용자에게 제공함으로써, 검색 결과 제시 건수를 줄여 검색 정확률을 향상시키는 방안을 마련할 필요가 있다.

5. 결론

본 연구는 역사적으로 중요한 가치가 있는 대통령기록물을 보존하고, 이용을 활성화하기 위한 큐레이션 방안을 제시하기 위해 먼저 대통령기록포털의 검색 로그 분석을 통해 이용자의 정보요구를 파악하는 것을 목표로 진행된 연구이다. 이용자의 정보요구를 파악하기 위해 먼저 시계열 분석을 통해 이상치를 파악, 이를 제거하여 분석 대상 검색어 목록을 도출하였으며, 이용자의 관심이 높은 5개 주제 분야를 대상으로 주제 분석과 워드 네트워크 분석을 수행하여 분야별 이용자의 정보요구를 파악하였다.

대통령기록포털의 주제 분석 결과를 살펴보면, "정책 및 업무수행" 분야는 ‘국무회의’, ‘국정백서’, ‘국정과제’, ‘공약’ 등 정책 관련 기록물에 대한 이용자의 관심이 높게 나타났으며, "인명 및 직책" 분야는 영부인 관련 기록물이 특징적으로 나타났다. "외교" 분야에서는 남북정상회담 관련 기록물을 찾는 이용자가 많았으며, "행사" 분야에서는 대통령 연설 관련 기록물을, "사건" 분야에서는 민주화운동 관련 대통령 기록물에 대한 수요가 높았다.

다음으로 워드 네트워크 분석을 통해 규정의 개정령, 정책 관련 회의록을 찾는 "정책 및 업무수행" 분야의 정보요구를 파악하였다. "인명 및 직책" 분야는 대통령의 연설문과 업무, 외교 기록을 찾는 이용자의 정보요구가 높게

https://doi.org/10.14404/JKSARM.2024.24.3.023

나타났으며, "외교"의 경우, 남북정상회담과 같이 특정 시기 발생한 외교 행사, 또는 이와 관련된 연설문, 사진 등 다양한 유형의 기록물에 대한 관심이 높게 나타났다. "행사" 분야는 매년 반복적으로 발생하는 행사의 대통령 연설문을 찾는 이용자의 정보요구가, "사건" 분야는 사건 기념 연설문을 찾는 이용자의 정보요구가 있었다.

상기 파악한 이용자 정보요구를 바탕으로 대통령기록포털에서 이용자 정보요구에 부합하는 서비스를 제공하고 있는지 확인하기 위해 대통령기록포털 검색 서비스의 품질 점검 테스트를 정확도 관점에서 진행하였다. 그 결과, 대통령기록포털의 "기록물목록 검색"을 통해 100개의 질의어 중 49개의 질의어는 1차 검색만으로 하나 이상의 맞는 문서를 찾을 수 있었으며, 3차 검색까지 진행했을 때 총 97개 질의어가 맞는 문서를 하나 이상 찾을 수 있는 것으로 나타났다. 그러나, 1차 검색까지의 정확률은 0.123, 3차 검색까지의 정확률은 0.403 수준으로 상당히 저조하게 나타났으며, MRR은 0.500 수준으로 평가되었다. 상세한 분석 결과를 얻기 위해 질의어의 주제 분야별로 "기록물목록 검색"에서 검색했을 때의 성능을 평가하여 검색 성능에 영향을 미치는 요소를 파악하였다.

분석 결과를 통해 도출한 시사점을 종합하면 다음과 같다. 대통령기록포털은 첫째, 이용자의 관심이 높은 주제와 관련 있는 기록물의 연계가 부족했다. 특히 정기적으로 발생하는 행사에 대한 연설문의 연계 관리 방안이 필요하다. 둘째, 대통령과 영부인 기록물에 대한 체계적 관리와 더불어 이와 관련된 다양한 주제의 기록물을 한 곳에서 관리해야 하며, 셋째, 이용자의 질의 확장을 돕는 서비스가 필요하다. 넷째, 이용자의 관심 분야 기록물들을 한 곳에서 모아서 관리하고, 이용자에게 제공하기 위한 방안이 마련되어야 한다. 다섯째, 기록의 위치를 모르는 상황에서도 이용자가 기록물을 찾을 수 있도록 검색 시스템을 개선해야 한다. 마지막으로 이용자가 자주 찾는 주제 기록물에 대한 다양한 접근점이 제공되어야 할 것으로 보인다.

본 연구는 대통령기록포털의 검색어 로그 빅데이터를 기반으로 최근 8년간의 대통령기록포털 이용자의 정보요구를 파악하였다는 점에서 차별성이 있다. 특히 주제 분석, 워드 네트워크 분석을 통해 이용자의 정보요구를 보다 세부적이고 다양한 관점에서 파악하였다는 점과 검색 품질을 정량적으로 수치화하였다는 점에서 실효성이 높다. 그러나 전체 문서셋에 대한 정답셋의 부재로 인해 대통령기록포털의 검색 성능 평가 시 재현율 측정이 불가하였던 점이 한계점으로 꼽힌다. 본 연구는 대통령 기록물 대상 디지털 기록 큐레이션을 위한 사전 연구로서, 후속 연구를 통해 보다 구체적인 대통령기록포털의 디지털 기록 큐레이션 방안을 제안할 예정이다. 이는 궁극적으로 이용자 만족도 향상과 기록물의 활용 가치 제고에 긍정적인 영향을 미칠 것으로 예상되며, 향후 대통령기록포털의 서비스 개선을 위한 연구의 기초 자료로 활용되기를 고대한다.

참고문헌

김대원 (2021). 국가기록원 검색어 로그 분석을 통한 기록정보서비스 개발 –큐레이션 서비스를 중심으로–. 석사학위논문, 경북대학교.

나정호, 소현기, 염경록, 이정옥, 오효정 (2022). 국가기록포털 검색서비스 품질 점검을 위한 평가셋 구축 및 현황 분석. 한국기록관리학회지, 22(4), 25–43. https://doi.org/10.14404/JKSARM.2022.22.4.025

백지연 (2020). 국가기록원 질의로그 빅데이터 기반 이용자 검색 만족도 향상 제고 방안. 석사학위논문, 전북대학교.

이상진, 정재승 (2011). 검색엔진 최적화(SEO) 기법을 활용한 전자무역 마케팅 성과에 관한 실증연구. 통상정보연구, 13(1), 3–28. https://doi.org/10.15798/kaici.13.1.201103.3

이해영 (2011). 기록관 웹사이트 검색어의 분석과 활용 – 대통령기록관을 중심으로 –. 한국기록관리학회지, 11(1), 93–112. https://doi.org/10.14404/JKSARM.2011.11.1.093

정광훈, 남영준. (2013). 우리나라 대통령기록물의 분류기준에 관한 연구. 한국문헌정보학회지, 47(1), 419–448.

https://doi.org/10.4275/KSLIS.2013.47.1.419
조민지 (2020). 디지털 기억 시대의 기록과 정보서비스. 기록과 정보·문화 연구, (10), 181-215. https://doi.org/10.23035/kaics.2020.1.10.181
진주영 (2018). 국가기록원 웹사이트의 빅 데이터 분석과 활용. 석사학위논문, 명지대학교.

• 국문 참고자료의 영어 표기
(English translation / romanization of references originally written in Korean)

Jin, Juyeong (2018). Analysis and Utilization of Big Data from the website of National Archives of Korea. Master's thesis, Myeongji University, Korea.

Jo, Min-Ji (2020). Archives and Information Services on the Digital Memory Era. The Korean Journal of Archival, Information and Cultural Studies, (10), 181-215. https://doi.org/10.23035/kaics.2020.1.10.181

Jung, Kwang-Hum, & Nam, Young-Joon (2013). Study on the Classification Guideline for the Korean Presidential Records. Journal of the Korean Society for Library and Information Science, 47(1), 419-448. https://doi.org/10.4275/KSLIS.2013.47.1.419

Kim, Dae-Won (2021). Development of Archival Information Services through Log Analysis of Search Word in National Archives of Korea -Focusing on curation services-. Master's thesis, Kyungpook National University, Korea.

Lee, Sangjin, & Chung, Jason (2011). An Empirical Study on the Marketing Performance of e-Tradeusing Search Engine Optimization. International Commerce and Information Review, 13(1), 3-28. https://doi.org/10.15798/kaici.13.1.201103.3

Na, Jeongho, So, Hyeon-Gi, Yeom, Gyungrok, Lee, Jung-Ok, & Oh, Hyo-Jung (2022). Test Set Construction for Quality Evaluation of NAK Portal's Search Service and the Status Analysis. Journal of Korean Society of Archives and Records Management, 22(4), 25-43. https://doi.org/10.14404/JKSARM.2022.22.4.025

Paek, Jiyeon (2020). A Study on Improving User's Search Satisfaction based on Query Log Big Data of the National Archives of Korea. Master's thesis, Jeonbuk National University, Korea.

Rieh, Hae-Young (2011). Analysis and Utilization of Search Terms in Archival Web Sites: A Case Study of Korean Presidential Archives. Journal of Korean Society of Archives and Records Management, 11(1), 93-112. https://doi.org/10.14404/JKSARM.2011.11.1.093