

# 국가기록원 질의로그 빅데이터 기반 이용자 정보요구 유형 분석\*

## User Information Needs Analysis based on Query Log Big Data of the National Archives of Korea

백지연 (Ji-yeon Baek)\*\*

오효정 (Hyo-Jung Oh)\*\*\*

### 초 록

이용자의 정보요구를 파악하기 위한 다양한 방법 중 로그 분석 방법은 이용자의 실제 검색 행위를 사실적으로 반영하고, 대다수 이용자의 전반적인 이용행태를 분석할 수 있다. 이에 본 연구에서는 국가기록원 웹 포털서비스를 통해 입수된 대량의 질의로그 빅데이터를 기반으로 이용자의 정보요구를 파악하기 위해 1) 질의에 내포된 정보요구 유형별과 2) 검색결과로 제공한 기록 유형별 분석을 진행하였다. 분석에 활용한 질의로그는 국가기록원이 웹을 통해 검색서비스를 제공한 2007년부터 2018년 12월까지, 총 141개월 동안 수집된 월별 상위 100개 질의어 1,571,547개를 대상으로 하였다. 나아가 분석결과를 토대로 이용자 검색 만족도를 향상시킬 수 있는 개선방안을 제안하였다. 본 연구의 결과는 국가기록원 검색 서비스 개선 및 고도화를 위한 구체적이고 실질적 방안을 제시했다는 점에서 의의가 있다.

### ABSTRACT

Among the various methods for identifying users' information needs, Log analysis methods can realistically reflect the users' actual search behavior and analyze the overall usage of most users. Based on the large quantity of query log big data obtained through the portal service of the National Archives of Korea, this study conducted an analysis by the information type and search result type in order to identify the users' information needs. The Query log used in analysis were based on 1,571,547 query data collected over a total of 141 months from 2007 to December 2018, when the National Archives of Korea provided search services via the web. Furthermore, based on the analysis results, improvement methods were proposed to improve user search satisfaction. The results of this study could actually be used to improve and upgrade the National Archives of Korea search service.

키워드: 국가기록원, 정보요구, 질의로그, 빅데이터

national archives of Korea, information needs, query log, big data

---

\* 이 논문은 이 논문은 2019년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2019S1A5B8099507).

\*\* 전북대학교 일반대학원 기록관리학과(yeon87602965@gmail.com) (제1저자)

\*\*\* 전북대학교 문헌정보학과 부교수, 문화융복합아카이빙연구소 연구원(ohj@jbnu.ac.kr) (교신저자)

■ 논문접수일자: 2019년 11월 17일 ■ 최초심사일자: 2019년 12월 18일 ■ 게재확정일자: 2019년 12월 23일

■ 정보관리학회지, 36(4), 183-205, 2019. <http://dx.doi.org/10.3743/KOSIM.2019.36.4.183>

## 1. 서론

2000년대 이후 디지털 시대가 도래하면서 웹 환경이 보편화 되었고, 이에 따라 이용자의 기록이용 행태에도 많은 변화를 가져오게 되었다. 기존에는 이용자가 직접 기록관에 방문하여 원하는 기록을 찾았다면, 이제는 기록관에 방문하지 않아도 언제, 어디서든 시공간에 제약 없이 웹을 통해 기록에 접근하고 활용할 수 있게 되었다(진주영, 2018). 이때 이용자가 웹사이트에서 입력하는 질의어는 이용자가 검색하는 시점에서 가장 관심 있는 주제이며, 정보에 대한 요구를 직접 표출하고 있는 표현이다. 따라서 이용자가 입력하는 질의어의 특징 변화를 분석하고 파악한다면 이용자의 향후 검색행태와 정보요구를 예측하는데 활용될 수 있다. 즉 이용자의 질의어에 내포된 의도를 파악할 수 있다면 이용자에게 더 만족스러운 양질의 검색결과를 제공해 줄 수 있을 것이다(김성희, 2011). 이와 같은 측면에서 기록관에서 웹을 통해 이용자가 접근하고 이용하기 원하는 기록은 어떤 종류와 주제가 될지를 이해하고, 이에 대해 적극적으로 대응하는 것은 매우 중요한 일이다(이해영, 2011).

이용자의 정보요구를 파악하는 방법으로는 설문이나 인터뷰, 관찰과 같은 질적 연구방법과 웹에서의 행적 및 질의로그를 활용하여 분석하는 정보학(informatic)적 방법 등 다양한 방법이 있다(이해영, 2011). 그 중 로그 분석 방법은 이용자의 실제 검색 행위를 사실적으로 반영하고, 대다수 이용자의 전반적인 이용행태를 분석할 수 있으므로, 이용자 계층이 다양하고 이질적인 웹 검색 분야의 연구에 적합한 방

법이다(박소연, 이준호, 2007). 또한 질의로그 분석을 통해 콘텐츠나 연관 질의어를 추천해줄 수 있으며, 개인별 맞춤형 서비스를 제공하거나 웹사이트의 검색 성능을 향상시킬 수 있다(이수상, 위성광, 2009). 따라서 본 연구에서는 기존에 이루어지던 인터뷰나 설문지와 달리 이용자의 실제 정보요구를 분석하기 위한 객관적이며 직접적인 기법인 로그 분석 방법이 적합하다고 판단하였다.

국내 대표 영구기록물관리기관인 국가기록원도 2007년부터 웹을 통해 일반 국민들에게 보존하고 있는 공공기록을 서비스하고 있다. 현재 기록관리 지침 등의 안내에서부터 기록정보콘텐츠 구축, 소장기록 정보 제공, 전시 등의 다양한 서비스를 통해 기록관의 전반적인 활동을 효과적으로 제공하고 있으며(장희정, 2012), 사전연구(오효정, 2019)를 통해 2007년부터 국가기록원 웹 포털사이트를 통해 입수된 방대한 양의 로그가 축적, 월별로 통계 처리되고 있는 것을 확인하였다. 이러한 국가기록원 웹 포털사이트 이용자들이 입력한 질의어를 분석해보고, 특징을 파악하는 것은 그 자체만으로도 의미가 있을 것이며, 이를 통해 이용자 서비스 개선에 활용한 만한 결과를 얻는다면 큰 의미가 있을 것이라 판단된다.

따라서 본 연구에서는 국가기록원의 포털서비스를 통해 입수된 대량의 질의로그 빅데이터로부터 이용자 정보요구의 특징을 파악하고자 한다. 나아가 분석결과를 토대로 이용자 검색 만족도를 향상시킬 수 있는 실질적이고 구체적인 개선방안을 제안하고자 한다.

이를 위해 수행한 세부 연구방법은 다음과 같다. 먼저 국가기록원 웹 포털서비스 현황을

조사하고 질의로그를 수집하였다. 수집 방법은 2019년도에 진행한 연구과제(오효정, 2019)와 정보공개청구를 통해 이루어졌으며, 입수 자료는 2007년 4월부터 2018년 12월까지의 질의로그이다. 질의로그는 날짜와 질의어, 검색 빈도 정보로 구성되어 있으며, 그중 정보의 누락이나 오류가 있는 부분을 제외한 후 분석을 실시하였다. 특히 이용자의 정보요구를 다각도로 파악하기 위해 1) 질의에 내포된 정보요구 특성과 2) 검색결과로 제시된 기록물의 특성을 유형별로 나누어 분석하였다. 마지막으로 유형별 분석결과를 바탕으로 국가기록원 이용자의 정보요구 특징을 파악하고, 이를 토대로 이용자의 검색 만족도 향상을 위한 개선방안을 제안하였다.

본 연구는 다음과 같이 구성되었다. 제2장 선행연구에서는 웹사이트 이용자 검색행태와 관련된 여러 분야의 로그 분석 연구를 살펴보았다. 제3장 연구방법론에서는 국가기록원 질의로그 수집 방법과 기초분석 결과를, 제4장 이용자 정보요구 세부 결과에서는 각 유형별 분석결과와 검색실패 질의어의 특성을 파악하였다. 끝으로 제5장 개선방안에서는 분석결과를 종합하여 도출된 시사점에 따라 국가기록원에서 적용 가능한 실질적 개선방안을 제시하였다.

## 2. 선행연구

웹사이트 이용자의 검색행태와 정보요구를 파악하기 위해 이용자 관련 로그를 분석한 연구는 기록관, 도서관, 포털사이트 등 전반적인 분야에서 다양하게 이루어졌다. 먼저 기록관

웹사이트를 대상으로 이루어진 연구는 이해영(2011), Prom(2011), 이효은(2015), 진주영(2018)이 있다. 이해영(2011)은 기록관 웹사이트 로그 분석결과를 이용자들에 대한 기록정보 서비스에 효과적으로 활용할 수 있는지 알아보려 하였다. 이를 위해 2010년 1월부터 11월까지의 대통령기록관의 로그를 수집하여 분석하였으며, 분석결과를 일반 검색어 분석, 대통령 이름별 분석, 검색어 주제 영역 분석 세 가지로 제시하였다. 해당 분석결과는 콘텐츠 서비스 및 웹사이트 메뉴 설정, 전시 등 다양한 서비스에 활용될 수 있을 것으로 판단되었다.

Prom(2011)은 일리노이 대학교 기록관 웹사이트의 구글애널리틱스를 이용하여 2007년 7월 한 달 동안의 유입경로와 이용자 탐색 행태 등을 분석하였다. 이효은(2015)은 웹 로그 분석 도구인 구글애널리틱스를 민주화운동기념사업회 오픈 아카이브에 2014년 8월 한 달 동안 실제 적용하여 웹 이용자 행태를 분석하고, 도입 현황과 문제점을 제시하였다. 이를 통해 세분화된 이용자 요구사항을 예측하고, 웹서비스 고도화와 기록정보서비스 개선방안에 대해서 논의하였다.

진주영(2018)은 2007년 4월부터 2017년 9월 월별 상위 100개 국가기록원 로그를 크게 외부 유입경로와 검색어로 나눠 분석하였다. 유입경로는 포털과 국가별, 유사기관, 모바일 접속으로 구분하여 분석하였으며, 검색어는 검색 횟수와 등장 횟수 분석, 주제별 분석, 최근 이슈별과 연계 분석으로 세분화하였다. 분석결과를 토대로 국가기록원 웹사이트 이용자 서비스 방안을 제안하였다.

다음으로 유사기관 로그 분석 선행연구로 박

소연, 이준호(2005)는 네이버에 1년 동안 입력된 질의어들의 표본과 각 질의에 대한 클릭 로그에 근거하여 국내 웹 이용자의 검색행태 추이를 분석하였다. 분석결과는 인터넷 검색 포털 업체들의 효과적인 콘텐츠 구축 및 검색 시스템 개발의 기초자료로 활용할 수 있다. 이수상, 위성광(2009)은 2008년 10월부터 12월까지 한국과학기술정보연구원(KISTI)에서 서비스하는 국가과학기술정보센터(National Digital Science Library, 이하 NDSL) 질의로그에 나타나는 이용자들의 네트워크 연결구조를 파악하고, 네트워크 분석 기법을 통해 이용자들의 검색 행위로 나타나는 다양한 특성을 파악하고자 하였다.

이성숙(2012)은 웹 기반 온라인목록 이용자들의 전반적인 검색행태를 파악하기 위해, 2006년부터 2012년까지의 대학도서관 트랜잭션 로그를 분석하였다. 온라인목록의 검색행태를 검색전략과 검색실패의 관점으로 구분하여 조사하였다. 남상은(2017)은 국립중앙도서관의 2015년 1월부터 12월까지의 질의로그를 대상으로 도서관 이용자의 검색 패턴을 분석하는 것을 연구의 목적으로 하였다. 이를 위해 질의로그를 검색 횟수와 검색어 내용, 검색어 형태별로 세부 분석하였으며, 인기 검색어와 실제 이용되는 인기 자료의 목록을 비교하여 자료 이용과 검색의 연관성을 분석하였다.

상기된 바와 같이 이용자 로그 분석에 관련된 연구는 여러 기관에서 다양한 주제와 방법을 통해 수행되었다. 그러나 이용자의 현실적이고 지속적인 검색 양상을 분석하기 위해서는 장기간에 걸쳐 누적된 빅데이터에 기반한 분석이 요구됨에도 불구하고, 대부분의 연구가 단

기간에 수집된 데이터를 대상으로 수행되었다. 또한 이용자의 검색행태가 전에 없이 다양하고 빠르게 바뀌는데도 불구하고 기존 연구는 단순히 검색 횟수나 주제 분석에 주로 치중되었거나 단기간의 부분 분석에 그치고 있었다. 이에 본 연구에서는 국가기록원에서 웹 서비스를 시작한 2007년 4월부터 2018년 12월까지, 근 12년간에 걸쳐 장기간 누적된 질의로그 빅데이터를 기반으로 이용자의 검색행태를 정보요구 유형별(기관명, 인명, 사건)과 검색결과 유형별(일반문서류, 도면류, 사진·필름류 및 녹음·동영상류)로 나누어 다각도로 분석하였다는데 그 차이가 있다.

### 3. 연구방법론

#### 3.1 국가기록원 질의로그 빅데이터

본 연구의 질의로그 수집은 오효정(2019)에서 사용한 데이터와 정보공개청구를 통해 입수한 데이터를 활용하였다. 데이터는 2007년부터 2012년까지는 월별 질의로그 상위 100개(총 1,500개)와 2013년부터 2018년까지는 월별 질의로그 전체 순위를 입수하여 활용하였다. 특히 본 연구에서는 장기간에 걸쳐 국가기록원에 입력된 질의어들의 전반적 유형 분석을 주안점으로 삼았기에 일별 통계보다는 월별 통계를 활용하였으며, 통일성을 위해 분석대상을 2007년 4월부터 2018년 12월까지 상위 100개 질의로그로 선정하였으며, 분석 데이터 통계 정보는 <표 1>과 같다. 총 141개월 동안 수집된 국가기록원 질의로그 전체 개수는 1,571,547개였

다. 그중 중복제거를 한 개수는 4,267개였으며, 총 검색 빈도는 3,920,269회였다.

〈표 1〉 분석 데이터 통계

연구대상 속성	개수
수집 기간	141개월 (2007년 4월~2018년 12월)
전체 질의로그 수	1,571,547개
질의어 수(중복제거)	4,267개
총 검색 빈도	3,920,269회

유형별 세부 분석에 들어가기에 앞서 전체 질의어 상위 100개를 분석하였으며, 그 결과는 〈표 2〉와 같다. 분석결과, 검색 빈도와 출현 개월 수가 가장 높은 질의어는 ‘토지조사부’<sup>1)</sup> (139,214)였다. ‘토지조사부’는 1910년부터 1918년 사이 9년여에 걸친 토지조사사업에서 만들어진 산출물로서 2위 검색어인 ‘관보’(51,536)와 비교해봤을 때 검색 빈도에서 87,678회의 큰 차이를 보인다. ‘토지조사부’ 외에도 ‘위토대장’, ‘토지’, ‘토지대장’ 등과 같이 토지 관련 질의어들이 다수 등장하였으며, 이를 통해 국가기록원 이용자들이 조상이 소유한 땅에 대한 정보를 찾기 위해 토지와 관련된 기록물을 검색하는 것을 알 수 있었다(진주영, 2018).

〈표 2〉의 세부 분석을 위해 질의어 상위 100개를 범주화하여 구분한 결과, 〈표 3〉과 같이 6가지 주제 분야로 나눌 수 있었다. 이렇게 주제 분야별로 질의어들을 묶어 보면, 이용자들의 관심 영역에 대해 좀 더 잘 이해할 수 있기 때문에 콘텐츠 서비스나 전시, 교육 등 일반 서비스 영역의 주제를 정할 때나 홈페이지의 메

뉴를 결정할 때에 도움을 받을 수 있다(이해영, 2011).

분석결과, 총 검색 빈도와 출현 개월 수가 가장 높은 주제는 “기관”으로 질의어 100개 중 56개를 묶을 수 있었다. “기관” 관련 질의어의 총 검색 빈도는 358,294회이며 총 출현 개월 수는 1,085회이었다. “기관”에 이어 두 번째로 많은 비중을 차지하는 주제인 “토지”의 경우 질의어 100개 중 12개를 묶을 수 있었으며 총 검색 빈도는 275,801회로 “기관”의 검색 빈도보다는 낮지만, 총 출현 개월 수는 1,087회로 높았다. “기관”과 “토지”에 해당하는 질의어 개수 대비 총 검색 빈도와 출현 개월 수를 비교해보았을 때, 국가기록원 이용자들이 실질적으로 더 관심 있어 하는 주제는 “토지”임을 알 수 있었다. 이용자들의 관심이 높은 주제인 만큼 “토지” 관련 기록물은 이미 “지적아카이브”와 같은 기록정보콘텐츠로 구축되어 국가기록원 웹사이트에서 이용자들에게 서비스되고 있다. 따라서 기록정보콘텐츠 구축 외에 검색의 정확성과 편리성을 높이는 방안을 추가로 고려할 필요가 있다.

상기한 바와 같이 일차적으로 국가기록원 이용자들의 정보요구를 파악한 결과, 이용자가 요구하는 정보가 “기관”, “인명” 등 유형별로 특징 있게 포착되었다. 또한 프로그래밍을 통해 국가기록원 웹 포털사이트에서 4,267개 질의를 입력한 후, 검색결과 페이지를 수집 및 파싱(parsing)하여 분석한 결과, 주제별로 요구하는 기록물의 형태가 다르게 파악되었다. 예를 들면 “기관” 유형에 해당하는 질의어를 통해 이용자에게 전달되어야 하는 기록물은 그

1) 본 논문에서는 개별 질의어는 ‘ ’로, 특정 주제나 유형을 강조하기 위해서는 “ ”로, 사이트 이동 경로는 [ ]로 표기함

〈표 2〉 전체 질의어 상위 100개

순위	질의어	검색 빈도	출현 개월 수	순위	질의어	검색 빈도	출현 개월 수
1	토지조사부	139,214	142	51	국유재산	5,495	33
2	관보	51,536	138	52	산림청	5,469	22
3	경기도	45,197	51	53	문교부	5,358	28
4	관리	41,294	33	54	상환대장	5,194	83
5	해양	30,613	30	55	올림픽	5,158	29
6	위토대장	27,081	75	56	통상산업부	5,092	13
7	박정희	25,672	138	57	국방부	5,053	20
8	설계	25,637	29	58	병무청	5,039	22
9	보존기간	13,251	119	59	태안군	5,025	26
10	행정자치부 기획관리실 법무담당관	13,050	7	60	금융위원회	5,014	16
11	조의순	12,278	39	61	대검찰청	4,975	19
12	토지	12,160	99	62	충무처의정국	4,941	16
13	일제강제 연행자명부	11,859	82	63	한국전쟁	4,909	98
14	감사원	11,001	44	64	공보처	4,899	17
15	특허청	10,940	22	65	내무부	4,864	19
16	국가보훈처	10,697	22	66	행정자치부	4,778	19
17	최순실	10,606	4	67	원주	4,737	26
18	광주군	10,254	33	68	상공자원부	4,704	12
19	국세청	10,120	22	69	행정자치부 기획관리실	4,502	6
20	충무처	10,015	21	70	국무총리실	4,425	16
21	경찰청	9,990	20	71	동력자원부	4,414	14
22	토지대장	9,975	116	72	노무현	4,368	81
23	조달청	9,948	21	73	체신부	4,357	20
24	관세청	9,579	23	74	6.25	4,348	50
25	교육인적자원부	9,543	19	75	국사편찬위원회	4,279	18
26	식품의약품안전청	9,512	22	76	감찰위원회	4,139	13
27	환경부	8,978	21	77	이승만	4,112	87
28	보건복지부	8,942	25	78	농촌진흥청	4,086	21
29	해양경찰청	8,865	21	79	조선총독부	4,055	83
30	입야조사부	8,782	117	80	국무총리비서실	4,038	18
31	소방방재청	8,486	19	81	국가보안법	4,018	18
32	문화재청	8,480	22	82	외무부	3,990	17
33	지적원도	8,383	110	83	분배농지부	3,932	75
34	외교통상부	8,234	22	84	공정거래위원회	3,833	19
35	기상청	8,224	20	85	중소기업청	3,805	20
36	광주	8,118	39	86	새마을운동	3,799	93
37	통일부	8,022	21	87	교육부	3,645	18
38	여성가족부	8,021	24	88	경제기획원	3,550	19
39	국가인권위원회	7,994	19	89	국정홍보처	3,497	17
40	통계청	7,776	22	90	농림부	3,463	18
41	정책	7,693	29	91	행정중심복합도시건설청	3,395	13
42	625	7,503	43	92	고시	3,384	83
43	서울	7,456	31	93	무주부동산 공고	3,292	53
44	과학기술부	6,816	20	94	건설부	3,219	20
45	대통령	6,484	87	95	지적도	3,162	78
46	독립운동 관결문	6,455	33	96	문화관광부	3,126	18
47	중요민속자료	6,164	14	97	상환 대장	3,090	1
48	법제처	6,095	23	98	국가정보원	3,082	14
49	법무부	5,935	21	99	대전	3,081	14
50	노무*	5,929	112	100	국무회의	3,073	73

〈표 3〉 전체 질의어 상위 100개 주제 분류

주제	질의어	총 검색 빈도	총 출현 개월 수
기관(56개)	행정자치부 기획관리실 법무담당관, 감사원, 특허청, 국가보훈처 등	358,294	1,085
토지(12개)	토지조사부, 위토대장, 토지, 토지대장, 입야 조사부 등	275,801	1,087
사건(9개)	한국전쟁, 새마을운동, 올림픽 등	58,340	544
인명(7개)	박정희, 조의순, 최순실, 노무현, 이승만 등	69,449	548
지역명(6개)	경기도, 광주, 서울, 태안군, 원주, 대전	73,614	187
기록관리 업무(1개)	보존기간	13,251	119
기타(9개)	관보, 국유재산, 중요민속자료, 국가보안법, 정책, 고시 등	175,834	407

기록에서 생산한 문서류, 각종 법률 관련 기록 물이고, “대통령”이나 특정 사건에 관련된 질 의는 당사자가 포함된 동영상을 비롯해 사건 기록 등 시청각기록물로 유추되었다. 이러한 기초 분석결과를 바탕으로 3장에서는 이용자의 정보요구를 유형별로 분석하고, 입력한 질의어 에 국가기록원에서 검색결과로 제공한 기록물 의 유형을 나누어 파악하고자 한다.

## 4. 이용자 정보요구 세부 결과

### 4.1 정보요구 유형별 분석결과

본 절에서는 국가기록원 이용자들의 정보요 구를 파악하기 위해 “기관명”, “인명”, “사건명” 별로 나누어 세부적으로 분석하였다.

#### 4.1.1 기관명

앞서 살펴보았듯이 국가기록원 질의어 특징 중 하나로 “기관명” 관련 질의어의 검색 빈도 가 높다는 것을 들 수 있다(〈표 3〉 참고). 따라 서 본 절에서는 정보요구 유형별 중 먼저 “기관 명” 질의어를 분석하였으며, 그중 상위 20개는

〈표 4〉와 같다.

분석결과, “기관명” 질의어는 학교나 민간기 관보다는 공공기관이 대부분이었으며, 그중 검 색 빈도가 가장 높은 질의어는 ‘행정자치부 기 획관리실 법무담당관’(13,050)이었다. 해당 질 의어는 〈표 4〉의 다른 질의어에 비해 출현 개월 수는 상대적으로 적지만 검색 빈도는 가장 높 은데, 이는 ‘행정자치부 기획관리실 법무담당관’ 이 단기간에 집중적으로 검색이 되었음을 의미 한다.

‘행정자치부 기획관리실 법무담당관’의 검색 빈도가 높은 이유를 파악하기 위해 해당 기관 의 업무를 연계하여 분석하였다. ‘행정자치부 기획관리실 법무담당관’은 2005년 5월 23일에 폐지되어, 현재 행정안전부 기획조정실 법무담 당관으로 명칭이 변경되어 운영되고 있다. 여 러 업무 중 분석 내용과 관련된 업무 위주로 살 펴보면, 먼저 ‘행정자치부 기획관리실 법무담당 관’은 “법령 및 조약의 공포와 관보의 발간”을 담당하고 있다. 이는 〈표 1〉에서 ‘관보’의 검색 빈도가 두 번째로 높았던 것과 연관 지어 생각 해 볼 수 있는데, 이용자들이 관보를 검색하면 서 생산기관인 ‘행정자치부 기획관리실 법무담 당관’을 함께 검색했을 것이라 유추된다.

〈표 4〉 “기관명” 질의어 상위 20개별 검색결과

순위	질의어	검색 빈도	출현 개월 수	검색결과	
				전체 기록물 검색 건수	[상세검색 > 생산기관] 탭 검색 시 기록물 결과 건수
1	행정자치부 기획관리실 법무담당관	13,050	7	0	255,945
2	감사원	11,001	44	18,589	8,100
3	특허청	10,940	22	3,354	2,584
4	국가보훈처	10,697	22	2,836	1,051
5	국세청	10,120	22	8,488	2,545
6	총무처	10,015	21	13,582	8,382
7	경찰청	9,990	20	13,355	37,995
8	조달청	9,948	21	38,871	2,755
9	관세청	9,579	23	9,063	2,023
10	교육인적자원부	9,543	19	814	3,087
11	식품의약품안전청	9,512	22	1,246	2,916
12	환경부	8,978	21	4,183	2,477
13	보건복지부	8,942	25	3,235	5,716
14	해양경찰청	8,865	21	1,175	467
15	소방방재청	8,486	19	454	830
16	문화재청	8,480	22	1,451	140,707
17	외교통상부	8,234	22	1,360	2,801
18	기상청	8,224	20	2,401	3,408
19	통일부	8,022	21	1,095	712
20	여성가족부	8,021	24	163	1,720

또한 ‘행정자치부 기획관리실 법무담당관’은 “국가기록원의 기록물 관련 법령 제정·개정 등의 업무 지원”, “국가기록원 업무의 운영 지원”과 같은 국가기록원 관련 업무를 맡고 있다. 따라서 이용자들이 해당 업무와 관련된 기록물을 찾기 위해 국가기록원 웹 포털사이트 내에서 검색이 많이 한 것이라 짐작할 수 있다.

한편, 해당 기관의 업무와 연계 분석 후, 실제 이용자가 어떤 기록물을 찾기를 원했는지를 파악하기 위해 국가기록원 웹 포털사이트 통합검색에서 행정자치부 기획관리실 법무담당관을 검색해보았다. 그러나 검색결과는 0건이었으며, [상세검색 > 생산기관] 탭의 특화 검색을 사용해야 제대로 된 검색결과를 얻을 수 있었다(〈표

4〉 검색결과 참고). 생산기관 검색 시 해당 질의어의 검색결과는 255,945건이었으며, 다른 기관명 질의어들도 더 많은 검색결과를 얻을 수 있었다.

상기 분석결과를 종합해보면, 이용자들이 검색하는 기관명이 생산기관을 의미하는 경우가 대다수임을 파악할 수 있다. 따라서 정확한 검색결과 제공을 위해 이용자가 통합검색에서 기관명 검색 시 [상세검색 > 생산기관] 검색결과를 먼저 제시할 필요가 있다.

#### 4.1.2 인명

〈표 5〉는 “인명” 관련 질의어를 분석한 상위 20개 결과로, 가장 검색 빈도가 높은 질의어는



〈표 5〉 “인명” 질의어 상위 20개

순위	질의어	분류	검색 빈도	출현 개월 수	전체 기록물 건수
1	박정희	대통령	25,672	138	66,215
2	조의순	독립운동가	12,278	39	1
3	최순실	일반인	10,606	4	1
4	노무현	대통령	4,368	81	15,844
5	이승만	대통령	4,112	87	12,810
6	조의순 선생	독립운동가	1,751	26	0
7	전두환	대통령	1,673	40	73,993
8	김대중	대통령	1,667	47	64,861
9	김창룡	친일파	1,417	7	9
10	이순신	역사적 인물	1,412	5	523
11	김진수 한국계	일반인	1,280	1	1
12	김재규	정치인	793	8	1,454
13	장준하	독립운동가	697	4	523
14	노태우	대통령	506	1	55,434
15	최태민	일반인	464	12	3
16	육영수	영부인	462	4	10,717
17	김영삼	대통령	417	8	33,452
18	최규하	대통령	364	12	16,427
19	노무현 대통령	대통령	324	8	15,786
20	이명박	대통령	263	8	2,426

‘박정희’(25,672)였다. 이는 그다음 순위인 ‘조의순’과도 13,394번의 큰 차이를 보였으며, ‘육영수’나 ‘김재규’와 같이 관련 인물들도 함께 검색된 것을 확인할 수 있었다. 이러한 현상은 국가기록원 이용자들이 ‘박정희’에 대한 관심이 많음을 시사한다.

또한 ‘박정희’ 뿐만 아니라 ‘노무현’(4,368), ‘이승만’(4,112), ‘전두환’(1,673) 등 대통령 이름이 많이 검색되는 것을 확인할 수 있었는데, 〈표 5〉의 질의어 20개 중 절반에 가까운 9개가 대통령 이름이었다.

한편 그 외 다른 “인명” 질의어를 살펴보면, 대통령뿐 아니라 독립운동가나 일반인에 대한 질의어도 발견할 수 있었는데, 일례로 여러 독립운동가 중 유독 ‘조의순’(12,278)에 대한 검색 빈도가 높았다.

이러한 이용자의 높은 관심과는 달리 제공된 검색결과에는 1건 또는 0건으로 파악되었다. 세부 분석을 위해 전체 질의어 중 ‘조의순’이 포함된 질의어를 〈표 6〉과 같이 도출한 후 국가기록원 웹 포털사이트에서 실제 검색결과를 확인하였지만, 관련 기록물은 존재하지 않았다. 검색된 1건도 독립운동가 ‘조의순’에 관한 기록물이 아니라 ‘관세청 정책홍보관리단 인사기획단 인사발령 보고’ 기록물이었다.

다음과 같은 상황은 이용자들의 관심은 높지만, 실제 기록물은 없는 경우로 이용자들의 검색 만족도를 저하시킬 수 있다. 원인 파악을 위해 ‘조의순’ 관련 기록물 소재를 파악해보니, 현재 ‘조의순’ 관련 기록물은 국가보훈처에서 운영하는 [공훈전자사료관] 웹사이트에서 서비

〈표 6〉 ‘조의순’이 포함된 질의어

순위	질의어	검색 빈도	출현 개월 수	전체 기록물 건수
1	조의순	12,278	39	1
2	조의순 선생의 판결날짜	2,012	23	0
3	조의순 선생	1,751	26	0
4	조의순 판결날짜	468	11	0
5	조의순 선생 판결날짜	216	6	0
6	조의순선생	169	4	0
7	독립운동 판결문`조의순선생	84	1	0
8	조의순 선생의 판결 날짜	79	2	0
9	조의순 선생님의 징역10월의 판결날짜	72	1	0
10	조의순판결문	64	1	0
11	조의순선생의 판결날짜	56	2	0
12	조의순의 판결날짜	35	1	0
13	조의순 선생의 징역 판결날짜	32	1	0
14	조의순 선생님의 판결날짜	30	1	0
15	독립운동 판결문`독립운동 판결문 징역 10월의 판결을 받은 조의순선생 & #40: 당시나이46세 & #41:의 판결날짜	30	1	0
16	조의순선생의판결날짜	24	1	0
17	독립운동가 조의순선생 판결날짜	21	1	0
18	대한독립 만세를 부르며 치안을 방해한 혐의로 재판에서 징역 10월의 판결을 받는 조의순 선생의 판결날짜	19	1	0

스되고 있다(공훈전자자료관, 2019). 따라서 ‘조의순’ 관련 기록물을 찾고자 하는 이용자들은 국가기록원이 아닌 공훈전자자료관에서 검색해야 하지만, 국가기록원 웹 포털사이트에서는 이에 대한 안내가 따로 없는 실정이다. 이처럼 국가기록원 이용자들이 타 기관이 소장하고 있는 기록물에 대한 검색 요구가 지속된다면, 이용자의 편의를 위해 타 기관 기록물을 연결(linking) 하는 방안을 고려해야 한다.

#### 4.1.3 사건

정보요구 유형별 분석 중 마지막 관점으로 “사건” 관련 질의어를 분석하였으며, 그중 상위 20개는 〈표 7〉과 같다. “사건” 질의어에서 가장 검색 빈도가 높은 질의어는 ‘일제강제연행자명

부’(11,859)이지만, 기록물 건수는 0건이었다. 이는 ‘일제강제연행자명부’ 관련 기록물은 현재 기록정보콘텐츠 “일제 강점기 피해자 명부”에서 서비스되고 있으나(국가기록원 기록정보콘텐츠, 2019), 이 같은 내용이 통합검색 결과에서 기록정보콘텐츠로 바로 연결되지 않기 때문에 여전히 이용자는 원하는 기록물을 찾기 어렵다. 이러한 문제점은 선행연구(진주영, 2018)에서도 지적된 사항이기 때문에 국가기록원은 통합검색 시스템 개선을 피할 필요가 있다.

“사건” 관련 질의어 세부 분석을 위해 선행연구와 국가기록원 웹 포털사이트 등을 참고하여 전체 “사건” 질의어를 범주화하여 구분한 결과, 〈표 8〉과 같이 4가지 주요 사건으로 나눌 수 있었다. 이렇게 나누어 본 주요 사건은 1) 일제

〈표 7〉 “사건” 질의어 상위 20개

순위	질의어		검색 빈도	출현 개월 수	전체 기록물 건수
1	일제강제 연행자명부	일제강점기	11,859	82	0
2	625	6.25 전쟁	7,503	43	62,333
3	독립운동 판결문	일제강점기	6,455	33	21
4	올림픽	스포츠	5,158	29	45,062
5	한국전쟁	6.25 전쟁	4,909	98	1,767
6	6.25	6.25 전쟁	4,348	50	4,503
7	새마을운동	새마을운동	3,799	93	2,636
8	대통령선거	기타	2,917	14	6,146
9	조의순 선생의 판결날짜	일제강점기	2,012	23	0
10	5.18	5.18 민주화 운동	1,634	27	1,985
11	518	5.18 민주화 운동	1,587	29	70,436
12	강제징용	일제강점기	1,542	48	10
13	독립운동	일제강점기	1,183	34	2,167
14	일제강점기	일제강점기	1,022	30	228
15	광복절	일제강점기	1,011	17	2,837
16	새마을	새마을운동	1,010	30	34,489
17	625전쟁	6.25 전쟁	968	15	1
18	세월호	기타	873	22	3,081
19	월드컵	스포츠	872	16	26,011
20	독립운동판결문	일제강점기	821	14	2

〈표 8〉 전체 “사건” 질의어 범주화

사건	관련 질의어	기록정보 콘텐츠 구축 여부
일제강점기	일제강제연행자명부, 독립운동 판결문, 조의순 선생의 판결날짜, 강제징용, 독립운동, 일제강점기, 광복절, 독립운동판결문, 일제, 위안부 등	○
6.25 전쟁	625, 한국전쟁, 6.25, 625전쟁, 이산가족, 인천상륙작전, 625사진, 흥남철수, 휴전협정문, 정전협정 등	
5.18 민주화 운동	5.18, 518, 광주민주화운동, 광주사태, 5.18 광주 민주화운동, 광주사태 사망자철, 5공청문회, 광주민주화운동에관한특별담화, 5.18 사망자, 광주항쟁 등	
새마을운동	새마을운동, 새마을, 새마을 운동, 새마을 교재, 새마을총람, 새마을운동 시작에서, 새마을운동 길잡이, 새마을노래	
스포츠	올림픽, 월드컵, 서울올림픽, 아시안게임, 88올림픽, 2002년 월드컵, 올림픽 해산, 북한 성화봉송 등	X
베트남 전쟁	월남전, 월남전쟁, 고엽제	
세월호 및 국정농단	세월호, 구국봉사단, 구국선교단, 구국여성봉사단 등	

강점기, 2) 6.25 전쟁, 3) 5.18 민주화 운동, 4) 새마을운동이었으며, 이외에도 “스포츠”, “베트남전쟁”, “세월호 및 국정농단” 등으로도 분류할 수 있었다.

분석결과, 주요 사건인 “일제강점기”, “6.25 전쟁”, “5.18 민주화 운동”, “새마을운동”은 이용자들이 관심을 가지고 많이 검색하는 만큼 이미 기록정보콘텐츠로 구축되어 제공되고 있

었다. 하지만 4가지 주요 사건 외에도 이용자들은 “스포츠”나 “베트남전쟁”, 사회적 이슈가 된 “세월호 및 국정농단” 등과 관련된 기록물도 많이 검색하지만, 아직 기록정보콘텐츠로 서비스되고 있지 않다. 국가기록원 기록정보콘텐츠가 사회적 관심이나 열람 요청이 많은 기록물을 대상으로 구축됨에도 불구하고, 다양한 사건을 포괄하지 못하는 문제점은 재고해볼 필요가 있다.

따라서 국가기록원은 질의로그 분석결과를 토대로 검색 빈도가 높거나 사회적 관심이 많았던 사건들을 대상으로 기록정보콘텐츠 구축을 추진할 필요가 있다. 또한 최신 이슈를 반영한 기록정보콘텐츠를 구축 및 제공하는 것도 굉장히 중요한데, 이를 통해 국가에서 이슈가 발생했을 때 이용자들이 복잡한 검색 과정을 거치지 않고도 관련 기록물을 빠르고 쉽게 찾아볼 수 있다.

#### 4.2 검색결과 유형별 분석결과

본 절에서는 이용자들이 많이 찾는 질의어에 대해 국가기록원 포털에서 검색결과로 제시한 기록물의 유형을 파악하기 위해 “일반문서류”, “도면류”, “사진·필름류, 녹음·동영상류”로 나누어 분석하였다.

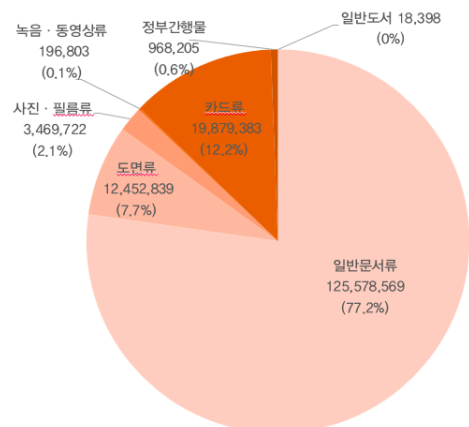
현재 국가기록원 웹 포털사이트에서는 검색결과를 10가지 유형으로 나누어 제공하고 있으며, 그 유형은 다음과 같이 1) 일반문서류, 2) 도면류, 3) 사진·필름류, 4) 녹음·동영상류, 5) 카드류, 6) 국무회의록, 7) 정부간행물, 8) 행정박물, 9) 일반도서, 10) 총독부간행물이다. 여기에서 기록물 건수가 적은 “국무회의”와 “행

정박물”, 그리고 일제 강점기 시대에만 해당하는 “총독부간행물”은 분석대상에서 제외하였다.

분석대상을 선정한 후 <표 1>에서 제시한 바와 같이, 전체 질의어 중 중복을 제거한 4,267개 질의어를 대상으로 검색결과로 제시된 기록물 유형별 기록물 건수를 살펴보았다(<표 9>, <그림 1> 참고). 전체 7가지 유형 중 “일반문서류”(125,578,569/77.2%)가 대다수를 차지하며, 그다음으로 “카드류”(19,879,383/12.2%)가 높은 비중을 차지했다. 이에 비해 시청각기록물인 “사진·필름류와 녹음·동영상류”는 상대적으로 적은 편이었다.

<표 9> 검색결과 유형별 기록물 건수

유형	전체 기록물 건수
일반문서류	125,578,569 (77.2%)
도면류	12,452,839 (7.7%)
사진·필름류	3,469,722 (2.1%)
녹음·동영상류	196,803 (0.1%)
카드류	19,879,383 (12.2%)
정부간행물	968,205 (0.6%)
일반도서	18,398 (0.0%)
합계	162,563,919



<그림 1> 검색결과 유형별 기록물 건수

〈표 10〉은 검색결과 수 상위 20개의 질의어 별로 제시된 기록 유형을 분석한 결과이다. \*로 표시된 질의어들은 검색 빈도 상위 100개 안에 포함된 것을 의미한다.

분석결과, 검색결과가 가장 많은 질의어는 ‘대장’(12,322,973)이었다. 해당 질의어는 검색 빈도 상위 100개 안에는 포함되지는 않지만, 검색결과는 12,322,973건으로 가장 많아 국가기록원에서 대량으로 소장하고 있는 유형의 기록물이 문서류나 카드류와 같이 ‘대장’ 유형임을 알 수 있다. 그 밖에도 ‘토지’, ‘토지대장’ 등과 같이 “토지” 관련 질의어도 대다수였다. 즉 이용자들이 관심을 가지면서 검색 빈도도 높은 “토지” 관

련 질의어가 검색되는 기록물도 많음을 확인할 수 있었다.

〈표 10〉에서 검색결과가 많았던 유형인 “일반문서류”, “도면류”, “사진·필름류 및 녹음·동영상류”를 대상으로 질의어 특성을 세부 분석하였다. 〈표 11〉은 4가지 유형을 대상으로 검색결과가 많은 순으로 상위 20개 질의어 분석결과이다.

#### 4.2.1 일반문서류

“일반문서류” 분석결과, 〈표 11〉에서 나타난 바와 같이 ‘대장’과 토지 관련 질의어들이 검색 결과의 대부분을 차지하고 있었다. 다음으로

〈표 10〉 검색결과 유형별 대상 질의어 상위 20개

순위	질의어	전체	일반 문서류	도면류	사진· 필름류	녹음· 동영상류	카드류	정부 간행물	일반 도서
1	대장	12,322,973	8,078,433	6,687	3,305	183	4,233,956	332	-
2	토지*	9,880,862	5,979,160	9,190	308	30	3,891,460	593	3
3	토지대장*	8,326,475	4,936,379	1	41	-	3,390,054	-	-
4	법률	5,976,931	5,956,897	1	412	28	18,808	743	8
5	도로	5,384,694	5,365,445	7,209	6,538	412	3,661	1,417	2
6	수용	2,723,146	2,599,989	66	674	38	122,042	326	1
7	카드	2,587,276	765,600	87	519	7	1,820,881	174	-
8	경기	2,158,373	177,412	1,951,608	19,648	5,662	53	3,626	8
9	신분장	2,105,245	2,105,067	-	83	-	95	-	-
10	경기도*	2,067,919	112,090	1,951,493	2,720	194	50	1,369	-
11	사고	2,060,010	2,043,031	72	11,891	1,697	2,411	897	1
12	교통사고	2,015,110	2,012,322	-	217	9	2,402	157	1
13	보고	1,939,116	1,850,600	992	31,789	3,106	163	52,147	85
14	교통사고처리	1,915,038	1,912,668	-	-	-	2,368	2	-
15	기록	1,793,710	810,721	681	19,464	7,967	953,357	1,495	11
16	공무원	1,572,817	766,828	78	6,178	348	798,298	1,074	11
17	설치	1,539,410	1,533,636	1,935	2,389	173	754	521	-
18	농지	1,499,076	1,488,950	239	81	9	9,539	255	1
19	상환	1,480,553	1,479,399	-	289	4	770	91	-
20	임야	1,426,439	972,234	753	1	2	453,427	16	-

\*는 검색 빈도 상위 100개 안에 포함된 질의어

〈표 11〉 4가지 유형 대상 질의어 상위 20개

순위	일반문서류		도면류		사진·필름류		녹음·동영상류	
	질의어	검색 건수	질의어	검색 건수	질의어	검색 건수	질의어	검색 건수
1	대장	8,078,433	경기	1,951,608	대통령*	323,864	대회	15,140
2	토지*	5,979,160	경기도*	1,951,493	대회	98,754	기록	7,967
3	법률	5,956,897	북면	850,592	전두환	72,811	경기	5,662
4	도로	5,365,445	도면	309,267	사진	69,172	위원회	4,578
5	토지대장*	4,936,379	수원	249,660	관계	68,987	대통령*	4,518
6	수용	2,599,989	수원군	243,721	문화	65,726	보고	3,106
7	신분장	2,105,067	광주*	226,424	박정희*	64,284	인천	2,363
8	사고	2,043,031	광주군*	225,497	김대중	60,241	자료	2,199
9	교통사고	2,012,322	이천	156,522	박정희 대통령	58,557	한국	2,164
10	교통사고 처리	1,912,668	안성	132,212	노태우	54,770	서울*	1,991
11	보고	1,850,600	경기도 광주	120,916	공보	47,246	문화	1,969
12	농지	1,488,950	강화	118,354	방문	45,835	직할시	1,781
13	상환	1,479,399	양주	113,965	공보부	35,748	대전*	1,758
14	설치	1,533,636	양주군	113,907	한국	35,720	사고	1,697
15	상환 대장*	1,381,952	경기도 안성	111,267	영부인	34,599	올림픽	1,665
16	상환대장*	1,361,933	양평	109,223	공사	33,666	대한	1,632
17	예비군	1,321,469	양평군	103,711	김영삼	32,454	용산	1,587
18	사업	1,194,020	용인	102,608	보고	31,789	방송	1,561
19	신고	1,146,519	연천	102,003	김종필	30,402	방문	1,466
20	등록	1,082,133	연천군	100,095	문화공보부	29,798	표어	1,437

\*는 검색 빈도 상위 100개 안에 포함된 질의어

‘법률’(5,956,897)의 검색 건수가 많았는데, 국가기록원에서 보유하고 있는 “법률” 관련 기록물은 각 지역 고등검찰청에서 생산한 “형사판결 원본”이나 “판결문”, “약식명령원본” 위주이다. 특히 최근 기록물보다는 1980~1990년대에 생산된 기록물이 대다수이다.

이렇듯 이용자들이 “법률” 관련 기록물을 법원이 아닌 국가기록원에서 많이 찾는 이유는 [대한민국 법원 대국민서비스 > 판결서 인터넷열람] 웹사이트에서는 2013년 1월 1일부터 확정된 사건에 대해서만 열람이 가능하기 때문이다(대한민국 법원 대국민서비스, 2019). 즉 국가기록원으로 이관된 2013년 이전 “사건” 관련 기록물을 입수하기 위한 것으로 파악된다.

국가기록원에서는 이와 같은 특징을 반영하여 이용자들이 ‘법률’을 검색하면 “판결문” 위주로 검색결과를 보여줄 필요가 있다. 또한 검색결과 내에서도 기록물을 단순 리스트 형식으로 보여주기보다는 사건별로 관련 기록물을 그룹화하는 방안을 고민해볼 수 있다. 그 외에도 [대한민국 법원 대국민서비스] 링크 정보를 제공하여, 해당 웹사이트의 다양한 정보에 대해 안내를 하는 방법도 검토될 수 있다.

#### 4.2.2 도면류

“도면류” 분석결과, ‘북면’과 ‘도면’을 제외한 나머지 질의어들 모두가 “지역명”과 관련된 것이었다. 특히 “지역명” 관련 질의어 중에서도

‘경기’(1,951,608)와 ‘경기도’(1,951,493)는 다른 질의어보다 검색 빈도도 높을뿐더러 검색결과도 많은 편이다.

〈표 12〉에서 보여지듯이 “지역명” 관련 질의어 검색결과는 주로 공사 및 건물 이전, 신·증축 등과 관련된 설계도면, 경계도, 기본도 즉 지도 위주임을 알 수 있다. 따라서 국가기록원은 이용자의 편의성을 위해 “지역명”이 포함된 질의어 검색결과가 “도면류” 위주로 제시되게끔 우선순위를 변동할 필요가 있다.

#### 4.2.3 사진·필름류 및 녹음·동영상류

〈표 11〉의 “사진·필름류 및 녹음·동영상류” 분석결과, 전반적으로 ‘대통령’, ‘대회’, ‘경기’,

‘회의’와 같은 질의어들이 다수 등장하였다. 특히, “사진·필름류”에서는 “대통령”(323,864) 관련 질의어가, “녹음·동영상류”에서는 “대회”(15,140), “경기”(5,662) 관련 질의어가 대부분인 것으로 파악되었다. 〈표 13〉의 예시와 같이 특정 대통령 관련 질의어에 대한 검색결과를 살펴보면, 대통령 업무 관련 활동 이미지를 제시하기 위한 “사진·필름류”가 가장 많았으며, 대통령과 관련된 영부인이나 외국 원수의 사진도 많았다. 그 밖에도 올림픽이나 아시안게임, EXPO와 같이 국가 주최로 열린 대회나 체육행사과 관련된 질의어에 대한 검색결과에는 해당 행사나 사건의 진행 상황이 포착된 “녹음·동영상류”가 다수였다(〈표 14〉 참고).

〈표 12〉 “도면류” 대상 검색결과 예시

질의어	검색결과(도면류) 예시	생산기관
경기도	경기도 삼각산 국유림경계일람도-0	경기도
	경기도 평택군 수월암 지구 농지 확대 개발 기본 계획도	농업진흥공사
	1969년도 도면-보안림기본도(경기도 용인군)	산림청
광주	광주 감옥 도로 교량 설계도	미상
	광주군 실촌면 수양리 원도 3매지내 1호 20필	조선총독부
	광주(1:5000 기본도)	건설부 국립지리원
광주군	광주군 도척면 노곡리 원도 1매지내 1호 16필	조선총독부
	경기도 광주군 서부면 연귀봉 국유림경계도-001	경기도
	전라남도 광주군 송정면 어등산 국유림경계도-001	전라남도

〈표 13〉 “사진·필름류” 대상 검색결과 예시

질의어	검색결과(사진·필름류) 예시	생산기관
대통령	노태우 대통령 내외분 멕시코 대통령궁 방문 멕시코 대통령과 악수	공보처 홍보국 사진담당관
전두환	이임 대통령(전두환) 국립묘지 참배 02 이임 대통령(전두환) 분향	공보처 홍보국 사진담당관
박정희	제5대 박정희 대통령 취임식 행사 - 박정희 대통령 취임사	공보처 홍보국 사진담당관
김대중	김대중 대통령 ‘김대중선생 생환기념 미사’ 참석3	대통령비서실
박정희대통령	박정희대통령 카터 미국 대통령 접견2	공보처 홍보국 사진담당관
노태우	노태우 대통령 모습	공보처 홍보국 사진담당관
김영삼	손을 흔드는 김영삼 대통령과 인사를 하는 김영삼 대통령 부인 손명순 여사	공보처

〈표 14〉 “녹음·동영상류” 대상 검색결과 예시

질의어	검색결과(녹음·동영상류) 예시	생산기관
대회	2015 광주하계U대회 비전홍보영상 PR_1분(국문)	2015 광주하계 유니버사이드 대회 조직위원회
기록	2012 여수EXPO 근접영상 기록104	2012 여수세계박람회 조직위원회
경기	2014 인천 아시안게임 경기 전 후 상영 이벤트 연출 영상(실내경기장) 경기전 1 실내 경기장 고양	2014 인천아시아경기 대회 조직위원회
올림픽	평창동계올림픽 CF	2018 평창동계올림픽 대회 조직위원회

〈표 15〉 시청각기록물 및 동영상기록물 작성 대상

시청각기록물 작성 대상
<ul style="list-style-type: none"> <li>• 대통령·국무총리 및 중앙행정기관의 장, 지방자치단체장 및 교육감, 교육장 등 주요 직위자의 업무 관련 활동과 인물사진</li> <li>• 외국의 원수·수상, 그 밖에 주요 외국인사의 주요 동정 중 대한민국과 관련되는 사항</li> <li>• 국가 및 지방자치단체의 주요 행사</li> <li>• 국제기구 또는 외국과의 조약·협약·협정·의정서·교류 등의 추진과 관련된 주요 활동</li> <li>• 「국가재정법」 제38조제1항에 해당하는 대규모사업 또는 같은 법 시행령 제14조에 해당하는 대규모 개발사업</li> <li>• 대규모의 토목·건축공사 등의 실시로 본래의 모습을 찾기 어렵게 되는 사항</li> <li>• 철거 또는 개축 등으로 사라지게 되는 건축물이나 각종 형태의 구조물이 사료적 가치가 높아 시청각기록물로 그 모습을 보존할 필요가 있는 사항</li> <li>• 다수 국민의 관심사항이 되는 주요사건 또는 사고로서 공공기관의 장이 시청각기록물의 작성·보존이 필요하다고 인정하는 사항</li> <li>• 증명적 가치가 매우 높아 그 현장 또는 형상을 시청각기록물로 보존할 필요가 있는 사항</li> <li>• 국내 최초의 출현물로서 사료적 가치가 높은 사항</li> <li>• 그 밖에 시청각기록물의 생산이 필요하다고 인정되는 사항</li> </ul>
동영상기록물 작성 대상
<ul style="list-style-type: none"> <li>• 제1항제1호 중 대통령 취임식</li> <li>• 제1항제3호 중 「국가장법」에 따른 장의행사(葬儀行事)와 국가적 차원에서 추진이 필요하다고 인정되어 특별법으로 정한 국제행사 또는 체육행사</li> <li>• 제1항제4호 중 다수의 외국 국가원수 또는 행정수반이 참석하는 국제회의</li> <li>• 제1항제5호 중 공공기관의 장과 중앙기록물관리기관의 장이 협의하여 정한 대규모 사업·공사</li> <li>• 그 밖에 동영상기록물의 생산이 필요하다고 인정되는 사항</li> </ul>

출처: 공공기록물 관리에 관한 법률

이는 〈표 15〉와 같이 『공공기록물 관리에 관한 법률』(이하 공공기록법)에서 정한 시청각 기록물 및 동영상기록물 작성 대상과 그 맥을 같이 하는 것으로, 법상으로는 대통령과 같이 주요 직위자의 업무 관련 활동과 인물사진 및 동영상을 남기게 되어있다. 또한 국가적 차원

에서 추진한 국제행사나 체육행사도 시청각 기록물로 생산하여야 한다.

#### 4.3 검색실패 질의어 특성 분석

국가기록원에서는 지금도 매달 웹 서비스에



대한 이용자 만족도 평가를 수행하고 있다. 이때 제기된 많은 문제점 중 하나로 “검색실패”를 들 수 있는데, 본 연구에서는 그중에서도 검색결과를 하나도 제시하지 못한 경우를 분석하고자 한다(국가기록 설문조사, 2019; 장희정, 2012).

〈표 16〉은 이용자들이 많이 입력하지만 검색결과가 0건인 질의어 상위 10개이며, 자연어 질의나 오타, 국가기록원 이용방법 관련 질의는 제외하고 분석하였다. 분석결과, ‘행정자치부 기획관리실 법무담당관’이나 ‘보건사회부 보건국 위생과’ 같이 “기관명 + 세부 부서”로 입력하면 검색결과가 없었다. “기관명” 분석결과에서 알 수 있듯이 이용자가 검색하는 “기관명” 질의어는 생산기관일 확률이 높으므로, [상세검색 > 생산기관] 탭의 검색결과를 먼저 제시해야 한다(〈표 4〉 참고).

다음으로 ‘일제강제연행자명부’, ‘임야조사부 답동’, ‘중선거구제’와 같이 관련 기록물이 기록정보콘텐츠에서 서비스되고 있는 경우, 통합검색 결과에서 기록정보콘텐츠로 바로 연결되지 않기 때문에 검색결과가 0건이었다. 이러한 문

제점은 통합검색 시스템의 개선이 필요함을 시사한다.

#### 4.4 시사점

본 연구는 국가기록원 이용자 정보요구의 특징을 파악하여 이용자 서비스 방안을 제안하고자 국가기록원 질의로그를 다양한 관점에서 분석하였다. 분석결과를 종합하여 시사점을 도출한 내용은 다음과 같다.

첫째, 통합검색 시스템이 국가기록원 웹 포털 사이트 내의 기록물이나 기록정보콘텐츠와 제대로 연계되지 않았다. 국가기록원 통합검색 시스템은 [상세검색] 결과와 연동되지 않을뿐더러 가장 먼저 보이는 [건 기록물] 검색결과에 기록정보콘텐츠를 포함해서 보여주지 않고 있다. 만약 이용자가 원하는 기록물이 기록정보콘텐츠에서 제공된다면, [웹페이지] 검색결과를 클릭하거나 [기록정보콘텐츠] 사이트로 이동해야 원하는 자료를 찾을 수 있는 분리된 구조이다. 통합검색은 국가기록원 웹 포털사이트 메인화면에

〈표 16〉 “검색실패” 질의어 상위 10개

순위	질의어	검색 빈도	출현 개월 수	전체 기록물 건수
1	행정자치부 기획관리실 법무담당관*	13,050	7	0
2	일제강제연행자명부*	11,859	82	0
3	보건사회부 보건국 위생과	2,439	11	0
4	감찰위원회 비서실	1,998	10	0
5	조의순 선생	1,751	26	0
6	화랑무공훈장철	1,545	4	0
7	감찰위원회 총무과	1,466	8	0
8	임야조사부 답동	1,151	1	0
9	세운 진양	1,139	1	0
10	중선거구제	874	17	0

\*는 검색 빈도 상위 100개 안에 포함된 질의어

서 제공되는 서비스로 국가기록원 웹 포털사이트 구조를 잘 모르는 이용자들이 접속한다면 제일 먼저 이용할 서비스이다(진주영, 2018). 따라서 국가기록원은 이용자들의 편의성을 위해 전체적인 통합검색 시스템을 개선하고 고도화하는 방안을 모색해야 한다.

둘째, 이용자가 많이 검색하는 기록물을 잘 찾아주지 못하거나, 보유하고 있지 않았다. 정보요구 유형별 중 “인명” 분석결과에서 검색 빈도가 높았던 ‘조의순’ 관련 질의어의 검색 결과는 0건이었는데, 이를 실제 확인해보니 ‘조의순’ 관련 기록물은 국가기록원이 아닌 국가보훈처에서 보유하고 있었다. 이러한 상황은 이용자들의 검색 만족도를 현저히 저하시킬 수 있으므로 국가기록원에서 이용자에게 풍부한 검색결과를 제공하기 위해 타 기관의 기록물을 연결(linking)하는 방안을 고려해야 한다.

셋째, 국가기록원 기록정보콘텐츠가 최신 사회적 이슈를 반영하지 못하고 있음을 알 수 있었다. 국가기록원은 이용자들이 관심을 가지고 많이 검색하는 역사적 사건들은 이미 기록정보콘텐츠로 구축하여 제공하고 있지만, 검색 빈도도 높으면서 큰 사회적 이슈가 된 사건들은 포괄하고 있지 못했다. 이는 국가기록원이 신규 기록정보콘텐츠 기획 및 구축 시에 이용자 요구사항을 반영하고 사회적 관심이 많았던 사건들을 대상으로 해야 함을 시사한다.

넷째, 국가기록원 웹 포털사이트는 단순 리스트 형식의 일률적인 검색결과 화면을 제시하고 있었다. 정보요구 유형별과 검색결과 유형별 분석결과, 국가기록원 웹 포털서비스는 질의어의 특성과 상관없이 검색결과를 항상 똑같은 순서로 제공하고 있었다. 이는 이용자들의

흥미를 유발하지 못하며, 이용자가 원하는 기록을 신속하게 찾기 힘들다는 문제점이 있다. 또한 검색결과 화면에 연관 질의어 추천과 같은 서비스가 없어 이용자는 폭넓은 질의어를 선택하는 데 한계가 있다. 따라서 국가기록원은 검색결과 화면에서 질의어 추천과 함께 시각화 기법을 활용하는 방안을 고려할 수 있다.

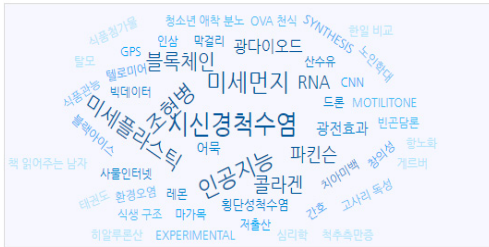
## 5. 개선방안

상기 도출된 시사점을 바탕으로 본 연구에서 제안하는 개선방안은 다음과 같으며, 도식화한 것은 <그림 2>와 같다.

첫째, 국가기록원 웹 포털사이트의 통합검색 시스템 자체에 대한 개선이 시급하다. 앞서 분석한 결과를 통해 국가기록원 통합검색 시스템이 웹 포털사이트 내의 기록물이나 기록정보콘텐츠와 제대로 연계되지 않는다는 문제점이 있으며, 분리된 구조를 가지고 있어 웹 포털사이트 이용이 익숙하지 않은 이용자들에게 어려움을 줄 수 있다는 점을 확인하였다. 따라서 통합검색 시스템이 여러 기록물과 기록정보콘텐츠와 상호 연계되어, 일원화된 구조를 갖출 수 있도록 시스템 고도화가 필요하다. 또한 시스템 개편 측면에서 국가기록원이 보유하고 있지 않은 타 기관 기록물을 연결(linking)하여 제공한다면 이용자는 검색 시간을 줄일 수 있을 뿐만 아니라 풍부한 검색결과를 얻을 수 있다는 장점이 있다.

둘째, 기존에 서비스되고 있는 기록정보콘텐츠의 현행화가 필요하다. 정보요구 유형별 분석을 통해 국가기록원 이용자들은 사회에 이슈





〈그림 4〉 NDSL의 Word Cloud

넷째, 질의어의 특성에 따라 검색결과로 먼저 제공하기 위한 유형별 순서를 변동하여야 한다. 현재 국가기록원 웹 포털 결과화면은 “일반문서류”, “도면류”, “사진·필름류 및 녹음·동영상류”순으로 항상 고정되어 있다. 그러나 검색결과 유형별 분석에 따르면 질의어의 특성에 따라 이용자에게 검색결과로 먼저 제시되는 기록물의 유형이 달라져야 함을 알 수 있었다. “일반문서류”에서는 “도지”와 “법률” 관련 질의어가, “도면류”에서는 “지역명” 관련 질의어의 검색결과가 가장 많았다. “사진·필름류 및 녹음·동영상류”에서는 공공기록법 시청각기록물 및 동영상기록물 작성 대상과 관련하여 “대통령”, “대회”, “경기”와 관련된 질의어의 검색결과가 대다수였다. 따라서 이용자가 입력한 질의어에 따라 검색결과 유형의 우선순위가 변동된다면, 이용자는 검색 과정을 단축할 수 있으며 원하는 기록물을 빠르게 찾을 수 있을 것이다.

## 6. 결론

디지털 시대는 이용자의 기록이용 행태를 바꾸었고, 이용자와 직접적인 커뮤니케이션을 할

수 없는 웹 환경에서 이용자가 남긴 질의로그 빅데이터를 활용하면 이용자의 정보요구를 분석하고 예측할 수 있다(진주영, 2018).

본 연구에서는 국가기록원 빅데이터인 질의로그를 다각적으로 분석하여 이용자 정보요구의 특징을 파악하고, 분석결과를 토대로 이용자 검색 만족도를 향상시킬 수 있는 개선방안을 제안하는 것을 목적으로 수행되었다. 질의로그 분석은 크게 정보요구 유형별과 검색결과 유형별로 나눌 수 있다. 정보요구 유형별 분석은 “기관명”, “인명”, “사건” 별로 수행되었다. 검색결과 유형별에서는 “일반문서류”, “도면류”, “사진·필름류, 녹음·동영상류” 분석이 진행되었으며, 그 분석결과는 다음과 같다.

첫째, 통합검색 시스템이 국가기록원 웹 포털 사이트 내의 기록물이나 기록정보콘텐츠와 제대로 연계되지 않고, 이용자가 이용하기 어려운 분리된 구조였다. 둘째, 이용자가 많이 검색하는 기록물을 잘 찾아주지 못하거나, 국가기록원에서 보유하고 있지 않았다. 셋째, 기록정보콘텐츠는 최신 사회적 이슈를 반영하지 못하고 있었다. 넷째, 국가기록원 웹 포털사이트는 단순 리스트 형식의 일률적인 검색결과 화면만을 제시하고 있었다.

이러한 분석결과를 바탕으로 본 연구에서는 다음과 같은 개선방안을 제안하였다. 첫째, 국가기록원 웹 포털사이트의 통합검색 시스템을 개선하고 고도화해야 한다. 둘째, 이용자의 요구사항과 최신 사회적 이슈를 반영하여 기록정보콘텐츠를 현행화해야 한다. 셋째, 시각화 기법을 활용하여 검색결과를 다양하게 제시하고, 질의어 추천 서비스를 제공해야 한다. 넷째, 이용자가 검색한 질의어의 특성에 따라 먼저 보

이는 기록 유형별 순서를 변동해야 한다.

본 연구는 기존 선행연구들에 비해 12년에 걸쳐 장기간 누적된 질의로그 빅데이터를 기반으로 이용자의 검색행태를 다각도로 분석하였다는 점에서 의의가 있다. 그러나 질의로그 분석과 같은 계량적인 방법만으로는 이용자의 요구를 파악하는데 한계가 있기 때문에, 포커스

그룹 인터뷰(FGI)와 같은 질적 연구방법을 병행하는 것이 필요하다. 또한 심층적인 분석을 위해 오타나 숫자, 한자 등과 같이 이용자들이 질의를 입력하는 행태에 대한 분석도 요구된다. 향후에 이러한 내용이 심도 있게 분석된다면 이용자들에게 더 의미 있는 기록정보서비스를 제공할 수 있을 것이다.

## 참 고 문 헌

- 김성희 (2011). 웹 검색질의어 분석을 통한 사회문화적 특성에 관한 연구. *정보관리연구*, 42(4), 155-174.  
<http://dx.doi.org/10.1633/JIM.2011.42.4.155>
- 남상은 (2017). 도서관 이용자의 검색어 활용에 관한 연구. 석사학위논문, 중앙대학교 대학원, 문헌정보학과 정보학전공.
- 박소연, 이준호 (2005). 국내 웹 이용자의 검색행태 추이 분석. *한국문헌정보학회지*, 39(2), 147-160.  
<http://dx.doi.org/10.4275/KSLIS.2005.39.2.147>
- 박소연, 이준호 (2007). 웹 검색 분야에서의 로그 분석 방법론의 활용도. *한국문헌정보학회지*, 41(1), 231-242. <http://dx.doi.org/10.4275/KSLIS.2007.41.1.231>
- 오효정 (2019). 지능형 전자기록관리 기술연구 개발 기획연구. 대전: 국가기록원.
- 이성숙 (2012). 트랜잭션 로그 분석을 통한 웹기반 온라인목록의 검색행태 추이 분석. *한국비블리아학회지*, 23(2), 209-233. <https://doi.org/10.14699/kbiblia.2012.23.2.209>
- 이수상, 위성광 (2009). 디지털 도서관 이용자의 검색행태 연구. *한국도서관·정보학회지*, 40(4), 139-158.
- 이해영 (2011). 기록관 웹사이트 검색어의 분석과 활용. *한국기록관리학회지*, 11(1), 93-112.  
<http://dx.doi.org/10.14404/JKSARM.2011.11.1.093>
- 이효은 (2015). 웹로그분석을 통한 아카이브 이용자 행태 연구. 석사학위논문, 명지대학교 기록정보과학전문대학원, 기록관리전공.
- 장희정 (2012). 국가기록관 웹사이트의 평가에 관한 연구. *한국기록관리학회지*, 12(2), 51-70.  
<http://dx.doi.org/10.14404/JKSARM.2012.12.2.051>
- 진주영 (2018). 국가기록원 웹사이트의 빅 데이터 분석과 활용. 석사학위논문, 명지대학교 기록정보과학전문대학원, 기록관리전공.
- Christopher J. Prom (2011). Using web analytics to improve online access to archival resources.

The American Archivist, 74, 158-184.

[ 법령 ]

“공공기록물 관리에 관한 법률”(법률 제14613호, 2017.9.22.).

[ 웹사이트 및 인터넷 자료 ]

공훈전자사료관 (2019.11.07.) Retrieved from <https://e-gonghun.mpva.go.kr/user/index.do>

국가과학기술정보센터 (2019.11.10.) Retrieved from <https://www.ndsl.kr/index.do>

국가기록 설문조사 (2019.11.04.) Retrieved from

<http://www.archives.go.kr/next/survey/votePoll.do?searchType=1&keyWord=&searchFlag=&Listsize=10&page=3&id=000230&flag=R>

국가기록원 기록정보콘텐츠 (2019.11.07.) Retrieved from

<http://www.archives.go.kr/next/theme/themeView01.do>

대한민국 법원 대국민서비스 (2019.11.07.) Retrieved from

<https://www.scourt.go.kr/portal/information/finalruling/guide/index.html>

한국학술지인용색인 (2019.11.10.) Retrieved from <https://www.kci.go.kr/kciportal/aboutKci.kci>

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

Jang, Hee-Jung (2012). A study on evaluation of national archives websites. Journal of Korean Society of Archives and Records Management, 12(2), 51-70.

<http://dx.doi.org/10.14404/JKSARM.2012.12.2.051>

Jin, Ju-yeong (2018). Analysis and utilization of big data from the website of national archives of Korea. Unpublished master's thesis, Graduate School of Records, Archives & Information Science at Myongji University, Department of Records and Archival Information Management.

Kim, Seong-hee (2011). A study on the social and cultural characteristics of web queries. Journal of Information Science Theory and Practice, 42(4), 155-174.

<http://dx.doi.org/10.1633/JIM.2011.42.4.155>

Lee, Hyo Eun (2015). A study of user behavior of archive using web analytics. Unpublished Master's thesis, Graduate School of Records, Archives & Information Science at Myongji University, Department of Records and Archival Information Management.

Lee, Soo-Sang, & Wei, Cheng-Guang (2009). A study on the search behavior of digital library

- users. *Journal of Korean Library and Information Science Society*, 40(4), 139-158.
- Lee, Sung-Sook (2012). Trends of web-based OPAC search behavior via transaction log analysis. *Journal of the Korean Biblia Society for Library and Information Science*, 23(2), 209-233. <https://doi.org/10.14699/kbiblia.2012.23.2.209>
- Nam, Sang eun (2017). A study on utilization of search terms in library users. Unpublished Master's thesis, Graduate School of Chung-Ang University, Department of Library & Information Science.
- Oh, Hyo-Jung (2019). A study on development of intelligent electronic records management technology. Daejeon: National Archives of Korea
- Park, Soyeon, & Lee, Joon-Ho (2005). Trends of search behavior of Korean web users. *Journal of the Korean Society for Library and Information Science*, 39(2), 147-160. <http://dx.doi.org/10.4275/KSLIS.2005.39.2.147>
- Park, Soyeon, & Lee, Joon-Ho (2007). Applications of transaction log analysis for the web searching field. *Journal of the Korean Society for Library and Information Science*, 41(1), 231-242. <http://dx.doi.org/10.4275/KSLIS.2007.41.1.231>
- Rieh, Hae-Young (2011). Analysis and utilization of search terms in archival web sites. *Journal of Korean Society of Archives and Records Management*, 11(1), 93-112. <http://dx.doi.org/10.14404/JKSARM.2011.11.1.093>

[ Web Site ]

- Archival Contents of the National Archives of Korea (2019, November 7). Retrieved from <http://www.archives.go.kr/next/theme/contentsOutline.do>
- Gonghun Digital Archive (2019, November 7). Retrieved from <https://e-gonghun.mpva.go.kr/user/index.do>
- Korea Citation Index (2019, November 10). Retrieved from <https://www.kci.go.kr/kciportal/aboutKci.kci>
- National Digital Science Library(2019, November 10). Retrieved from <https://www.ndsl.kr/index.do>
- Supreme Court of Korea (2019, November 7). Retrieved from <https://www.scourt.go.kr/portal/information/finalruling/guide/index.html>
- Survey on National Archives (2019, November 4). Retrieved from <http://www.archives.go.kr/next/survey/votePoll.do?searchType=1&keyWord=&searchFlag=&Listsize=10&page=3&id=000230&flag=R>

