

국가기록원 웹사이트 유입경로와 이용자 검색어 분석*

Analysis of Users' Inflow Route and Search Terms of the Korea National Archives' Web Site

진주영 (Ju Yeong Jin)**

이해영 (Hae-young Rieh)***

초 록

이용자 정보이용환경이 웹으로 변화하면서, 기록관리기관들도 종전보다 더 많은 서비스를 웹을 통해 제공하고 있다. 이 연구는 국가기록원 웹사이트를 이용하는 이용자의 최근 유입경로와 10년 반 동안 매달 상위 100개의 검색어 데이터를 분석하고 그에 맞는 기록정보서비스방안을 제시한다. 분석 결과 유입경로는 크게 포털 사이트별 접속, 국가별 접속, 유사기관 별 접속 및 모바일 접속으로 분류할 수 있었다. 이용자들의 검색어를 분석한 결과 이용자들이 10년 반 동안 가장 많이 검색한 검색어는 '토지조사부'였으며, 꾸준한 관심을 갖고 검색한 검색어 또한 '토지조사부'였다. 그 외에 정부 기관에서 생산한 문서 또는 관보에 대한 관심도 높음을 확인하였다. 그리고 이용자들이 가장 많이 검색한 검색어와 꾸준히 등장한 검색어를 파악한 결과, 검색어들을 크게 토지, 일제강점기, 6·25 전쟁과 남북관계, 기록관리 및 이용으로 묶어볼 수 있었다. 분석결과를 토대로 포털 사이트 및 모바일 등을 활용한 국가기록원 웹사이트의 연계성 강화와 검색 서비스의 고도화 방안을 제안하였다. 이 연구는 이렇게 웹로그나 이용자 검색어 등을 분석하면 기록관리기관 이용자들에게 제공하는 서비스를 고도화할 수 있는 의미 있는 결과를 얻을 수 있음을 확인하였다.

ABSTRACT

As the users' information use environment changes to the Web, the archives are providing more services on the Web than before. This study analyzes the users' recent inflow route and the highly ranked 100 search terms of each month for 10 and half years in the Web site of National Archives of Korea, and suggests suitable information services. As a result of the analysis, it was found out that the inflow route could be divided into access from portal site, by country, from related institutions, and via mobile platform. As a result of analyzing the search terms of users for the last 10 and half years, the most frequently searched term turned out to be 'Land Survey Register', which was also the search term that was searched for with steady interests for 10 and half years. Also, other government documents or official gazettes were of great interests to users. As results of identifying the most frequently searched and steadily searched terms, we were able to categorize the search terms largely in terms of land, Japanese colonial period, the Korean war and relationship of North Korea and South Korea, and records management and use. Based on the results of the analysis, we suggested strengthening connection of the National Archives Web site with portal sites and mobile, and upgrading and improving search services of the National Archives. This study confirmed that the analysis of Web log and user search terms would yield meaningful results that could enhance the user services in archives.

키워드: 기록정보서비스, 웹로그 분석, 이용자 검색어 분석, 유입경로 분석, 빅데이터 분석
archival information services, Web log analysis, users' search terms analysis,
inflow route analysis, big data analysis

* 이 논문은 진주영의 석사학위 논문 '국가기록원 웹사이트의 빅 데이터 분석과 활용' 내용 일부를 정리한 것임.

** 명지대학교 기록정보과학전문대학원 석사(sky8010jy@naver.com) (제1저자)

*** 명지대학교 기록정보과학전문대학원 교수(hyrieh@mju.ac.kr) (교신저자)

■ 논문접수일자: 2018년 2월 19일 ■ 최초심사일자: 2018년 3월 8일 ■ 게재확정일자: 2018년 3월 15일

■ 정보관리학회지, 35(1), 183-203, 2018. [http://dx.doi.org/10.3743/KOSIM.2018.35.1.183]

1. 서론

1.1 연구 배경 및 목적

기록을 보존하고 관리하는 이유는 궁극적으로 기록에 담긴 정보를 이용자들이 잘 활용하도록 하기 위해서이다. 과거에 이용자들은 직접 기록관리기관을 방문하여야만 기록 속의 정보를 찾을 수 있었다. 그러나 디지털 시대에 도래하면서 웹 환경이 보편화 되었고, 이용자들은 직접 기록관리기관에 방문하지 않아도 시공간에 제약 없이 웹을 통해 정보를 이용하고 활용할 수 있다고 생각한다. 그러나 웹에서 찾을 수 있는 기록은 제한적이며, 또 기록관리기관 입장에서는 웹 환경에서 이용자와의 직접적인 커뮤니케이션이 어려워 이용자들이 원하는 기록을 찾을 수 있도록 도와주기가 어렵다. 그러므로 기록관리기관은 웹으로 찾아오는 이용자들의 특징을 잘 파악하여, 그들이 많이 원하는 기록에 대한 서비스를 강화하고, 웹을 통해서 많이 요구되고 활용되는 기록에 대한 더 적극적인 서비스가 가능하도록 해야 할 것이다.

이렇게 웹으로 방문하는 이용자들의 정보 행위를 파악하는 방법으로, 이용자들이 웹사이트를 방문할 때 남겨진 유입경로나 검색어, 클릭한 게시물 등 이용자 흔적이 담긴 로그 데이터의 분석이 있다. 박소연과 이준호(2007)는 로그 데이터 분석을 통하여 대다수 이용자의 전반적인 이용형태를 분석할 수 있다고 하였다. 또한 로그 데이터 분석을 통해서도 이용자들이 자주 찾는 콘텐츠, 주로 입력하는 키워드 등을 토대로 콘텐츠나 키워드 등을 추천해 줄 수 있으며, 이용자들의 접속 경로나 내비게이션 행

태를 파악하여 웹사이트의 구조를 조정하여 이용자에게 유용한 웹서비스를 구축할 수 있다(이수상, 위성광, 2009).

Duff(2001)는 웹이 사회를 혁명적으로 변화시키고 있다고 하며 인터넷을 통한 기록정보서비스의 중요성을 강조하였다. 정동훈(2010)은 정부기관의 정보는 그 사이트에서만 제공 가능한 특수한 정보이기 때문에 가능한 많은 정보를 제공해야 함과 동시에 이용자의 편의성을 고려해야 한다고 하였다.

국가기록원 웹사이트 또한 기록원에서 소장하고 있는 기록을 다양한 계층의 이용자에게 제공하는 공간이기 때문에, 이용자 중심의 편리성과 웹으로서의 유용성을 갖추고 있어야 한다. 2000년도에 만들어진 국가기록원 웹사이트를 통해서도 이용자 검색어 등 방대한 양의 웹 로그가 쌓여 있음을 확인하였다. 최근 정보기술의 발전으로 빅데이터를 분석하여, 그간 확인하지 못했던 의미 있는 결과를 도출하는 사례가 많아졌다. 국가기록원 이용자들에게 더 의미 있는 기록정보서비스를 제공하기 위해 웹 로그 빅데이터를 분석하여 이용자의 정보 이용 행위를 파악하는 것은 큰 의미가 있을 것이다.

장희정(2012)은 웹사이트 평가 기준을 도입하여 국가기록원 웹사이트 이용자에게 조사한 결과 콘텐츠 사용 대상자에 대한 구분이 전혀 없으며 대상이용자에 적합한 정보를 제공하지 못하고 있고, '기록물 검색', '기술계층별 검색'을 따로 나누어서 검색하도록 되어 있어 혼동을 준다고 하였다. 이와 같이 현재 국가기록원 웹사이트에서는 기록검색도구와 콘텐츠 등을 제공하고 있으나 이용자가 정보를 쉽게 찾을 수 있게 도와주는 검색 가이드나 이용자의 정

보 요구를 반영한 콘텐츠 등 이용자들이 원하는 정보를 편리하고 쉽게 찾도록 도와줄 기록 정보서비스는 부족한 실정이다.

이 연구에서는 국가기록원 웹사이트 이용자의 요구를 파악할 수 있는 이용자의 흔적이 담긴 웹로그 빅데이터를 분석하여 이용자들의 국가기록원 웹사이트 유입경로와 검색어를 분석하여 이용자들의 정보이용 행태를 확인해보고, 그 결과를 반영한 기록정보서비스의 고도화 방안과 이용자 맞춤형 웹서비스 제공 방안을 구상하여 제공하고자 한다.

1.2 선행연구

이 연구와 관련하여서는 웹로그 및 빅데이터 분석에 관한 연구들을 살펴보았다. 먼저 웹로그 분석을 진행한 연구들로 부터는 연구방법 및 결과를 확인하였으며, 웹로그 분석결과가 이용자의 이용 행태를 반영하여 서비스 개선 방향을 확인할 수 있었다. 그간 문헌정보학 및 기록학 분야에서 로그분석에 대한 연구는 웹 검색 중심으로 많이 진행되었다.

해외에서는 Silverstein et al.(1999), Jansen, Spink, Saracevic(2000), Spink et al.(2001, 2002), Koshman, Spink, Jansen(2006)과 Jansen, Spink, Koshman(2007) 등 많은 연구에서 검색 엔진 로그 데이터를 기반으로 웹 사용자들의 검색행태를 살펴보았다. 특히 Jansen(2006, 2009)은 기존의 검색 로그에 의한 연구들을 종합적으로 분석하고, 로그분석을 위한 노이즈 제거 방법 등 효율적인 분석 방법을 제시하였으며, 검색로그의 활용분야와 연구현황 등을 제시하였는데, 방법론이 시사하는 바가 크다.

또 Dumais et al.(2014)은 그간의 로그분석 연구들을 포괄적으로 살펴보았는데, 로그를 특성별로 언어나, 지역 별, 접근 장치 별, 시간이나 이용자 특성 별로 분할하여 살펴본 것을 특징으로 보았다. 예를 들면 Kotov et al.(2011) 등은 이용자들이 같은 검색을 이후에 되풀이 한다는 것을 밝혀내기도 했다. 또 Dumais et al.(2014)은 다양한 서비스의 개선을 위한 웹 로그 기반의 실험연구도 소개하고, 또한 웹로그 데이터를 수집하는 다양한 방법도 포괄적으로 살펴보아, 로그 분석 연구 전반에 대한 이해를 높여주었다.

국내에서는 특히 네이버의 로그 데이터를 기반으로 한 연구가 많이 진행되었고, 그 외에 웹포털과 개인 PC에서의 이용자 웹검색 행태 등 다양한 연구들이 진행되었으며, 기록학 분야에서도 웹로그 분석을 활용한 연구들이 진행되었다. 이들을 살펴보면, 먼저 박소연과 이준호(2002)는 로그분석을 위한 세션 정의 방법 및 로그 정제 방법과 질의 유형 분류 방법 및 검색어 정의 방법 등을 제안하였으며, 이후 박소연과 이준호(2005, 2007)는 웹로그 분석을 통하여 대다수 이용자의 전반적인 검색 행태를 분석할 수 있음을 확인하고, 효율적인 검색시스템 개발과 콘텐츠 구축에의 기여가 가능함을 제시하였다. 또한 박소연(2011)은 네이버 이용자들이 조회한 문서를 기록한 클릭 로그를 분석하였는데, 질의별로 클릭의 절반 정도가 한 문서에 집중된 것으로 나타났음을 확인하였다.

한편 이소영과 정영미(2006)는 다음 포털 사이트에서 12일간 수집한 이용자 로그 데이터를 사용하여 이용자의 관심 주제를 반영한 검색 결과 재순위화 및 관련 주제 카테고리화 질의어

추천으로 구성된 개인화 검색 서비스 모형을 제안하고, 결과의 만족도도 높음을 확인하였다. 또 이수상과 위성광(2009)은 총 3개월 간 NDSL에 로그인하여 검색한 로그 데이터에 대해 네트워크 분석방법을 통해 검색행위의 다양한 특성을 살펴보았다. 그 결과 특정한 검색자 네트워크에서 중심적인 위치를 차지하는 검색자들이 존재하였으며, 전체 검색자들은 다수의 하위 집단으로 군집되어 있었고, 중심 검색자들은 다른 검색자들과 검색 키워드를 공유하고 있음을 확인하였다.

이중식(2011)은 사용자 PC에 웹로그 캐처를 설치하고, 취합된 웹검색 기록을 이용자들이 몽치화하도록 하도록 하고, 설문조사를 통해 검색의 동기, 만족도 등을 조사하였는데, 이용자들의 행동 패턴을 웹로그를 통해 확인할 수 있음을 제시하였다. 또, 이성숙(2012)은 웹 기반 온라인목록 이용자들의 검색행태를 7년간의 트랜잭션로그 분석을 통해 확인하였다. 검색유형, 접근점, 연산자, 단어 사용횟수 등의 검색전략과 검색실패에 대해 분석하고 이들이 온라인 목록 시스템과 서비스개선에 결과 활용될 수 있음을 제시하였다.

또 Han, Joo, Wolfram(2014)은 이미지 기반 디지털 도서관에서 한 달간의 이용자 검색 로그데이터를 분석하여, 이용자들이 검색 결과 페이지와 아이템 보기를 가장 많이 행했으며, 좀 더 많은 메타데이터가 필요함을 확인하였다. 한편 박소연, 조기훈, 최기린(2015, 2017)은 각각 쇼핑 검색 사이트인 팜슈즈와 1300K 이용자들의 로그분석을 통해 이용자들의 정보검색 형태를 조사 분석하였다. 특히 1300K 이용자들의 PC와 모바일 정보검색 행태를 비교 분석한

결과, 모바일의 세션수가 PC보다 좀 더 많으며 검색어 수는 2배 이상 많고, PC에서의 브라우징 빈도는 모바일보다 약간 더 많은 것으로 나타났다 하였고.

한편 기록학 분야에서도 웹로그 분석 결과를 서비스 개선에 적용하고자 하는 연구들이 몇 개 진행되었다. Prom(2011)은 웹로그 분석이 아키비스트가 이용자들의 이용 흔적을 측정하고 이용자 행동을 이해하여 온라인 서비스를 개선하기 위해 사용할 수 있는 방법이라고 하였다. 이해영(2011)은 대통령기록관 웹사이트의 검색어분석 결과를 콘텐츠서비스 등 기록정보서비스 개발의 한 방법론으로 검토해보았으며, 이효은(2015)은 구글애널리틱스를 실제 아카이브 웹사이트에 적용하여 웹 이용자를 분석하고, 구체적인 이용자 세분화와 서비스방안을 제시하였다.

빅데이터를 활용한 연구는 기록학 분야에서는 아직 진행된 바 없었고, 문헌정보학 분야에서 다양하게 진행되었는데, 그 중에서도 도서관의 이용자서비스와 관련된 연구들을 살펴보았다. 이 연구들은 빅데이터 분석에 방법 면에서도 참고가 되었을 뿐 아니라, 결과의 서비스 활용 방안 확인에도 도움이 되었다.

먼저 대학도서관 웹사이트의 웹로그를 분석한 김서(2013)의 연구는 KAIST 도서관의 방문자, 검색어, 페이지, 발송을 통해 유입경로, 자료별 관심, 사용자환경, 시스템에 대한 웹로그를 빅데이터 분석기법으로 분석하고, 이에 따른 이용자 요구에 맞는 메뉴 개발, 시스템 점검 시간, 콘텐츠 개발 방안 등을 제시하였다. 이정미(2013)는 빅데이터가 다양한 도서관 정보서비스에 활용될 수 있을 것으로 판단하고, 도서관

정보센터 역할 규정 및 전략적 정책 수립에 도입하는 방안, 지역사회 지원을 위해 활용하는 방안, 도서관·정보센터의 정보자원 개발 계획에 활용하는 방안, 서비스 대상 이용자에 따른 맞춤형 정보서비스 계획에 활용하는 방안 등을 제안하였다.

표순희 외(2015)는 도서관 빅데이터의 유형별로 활용 가능한 빅데이터를 분석하기 위해, 도서관계 연구자 및 현장 사서와의 심층인터뷰, 표적집단 인터뷰(focus group interview), 그리고 사서 및 이용자 설문조사를 시행하였다. 그리고 빅데이터를 이용한 사서 의사결정 지원 서비스와 이용자 도서 추천 서비스 및 디지털 큐레이터 서비스 등을 제안하였다.

이 연구는 특히 대학도서관 웹사이트 웹로그를 분석한 김서(2013)의 연구 방법을 참고하여 국가기록원 웹사이트의 웹로그를 빅데이터로 보고 분석하였으며, 앞서 언급한 웹로그 및 빅데이터 분석과 관련된 선행 연구들을 통해 이용자 서비스 방안 등을 도출할 수 있었다. 그러나, 우리나라의 중추적인 기록관리기관인 국가기록원의 웹로그는 분석된 바 없었으며, 또 빅데이터 분석은 아직 기록학 분야에서는 연구되지 않아, 이 연구를 통해 그를 보완하고자 하였다.

1.3 연구범위 및 방법

이 연구에서는 국가기록원 웹사이트 이용자의 요구를 파악하기 위해 이용자의 흔적이 담긴 데이터 즉, 웹로그로 남은 유입경로와 검색어를 분석하여 국가기록원 웹사이트 서비스 적용 방안과 이용자 맞춤형 웹서비스 제공 방안을 제시

하고자 하였다. 이용자의 요구를 분석하기 위한 웹로그 데이터는 국가기록원에 정보공개청구를 하여, 2016년 1월부터 2017년 9월까지의 이용자 유입경로 URL 데이터 128,175개를, 검색어는 2007년 4월부터 2017년 9월까지의 매달 1위부터 100위까지의 로그 데이터로 총 12,600개를 엑셀 파일로 받았다. 유입경로 데이터는 기관의 기술적 한계로 1년 반 기간의 데이터만 확보가 가능하였고, 검색어 데이터는 10년 반 기간의 데이터를 분석할 수 있었다.

유입경로는 엑셀을 이용해 구글, 네이버, 다음, 그 외 포털 등으로 분류하였고 이중에서 모바일 접속을 했는지도 확인하였다. 또 도메인 주소에서 세계 여러 나라에서 국가기록원 웹사이트에 접속하는 것을 확인하고, 이들을 국가별로 분류하여 분석하였다. 유입경로에 IP 주소가 나오거나, URL이 명확하지 않은 데이터가 약 2,600개 정도 있었는데, 이들은 기타로 분류하였다. 분석결과는 표와 그래프 등을 사용하여 나타내었다.

다음으로 이용자들이 국가기록원 웹사이트에서 무엇을 찾는지 파악하기 위해서 이용자들이 입력하는 검색어를 확인하였다. 이를 위해 2007년 4월부터 2017년 9월까지 이용자들이 입력한 매달 1위부터 100위까지 총 12,600개의 검색어 데이터를 받았으며, 엑셀 피벗테이블을 이용하여 정리하였다. 검색어에는 11,~7.&&* 등과 같은 노이즈가 30개 정도 있었는데, 이러한 노이즈를 제외하고 12,570개를 분석하였다. 이 검색어들은 이용자들이 정보를 찾기 위하여 검색한 단어가 몇 번 검색 되었는지를 뜻하는 검색 횟수와 매달 1위부터 100위 중 2004년 4월부터 2017년 9월까지 총 126개월 동안 단어가

몇 번 등장하였는지 빈도를 체크한 등장 횟수로 구분하여 분석하였으며, 이들을 통해 이용자들의 요구를 파악하고자 하였다.

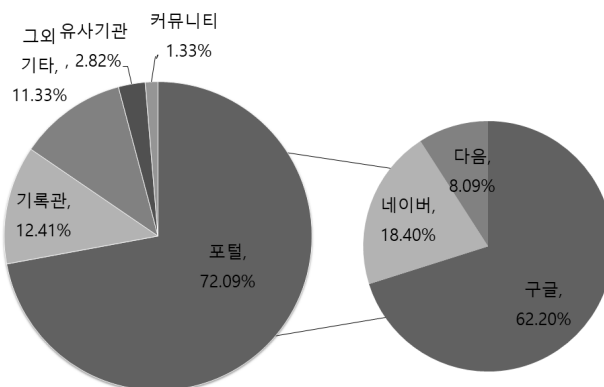
빅데이터를 분석하기 위해서는 R과 같은 통계프로그램을 많이 이용하지만, 이 연구에서는 간단히 엑셀 피벗테이블을 활용하였다. R 프로그램이 개인의 PC 환경에서 활용하기에는 제한이 있었고, 데이터 수가 대단히 많은 것은 아니었으며, 또한 전문가 두 명의 자문을 통해 확인해 본 결과 엑셀 피벗테이블을 활용하여 분석하여도 문제가 없다고 판단되었다. 또한 빅데이터 분석에서 데이터의 특징을 도출하기 위해 활용하는 워드클라우드를 활용하였다. 워드클라우드란 문서의 키워드 개념 등을 직관적으로 파악할 수 있도록 핵심 단어를 시각적으로 돋보이게 하는 기법이다. 예를 들면 많이 언급된 단어를 크게 표현해 한눈에 들어올 수 있게 하는 것이 가능하다(시사상식사전, 2015). 이 연구에서 쓰인 워드클라우드란 'Worditout'으로, 웹사이트에서 서비스를 제공하므로 별도의 설치를 할 필요가 없으며, 사용이 편리하여 쉽게 이용할 수 있다.

2. 국가기록원 웹사이트 분석 결과

2.1 유입경로 분석

유입경로란 이용자가 웹사이트에 어떤 경로를 통해 접속하였는지 나타내는 것으로, 유입 URL에서 호스트 이름과 도메인 이름, 모바일 접속 여부 등을 알 수 있어 국가기록원 웹사이트를 이용하는 이용자가 국가기록원 웹사이트에 어떻게, 무엇을 통해 들어오게 되었는지 경로를 알 수 있다.

이용자의 유입경로를 파악하기 위해 데이터를 분석한 결과, 구글, 네이버 다음 등 포털 사이트를 통한 접속이 전체의 72.1%(92,408번)를 차지하였으며, 공공기관 기록관에서의 접속이 12.4%(15,910번)고, 그 외 IP주소 접속, 노이즈 등을 포함한 기타 접속이 11.3%(14,533번)로 나타났다. 그 외 유사기관(도서관, 대학사료실 등)이 2.8%(3,618번), 커뮤니티(오늘의 유머, 일간베스트, MLBPARK, Facebook, Twitter 등)가 1.3%(1,745번)를 차지하였다. <그림 1>은 이 내용을 원형그래프로 나타낸 것이다.



<그림 1> 국가기록원 상세 유입경로

또 세계 여러 나라에서도 국가기록원 웹사이트에 접속한 것을 확인하였다. 한편 스마트 기기의 보급으로 모바일 기기로 접속한 횟수는 14,619번으로 전체이용의 11%를 차지했다. 이 유입경로를 크게 포털과 국가별, 유사기관으로 나누어 살펴보고, 또 모바일 접속에 대해서도 별도로 분석해 보았다.

2.1.1 포털 접속

국가기록원 유입경로 중 포털 접속을 자세히 살펴보면, 구글이 전체의 62.2%(57,530번)를 차지하였으며, 네이버가 18.4%(17,054번), 다음은 8.1%(7,482번), 그 외 포털(Zum, Bing, Wikipedia, 언론사 등)은 11.2%(10,342번)를 차지하였다(〈표 1〉 참조). 즉 이용자들은 포털 검색을 통해 많이 들어왔으며, 지식인, 블로그, 또는 학술정보를 통해서도 접속하는 등, 다양한 유입경로를 통해 국가기록원 웹사이트에 들어오는 것으로 나타났다. 구글을 통한 접속 57,530번 중 구글 메인 페이지에서 검색해서 들어오는 것이 57,504회로 압도적으로 많았다. 네이버는 전체 17,052번 중 10,600번이 메인페이지 검색을 통해 국가기록원 웹사이트에 접속된 것을 비롯해, 블로그, 지식인, 학술정보 등 다양한 유입경로를 통해 접속되었다. 다음 또한 네이버와 비슷한 유입경로를 나타냈다.

〈표 1〉 포털별 유입경로 횟수

포털이름	유입경로	횟수
구글	메인	57,504
	기타	26
네이버	검색	10,600
	지식인	2,944
	기타	3,510
다음	다음검색	5,592
	기타	1,890

2.1.2 국가별 접속

국가기록원 웹사이트 유입경로의 도메인을 국가별로 분석한 결과 미국, 일본, 영국 등을 비롯해 142개국에서 국가기록원 웹사이트에 접속한 것으로 나타났다. 접속횟수는 총 30,854번이었으며, 그 중 구글을 통해 접속하는 횟수가 20,245회로 가장 많았고, 야후와 한국을 알리는 블로그인 'Askakorean', 그리고 위키피디아 등을 통해 접속하였다. 〈표 2〉에 제시된 바와 같이 국가별로는 미국이 9,972번으로 제일 많았으며, 그 다음 일본, 영국, 캐나다, 독일, 호주, 베트남, 홍콩, 프랑스, 중국 순으로 나타났다. 일본과 중국을 제외한 나라들은 대부분 구글을 통해 접속하였으며, 일본은 구글, 위키피디아재팬, 야후재팬, 네이버재팬 등을 통해 접속하는 횟수가 높았으며, 중국은 위키피디아 또는 중국 자체 도메인으로 접속하는 횟수가 높았다.

〈표 2〉 국가별 접속 횟수

순위	나라이름	접속횟수
1	미국	9,972
2	일본	4,544
3	영국	1,597
4	캐나다	1,580
5	독일	967
6	호주	911
7	베트남	731
8	홍콩	684
9	프랑스	671
10	중국	586

2.1.3 유사기관을 통한 접속

국가기록원 웹사이트에 접속하기 전에 국가기록원과 유사한 기관에서 접속한 횟수는 3,618번으로 전체 유입경로 중 2.8%를 차지했다. 적은

숫자이긴 하지만 유입경로에서 유사기관 유입은 기록관이 어느 기관과 협력하여 이용자에게 정보를 제공해야 하는지 알 수 있는 중요한 경로이다. 먼저 각 지방자치단체 홈페이지를 비롯해 각 지방자치단체에서 하는 행정서비스 웹사이트나 정부산하기관 웹사이트를 통해 국가기록원 웹사이트에 접속한 것은 총 2,223번이었다(〈표 3〉 참조). 국립중앙도서관 및 그 외 도서관 홈페이지를 통해 국가기록원 웹사이트에 접속한 횟수는 445번이었고, 대학도서관 및 사료실을 통해 국가기록원 웹사이트에 접속한 횟수는 696번, 외국대학교 도서관에서 국가기록원 웹사이트에 접속한 횟수도 132번이었다.

〈표 3〉 유사기관별 유입경로 횟수

유입경로	횟수
지방자치단체 및 정부기관 (시청, 행정안전부, 환경부, 법제처 등)	2,223
국립중앙도서관 및 그 외 도서관	445
대학도서관 및 사료실	696
외국대학교 도서관	132

2.1.4 모바일 접속

스마트 기기의 보급으로 이용자들이 웹사이트를 이용할 때 모바일을 통해 접속하는 횟수가 점점 증가하고 있다. 모바일 접속은 유입경로에서 별도로 확인을 해보았다. 스마트 기기를 이용해 국가기록원을 이용한 횟수는 14,619회로 전체의 약 11%를 차지하였다. 전자정부서비스 이용 실태조사 결과 보고서(행정자치부, 한국정보화진흥원, 2016)에 의하면, 전자정부서비스 접속은 스마트폰이나 태블릿 PC 등의 모바일 접속이 2015년의 7.1%에 비해 2016년에는 20.5%로 늘어난 것을 확인할 수 있었다. 그러나 국가기

록원의 모바일 접속은 11%에 불과한데, 이는 국가기록원의 모바일 접속 환경의 미비에 기인한 것으로 보인다. 이는 국가기록원이 모바일을 이용하여 접속하는 이용자들을 위한 서비스 방안을 더 적극적으로 마련해야 함을 의미한다.

2.2 검색어 분석

검색어 로그 데이터는, 검색어가 몇 번 검색되었는지를 뜻하는 검색 횟수와 총 126개월 동안 100위 안에 검색어가 몇 번 등장하였는지를 확인한 등장 횟수로 구분하여 분석하였다. 또한 검색어들을 주제별로 분석하여 큰 주제 분야 별로 묶어보았다.

2.2.1 검색횟수와 등장횟수 상위 검색어

제일 먼저 이용자들이 국가기록원 웹사이트에서 가장 많이 찾고 꾸준히 찾는 검색어가 무엇일까 의문점을 갖고 매달 검색된 상위 100개 용어의 합인 12,567개의 검색어가 입력된 1,469,269번 중에 가장 많이 검색한 검색어 순위와 등장횟수 순위를 비교해 보았다. 예를 들면, '토지조사부'의 검색 횟수는 총 133,488번이었으며, 등장횟수는 126개월 동안 126번으로, 2004년 4월부터 2017년 9월까지의 기간 동안 매달 이용자들이 이 단어를 입력하여 검색했다는 것을 알 수 있다. 그 반면 최순실은 검색횟수는 10,595번이고 등장횟수는 4번으로 2016년 10월, 11월, 12월 2017년 1월에만 검색되어 4번만 등장한 것을 알 수 있다. 〈표 4〉는 검색횟수와 등장횟수 상위 30개 검색어를 나타낸 것이다.

〈그림 2〉는 'Worditout'을 이용하여 검색횟수와 등장횟수 상위 30권을 대표하는 검색어 클라

〈표 4〉 검색횟수와 등장횟수 상위 30개 검색어

순위	검색횟수 기반 상위 검색어	순위	등장횟수 기반 상위 검색어
1	토지조사부	1	토지조사부
2	관보	2	관보
3	경기도	3	박정희
4	관리	4	보존기간
5	해양	5	임야조사부
6	위토대장	6	토지대장
7	설계	7	지적원도
8	박정희	8	토지
9	행정자치부 기획관리실 법무담당관	9	한국전쟁
10	조의순	10	새마을운동
11	일제강제연행자명부	11	일제강제연행자명부
12	토지	12	대통령
13	보존기간	13	노무현
14	감사원	14	이승만
15	특허청	15	상환대장
16	국가보훈처	16	고시
17	최순실	17	위토대장
18	광주군	18	조선총독부
19	국세청	19	판결문
20	총무처	20	대한뉴스
21	경찰청	21	분배농지부
22	조달청	22	지적도
23	관세청	23	국무회의
24	교육인적자원부	24	기록물분류기준표
25	식품의약품안전청	25	지세명기장
26	환경부	26	무주부동산공고
27	보건복지부	27	무주부동산
28	해양경찰청	28	기록물관리지침
29	소방방재청	29	경기도
30	문화재청	30	발간등록번호

* 검색횟수: 이용자들이 2004년 7월부터 2017년 9월 까지 매달 하나의 검색어를 입력한 횟수

* 등장횟수: 하나의 검색어가 2004년 7월부터 2017년 9월 까지 매달 100개 검색어 안에 포함된 횟수



〈그림 2〉 'Worditout'을 이용한 검색횟수(왼쪽)와 등장횟수(오른쪽) 상위 30위의 검색어

우드를 나타낸 것이다. <그림 2>에서 검색횟수를 보여주는 왼쪽 클라우드는 가장 많이 검색된 '토지조사부'를 중심으로, 관보, 경기도, 관리, 위토대장, 박정희, 해양, 설계 등의 단어들에 눈에 띈다. 검색횟수의 등장횟수를 보여주는 오른쪽 클라우드는 가장 많이 검색된 '토지조사부'를 중심으로, 관보, 박정희, 보존기간, 임야조사부, 지적원도, 토지대장 등의 단어들에 눈에 띈다.

검색횟수와 등장횟수에서는 공통적으로 1, 2위가 토지조사부와 관보로 같으며, 1위부터 30위까지 공통적으로 들어가 있는 검색어는 이 두 용어를 포함해 경기도, 위토대장, 박정희, 일제강제연행자명부, 토지, 보존기간의 8개였다. 가장 많이 등장한 용어 1위인 '토지조사부'는 1910년에서 1918년까지 토지조사사업으로 작성된 명부이다. 2위인 '관보'는 국가가 국민에게 널리 알릴 사항을 편찬하여 발행하는 국가의 공고기관지로, '관보'라는 제호로 발간하는 정기간행물이다(대한민국 전자관보, 2018). 당시 국가의 정치·사회상이 반영된 역사적인 기록물로서 가치가 있으며, 헌법개정·법률·조약·대통령령·총리령 및 부령의 공포와 헌법개정안·예산 및 예산외국고부담계약의 공고 수단이다. 또한, 법령 공포와 대통령(국무총리)훈령 및 대통령(국무총리) 지시사항, 각급 기관의 인사발령 통지 등 정부 기관간의 공문 시행을 대체하는 역할을 한다. 이 두 검색어는 국가기록원이 생소한 잠재적 이용자에게 기록원이 소장한 기록에 대한 정보를 알릴 수 있는 중요한 용어이다. 일제강제연행자명부는 국가기록원 웹사이트에서 검색하면 바로 검색되는 기록물들은 없으나, 국가기록원 웹사이트에서 제공하는 일제강점기 피해자 명부 기록 콘텐츠로 제공되는

데, 국가기록원에서만 이용할 수 있다.

검색횟수와 등장횟수는 3위부터는 차이가 난다. 검색횟수와 등장횟수에서 3위부터 30위까지 공통적으로 들어가 있는 검색어들은 경기도, 위토대장, 박정희, 일제강제연행자명부, 토지, 보존기간의 6개이지만, 이들의 순위는 차이가 많이 났다. 검색횟수에서 3위를 차지한 경기도가 등장횟수에서는 29위였다. <표 4>를 보면 검색횟수에서 4위와 5위를 차지한 관리와 해양이라는 단어는 등장횟수 상위어에서는 찾아볼 수 없었으며, 설계, 행정자치부 기획관리실 법무담당관, 조의순, 감사원 등의 단어는 등장횟수 상위어에서 찾아볼 수 없었다. 반대로 등장횟수 상위어에는 있지만 검색횟수 상위어에는 없는 검색어도 있었다. 임야조사부, 토지대장, 지적원도, 한국전쟁, 새마을운동, 대통령 등이 이에 해당된다. 또한 검색횟수 상위어에는 행정자치부 기획관리실 법무담당관, 감사원, 특허청, 국세청, 총무처 등 정부기관과 관련된 검색어들이 많이 나타난 반면에, 등장횟수 상위어에는 토지 관련 검색어들인 임야조사부, 토지대장, 지적원도, 토지, 지세명기장 등이 포함된다.

2.2.2 검색어의 주제 영역별 분석

이용자들이 검색한 용어는 몇 가지 큰 주제로 묶어 볼 수 있었다. 예를 들어 이용자들은 토지에 관련된 정보를 찾기 위해 토지조사부, 토지 문서, 토지대장, 임야조사부 등을 검색하였으며, 일제강점기시대 정보를 찾기 위해 일제강제연행자명부, 독립운동 판결문, 조선총독부, 징용 등의 용어를 검색하였다. 이렇게 많이 검색된 용어들은 크게 토지, 일제강점기시대, 6·25 전쟁 및 북한 관련, 기록관리 등으로 묶을 수 있었

다. 주제 영역별로 묶인 검색어들은 연관검색어, 추천검색어, 관련용어사전 등을 통해 이용자들에게 기록정보서비스를 고도화 할 수 있는 방안으로 제시될 수 있을 것이다. 대통령 관련 검색어들도 하나의 주제로 묶을 수 있었지만, 이는 대통령 기록관이 따로 있기 때문에 이 연구에서는 분석에서 제외하였다.

1) 토지 관련 검색어

토지, 즉 땅과 관련된 단어들은 전체 검색 횟수 1,469,264번 중 243,913번이 검색되었다. 이 중 가장 많이 검색된 단어는 '토지조사부'였다. 그 다음은 위토대장이었으며 토지, 토지대장, 임야조사부, 지적원도, 무주 부동산 공고 등이 있었다. <표 5>는 토지 관련 검색어 중에서 검색 횟수가 2,000번 이상인 단어 10개와 등장횟수가 40번 이상인 단어들을 보여준다.

<표 5>에 제시된 단어들을 포함해서 토지 관련 된 검색어들은 총 172개였는데, 토지 정보를 찾고자 하는 검색어가 대부분을 차지하였다. 또한 이용자들은 검색을 할 때 '무주부동산공고'와 같이 지역명과 문서이름의 복합어로 토

지 관련 정보를 많이 검색한 것으로 나타났다.

2) 일제강점기시대 관련 검색어

일제강점기시대 관련된 단어들로 많이 검색된 용어는 '일제강제연행자명부'였다. 그 다음은 독립운동 판결문이었으며, 조선총독부, 징용, 광복, 강제징용, 독립운동 등이 포함되었다. 일제강점기시대에 대한 검색어는 총 96개로 확인할 수 있었으며, 이 단어들은 총 35,013번 검색되었고 573번 등장하였다. 일제강점기시대 관련 단어 중 500번 이상 검색된 단어들과 10번 이상 등장한 검색어들은 <표 6>과 같다.

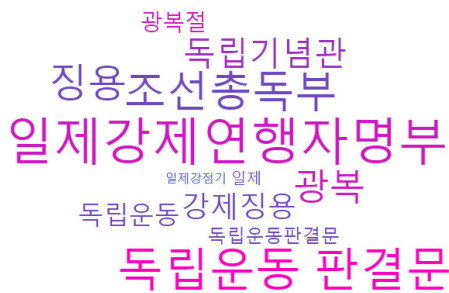
<표 6>에서 나온 검색어를 비롯해 일제강점기시대 전체 검색어들을 크게 분류하면 강제징용, 일제강제징용징병, 징용, 징용자, 왜정시피 징용자명부와 같은 징용명부 관련 검색어들과 독립운동판결문, 독립운동, 대한독립만세 등과 같은 독립운동 관련 검색어들로 분류할 수 있었다. 그 외 위안부, 총독부, 군함도 등의 검색어들이 있었다. <그림 3>은 일제강점기시대 검색어들의 등장횟수를 워드클라우드로 나타낸 것이다.

<표 5> 토지 관련 검색어

검색어	검색횟수	검색어	등장횟수
토지조사부	133,338	토지조사부	126
위토대장	26,566	임야조사부	101
토지	11,347	토지대장	100
토지대장	8,335	지적원도	94
임야조사부	7,325	토지	87
지적원도	6,608	위토대장	67
무주부동산공고	3,168	분배농지부	62
분배농지부	3,035	지적도	61
농지소표	2,541	지세명기장	55
분배농지상환대장	2,434	무주부동산공고	49

〈표 6〉 일제강점기 시대 관련 검색어

검색어	검색횟수	검색어	등장횟수
일제강제연행자명부	11,577	일제강제연행자명부	81
독립운동 판결문	6,448	조선총독부	67
조선총독부	2,812	강제징용	43
징용	1,311	징용	42
광복	1,240	독립운동 판결문	33
독립기념관	1,207	일제	28
강제징용	1,204	독립운동	28
독립운동	911	일제강점기	20
광복절	838	광복	19
독립운동판결문	821	광복절	14



〈그림 3〉 일제강점기 시대 관련 검색어

일제강점기시대 검색어들은 역사 교육이나 연구와 관련된 조선총독부관보, 서대문형무소,

3.1운동 등의 검색어들도 있었으며, 또한 2017년 7, 8월에는 '군함도'라는 용어가 영화와 TV프로그램에서 언급되면서 그 시기에 많이 검색된 것을 볼 수 있었다.

3) 6·25 전쟁 및 남북 관계 관련 검색어
6·25 전쟁 및 이산가족이나 남북정상회담 등 남북 관계와 관련된 검색어들은 총 45개로 확인되었으며 18,849번 검색되었다. 6·25 전쟁 관련하여서는, 625, 6.25, 625전쟁, 한국전쟁 등 여러 용어로 이용자들이 검색을 하였다. 사용된

〈표 7〉 6·25 전쟁 및 남북관계 관련 검색어

검색어	검색횟수	검색어	등장횟수
625	7,361	한국전쟁	82
한국전쟁	3,766	625	40
6.25	3,024	6.25	34
625전쟁	957	625전쟁	15
전쟁	595	전쟁	11
이산가족	326	북한	11
북한	288	이산가족	9
인천상륙작전	265	인천상륙작전	6
625사건	262	남북정상회담	6
판문점	252	625 전쟁	5

3. 분석결과에 기반한 기록정보 서비스 방안

여기서는 앞에서 분석한 국가기록원 웹사이트의 유입경로와 검색어 분석 결과에 기반하여 국가기록원에서 기록정보서비스에 활용할 수 있는 방안을 제안한다. 웹사이트 연계성 관련, 콘텐츠 서비스 및 검색 서비스 관련하여 제안이 가능하다고 판단하였다.

3.1 국가기록원 웹사이트의 연계성 및 접근성 강화

3.1.1 포털 사이트에 적극적 노출

포털을 이용해 국가기록원에 웹사이트를 이용한 이용자들은 대부분 구글, 네이버, 다음 등 포털 사이트를 통해 국가기록원 웹사이트에 접속하였다. 그런데, 구글에서는 국가기록원이 소장한 기록물에 대해 검색하면 국가기록원 웹문서가 검색되어 국가기록원으로 연결이 된다. 그러나 같은 검색어를 네이버에서 검색 했을 때는 검색결과에서 국가기록원 소장 기록물을

찾아볼 수가 없다(<그림 5> 참조).

네이버는 국내 1위 포털 사이트로 국내에서 가장 많이 이용되는 사이트이다. 그럼에도 불구하고 검색을 했을 때 국가기록원이 소장한 기록물에 대한 정보가 없다는 것은 네이버를 이용하는 잠재적인 이용자를 놓칠 수 있다는 것이다. 이러한 문제는 다음도 마찬가지였다. 국가기록원은 네이버와 다음을 이용하는 이용자들을 끌어들이기 위해 국가기록원이 소장한 소장물에 대한 정보를 네이버와 다음에 공유하도록 해야 할 것이다. 따라서 검색했을 때 국가기록원 홈페이지가 첫 화면에 노출 될 수 있도록 하고, 이용자들이 국가기록원이 소장한 기록물과 콘텐츠에 대한 정보도 볼 수 있게 해주어야 할 것이다. 이렇게 국가기록원이 소장한 기록물 정보를 노출시킴으로써 이용자에게 국가기록원의 역할에 대한 홍보는 물론 잠재적인 이용자를 더 끌어 들일 수 있을 것이다.

3.1.2 모바일 서비스 활성화

이용자들이 웹사이트를 이용할 때 모바일을 통해 접속하는 횟수가 점점 증가하고 있다. 스



<그림 5> 구글과 네이버에 같은 검색어를 넣고 검색한 결과

마트 기기를 이용해 국가기록원에 접속한 횟수는 전체의 약 11%를 차지하였는데, 앞서도 언급한 바와 같이 이는 2016년 전자정부에 대한 스마트기기 접속 비율 20.5%(행정자치부, 한국정보화진흥원, 2016)에 비해 매우 낮은 수치이다. 현재 국가기록원은 모바일 웹 서비스만 제공하고 어플리케이션 서비스는 제공하고 있지 않으며, 모바일 웹 서비스도 PC 접속 서비스에 비해 기능이 많이 부족하다. 그런데 모바일 웹은 페이지 단위 로딩에 따라서 속도가 저하되고, 브라우저 기반으로 운영되기 때문에 어플리케이션보다 제공할 수 있는 서비스가 제한되어 있다.

반면 어플리케이션은 사용자 기기에 직접 디바이스를 설치 후에 사용하는 방식으로 웹보다 쉽고 빠르게 이용이 가능하고, 일부는 인터넷을 연결하지 않아도 사용이 가능하다. 또한 특정 타겟의 이용자에게 맞는 다양한 콘텐츠를 제공할 수 있는데, 국가기록원도 다양한 이용자와 다양한 콘텐츠를 갖고 있으므로, 이러한 콘텐츠를 이용자에게 맞게 편리하게 제공하기 위해 어플리케이션을 개발하여 서비스를 제공하는 것이 바람직할 것이다.

3.1.3 해외 이용자 대상 영어 콘텐츠 제공

142개 국가에서 국가기록원 웹사이트에 접속한 흔적이 유입경로 데이터에서 발견됨으로써, 국가기록원 웹사이트가 세계 여러 나라에서 이용되고 있다는 것을 알 수 있었다. 따라서 국가기록원은 해외 이용자들을 위한 서비스에도 관심을 가지고 정보를 제공해야 할 것이다. 세계 각지에서 국가기록원에 접속하는 것은 한류 등 다양한 콘텐츠에 관심이 있어서 일 것이

다. 국가기록원은 현재 영어, 일본어, 중국어로 국가기록원에 대한 소개와 RMS, 국가기록원 검색 서비스 등에 관한 설명을 제공한다. 그러나 국가기록원 웹사이트에서 정보를 검색하는 방법이나 국가기록원 웹사이트에 구축되어 있는 콘텐츠 정보에 대해서는 외국어 서비스를 지원하지 않고 있다. 따라서 해외 이용자들이 한글을 모른다면 국가기록원 웹사이트를 이용하기에 어려울 것이다. 국가기록원 웹사이트가 외국어 서비스를 폭넓게 지원한다면 세계 각지에서 국가기록원 웹사이트를 이용하는 이용자가 더 증가할 것이다.

3.2 검색 서비스 고도화 방안

3.2.1 검색어 추천 서비스

이용자들이 제일 많이 찾은 검색어는 토지조사부, 관보, 정부기관명칭 등이었다. 이용자들이 검색한 검색어들을 분석한 결과, 크게 토지, 일제강점기, 6·25전쟁 및 북한 관련, 기록관리 등의 주제 분야가 많이 검색된 것을 확인할 수 있었다. 이렇게 많이 찾는 검색어들을 활용하여 실시간 검색어, 연관 검색어, 다른 이용자들이 찾는 검색어 등 검색어 추천 서비스를 제시할 수 있다. 예를 들어 이용자가 조상 땅을 검색했다면, 관련 검색어 또는 추천검색어로 토지조사부, 임야조사부, 지적원도를 추천해 줄 수 있을 것이다.

이는 이용자들의 실시간 관심사를 다른 이용자들에게 알려줄 수 있고, 이것을 본 이용자들은 찾고자하는 기록물 외에 다른 콘텐츠나 기록물을 찾아봄으로써 국가기록원에서 소장하고 있는 기록물에 대해 알 수 있는 방법으로, 국가

기록원의 역할을 다양한 이용자들에게 더 많이 알려줄 수 있다. 이처럼 실시간 검색어 혹은 연관 검색어는 다른 이용자들의 관심사를 알려주고 최근 이슈가 무엇인지 알 수 있도록 해줄 수 있다. 이러한 서비스를 제공함으로써 이용자들이 정보를 신속하게 이용할 수 있고 원하는 정보를 명확하게 찾을 수 있을 것이다.

또 주제별 검색어 서비스를 더 확장해 나가면 관련 용어 서비스도 분석된 검색어를 활용하여 좀 더 확장할 수 있을 것이고, 이용자들을 위한 더 개선된 서비스가 가능하다. 예를 들어 토지를 검색했을 때 관련용어로 이용자들이 검색한 단어들을 추가할 수 있다. 현재 관련용어 사전에서는 귀속농지, 농지소표, 지주별 농지확인, 조상 땅 등 이용자들이 검색한 단어가 없었다. 이러한 단어들을 '다른 이용자들이 찾는 단어'라는 항목을 추가하여 제공하면, 찾고자 하는 기록물의 정확한 명칭이나 키워드를 모르는 이용자들에게 큰 도움이 될 것이다.

3.2.2 검색 가이드 제공

검색 가이드는 기록물관리기관 내외의 다양한 정보원을 주제별로 모아줌으로써 특정 기록에 대한 '의도적 검색' 없이도 기록을 발견할 수 있도록 하며, 주제와 관련된 특정 기록유형의 역사적 배경, 기록에 포함된 정보의 의미 등 생산 배경과 기록과의 관계를 설명함으로써 기록의 의미를 이해할 수 있도록 한다(설문원, 2015). 이용자들이 기록을 발견하고 더 나아가 기록을 이용하는 방법을 알려주기 위해서 검색가이드를 이용하면, 또 다른 관련정보를 찾을 수 있도록 할 수 있을 것이다.

검색 가이드를 만들기 위해서는 제일 먼저 주

제별로 상세하게 분류하고 큰 주제 밑에 이용자들이 검색한 단어들을 연관 검색어들로 보여줄 수 있다. 예를 들면 6·25 전쟁에서 좀 더 세분화된 미·소의 한반도 정책과 남·북한 정세, 또는 연도 등으로 단계를 나누어 이와 관련된 기록을 보도록 할 수 있을 것이다. 그리고 덧붙여 기록을 이용하는 방법, 그리고 전쟁에 관한 개요와 국가기록원에서 소장한 기록물, 그리고 이 주제와 관련된 연관 검색어, 관련 정보원, 서적 등을 추천해 줄 수 있을 것이다.

4. 결론

디지털 시대는 이용자의 정보 이용환경을 바꾸었고, 빅데이터를 수집, 분석하여 많은 이용자들의 행동이나 의견을 분석하고 예측할 수 있게 해줄 수 있다. 이용자와 직접적인 커뮤니케이션을 할 수 없는 웹 환경에서 이용자들이 남긴 로그를 활용하면 이용자들의 정보이용 행태를 분석할 수 있으며, 몇 년간 쌓여온 빅데이터 로그분석은 이용자를 위한 서비스를 개발하는데 중요한 자료가 된다. 이 연구는 국가기록원 웹사이트의 이용자들이 남긴 로그의 빅데이터를 분석하면 어떤 결과를 얻을 수 있으며, 이 이용자들의 특성을 반영해 어떤 기록정보서비스를 할 수 있을까 하는 의문에서 시작되었다. 이용자 분석을 위해 2007년 4월부터 2017년 9월까지 입력한 검색어를 매달 1위부터 100위까지 총 12,600개, 그리고 2016년 1월부터 2017년 9월까지 이용자 유입경로 128,175개의 웹 로그 데이터를 엑셀 피벗테이블과 워드클라우드 Worditout을 활용하여 분석하였다.

그 결과 유입경로는 크게 포털 사이트별, 국가별, 유사기관별 접속 및 모바일 접속 등으로 분류할 수 있었다. 또한 검색어를 분석한 결과 이용자들이 10년 반 동안 가장 많이 검색한 단어는 '토지조사부'였으며, 꾸준한 관심을 갖고 검색한 단어 또한 '토지조사부'였다. 이를 통해 우리나라 국가기록원 웹사이트를 이용하는 이용자들은 토지에 관심이 많으며 조상 땅과 관련된 문서를 찾기 위해 국가기록원 웹사이트를 많이 이용하는 것을 다시 한 번 확인할 수 있었다. 많이 검색된 용어들을 주제별로 묶었을 때는 토지, 일제강점기, 6·25 전쟁 및 남북관계 관련, 기록관리 및 이용으로 묶어 분석할 수 있었다. 또한 정부기관에서 생산한 문서 또는 관보에 대한 관심이 높음을 확인하였다.

이를 통해 잠재적 이용자들을 끌어들이기 위해 포털 사이트에 적극적인 노출이 필요하며, 또 모바일 앱의 필요성과 외국인 연구자 또는 학생들을 위한 외국어 콘텐츠가 필요함을 확인하였다. 그리고 이용자들이 검색한 용어를 활용하여 실시간 검색어 및 연관 검색어 추천, 다른 이용자가 검색한 검색어 추천 및 검색가이드 제공 등의 서비스 고도화 방안이 필요함을 확인하였다.

이 연구에서 제시된 웹로그 빅데이터 분석

방법과 그 결과를 반영하여 제시된 서비스 고도화 방안들은 국가기록원뿐만 아니라 향후 다른 기록관리기관에서도 활용이 가능할 것이다. 이번 연구와 같은 데이터 분석이 다른 기록관리기관에도 확산되어, 다양한 분석과 활용 및 기록정보서비스 방안들이 이용자들을 위해 개발되길 바란다.

이 연구에서는 빅데이터를 분석하기 위해서 통계프로그램이 아닌 엑셀 피벗테이블을 활용했는데, 쉽게 접할 수 있는 프로그램으로 분석을 해보았다는 장점이 있으며, 향후 다른 연구에서는 빅데이터를 분석하는 다양한 통계 프로그램을 활용할 수도 있을 것이다. 이 연구에서는 이러한 프로그램들을 활용하여 기록연구자들이 웹로그나 이용자 검색어 등을 분석하면 기록관리기관 이용자들에게 제공하는 서비스를 고도화할 수 있는 의미 있는 결과를 얻을 수 있을 것임을 확인하였다.

앞으로 이러한 데이터를 기록관리의 기반으로 활용하고, 기록정보서비스를 다양하게 개발하는 것이 필요할 것이다. 앞으로 빅데이터는 마케팅을 비롯해 인공지능 등 다양한 분야에서 활용될 것이다. 기록학 분야에서도 빅데이터의 수집, 분석을 통해, 분류체계 고도화 방안 등 다양한 활용방안이 연구되고 발표되기 바란다.

참 고 문 헌

- 김서 (2013). 빅데이터 도서관 활용사례: KAIST 도서관 웹로그분석시스템을 중심으로, KSLA bulletin, 5, 45-52.
- 대한민국 전자관보 (2018). 행정안전부. 홈페이지. Retrieved from <http://gwanbo.mois.go.kr/user/intro/definePage.do>

- 박소연 (2011). 검색 포털의 클릭 집중 문서 분석 평가. 한국도서관·정보학회지, 42(1), 325-338.
- 박소연, 이준호 (2002). 로그 분석을 통한 이용자의 웹 문서 검색 행태에 관한 연구. 정보관리학회지, 19(3), 111-122. <https://doi.org/10.3743/KOSIM.2002.19.3.111>
- 박소연, 이준호 (2005). 국내 웹 이용자의 검색 행태 추이 분석. 한국문헌정보학회지, 39(2), 147-160. <https://doi.org/10.4275/KSLIS.2005.39.2.147>
- 박소연, 이준호 (2007). 웹 검색 분야에서의 로그 분석 방법론의 활용도. 한국문헌정보학회지, 41(1), 231-242. <https://doi.org/10.4275/KSLIS.2007.41.1.231>
- 박소연, 조기훈, 최기린 (2015). 이용자들의 쇼핑 검색 행태 분석: 팝슈즈 로그 분석을 중심으로. 정보관리학회지, 32(4), 289-305. <https://doi.org/10.3743/KOSIM.2015.32.4.289>
- 박소연, 조기훈, 최기린 (2017). 이용자들의 PC 검색 행태와 모바일 검색 행태 비교. 정보관리학회지, 34(3), 89-107. <https://doi.org/10.3743/KOSIM.2017.34.3.089>
- 설문원 (2015). 기록의 발견과 이해를 위한 온라인 검색가이드 연구. 한국기록관리학회지, 15(1), 53-87. <https://doi.org/10.14404/JKSARM.2015.15.1.053>
- 시사상식사전 (2015). 서울: 박문각. Retrieved from <http://terms.naver.com/list.nhn?cid=43667&categoryId=43667>
- 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>
- 이소영, 정영미 (2006). 웹 포털 이용자 로그 데이터에 기반한 개인화 검색 서비스 모형의 설계 및 평가. 정보관리학회지, 23(4), 179-196. <https://doi.org/10.3743/KOSIM.2006.23.4.179>
- 이수상, 위성광 (2009). 디지털 도서관 이용자의 검색행태 연구. 한국도서관·정보학회지, 40(4), 139-158.
- 이정미 (2013). 빅데이터의 이해와 도서관 정보서비스에의 활용. 한국비블리아학회지, 24(4), 53-73. <https://doi.org/10.14699/kbiblia.2013.24.4.053>
- 이중식 (2011). 웹검색 행태 연구. 한국문헌정보학회지, 45(2), 209-228. <https://doi.org/10.4275/KSLIS.2011.45.2.209>
- 이해영 (2011). 기록관 웹사이트 검색어의 분석과 활용: 대통령기록관을 중심으로. 한국기록관리학회지, 11(1), 93-112.
- 이효은, 임진희 (2015). 웹애널리틱스를 이용한 아카이브 이용자 분석 사례 연구. 기록학연구, 45, 83-120.
- 장희정 (2012). 국가기록관 웹사이트의 평가에 관한 연구. 한국기록관리학회지, 12(2), 51-70.
- 정동훈 (2010). 웹 2.0 특성에 따른 정부부처 웹페이지 평가. 사이버커뮤니케이션학보, 27(4), 209-255.
- 지적아카이브 (2017). 행정안전부 국가기록원. 홈페이지. 2017.12. Retrieved from <http://theme.archives.go.kr/next/acreage/typeArchive.do?type1=1&type2=1&type3=2>
- 표순희, 김윤형, 김혜선, 김완중 (2015). 도서관 빅데이터 서비스 모형 개발에 관한 연구. 정보관리학회지, 32(2), 63-86. <https://doi.org/10.3743/KOSIM.2015.32.2.063>

- 행정자치부, 한국정보화진흥원 (2016). 2016 전자정부서비스 이용실태조사 결과 요약보고서.
- Duff, W. (2001). Web을 통한 기록물 공개 열람 제공: 축복인가 저주인가? (이상민 역). 기록학연구, 4, 175-196.
- Dumais, S., Jeffries, R., Russell, D. M., Tang, D., & Teevan, J. (2014). Understanding user behavior through log data and analysis. In *Ways of Knowing in HCI* (pp. 349-372). New York: Springer. https://doi.org/10.1007/978-1-4939-0378-8_14
- Han, H., Joo, S., & Wolfram, D. (2014). Using transaction logs to better understand user search session patterns in an image-based digital library. *Journal of The Korean Biblia Society for Library and Information Science*, 25(1), 19-37. <https://doi.org/10.14699/kbiblia.2014.25.1.019>
- Jansen, B. J. (2006). Search log analysis: What it is, what's been done, how to do it. *Library and Information Science Research*, 28(3), 407-432. <https://doi.org/10.1016/j.lisr.2006.06.005>
- Jansen, B. J. (2009). *Understanding user: Web interactions via Web analytics*. (Synthesis Lectures on Information Concepts, Retrieval, and Services) Plymouth: Morgan and Claypool Publishers. <https://doi.org/10.2200/S00191ED1V01Y200904ICR006>
- Jansen, B. J., Spink, A., & Koshman, S. (2007). Web searcher interaction with the Dogpile. com metasearch engine. *Journal of the Association for Information Science and Technology*, 58(5), 744-755. <https://doi.org/10.1002/asi.20555>
- Jansen, B. J., Spink, A., & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Information Processing and Management*, 36(2), 207-227. [https://doi.org/10.1016/S0306-4573\(99\)00056-4](https://doi.org/10.1016/S0306-4573(99)00056-4)
- Koshman, S., Spink, A., & Jansen, B. J. (2006). Web searching on the Vivisimo search engine. *Journal of the Association for Information Science and Technology*, 57(14), 1875-1887. <https://doi.org/10.1002/asi.20408>
- Kotov, A., Bennett, P., White, R. W., Dumais, S. T., & Teevan, J. (2011). Modeling and analysis of cross-session search tasks. In *Proceedings of SIGIR 2011*, 5-14. <https://doi.org/10.1145/2009916.2009922>
- Prom, C. (2011). Using Web analytics to improve online access to archival resources. *The American Archivist*, 74(1), 158-184. <https://doi.org/10.17723/aarc.74.1.h56018515230417v>
- Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1999). Analysis of a very large Web search engine query log. *SIGIR Forum*, 33(1), 6-12. <https://doi.org/10.1145/331403.331405>
- Spink, A., Ozmutlu, S., Ozmutlu, H. C., & Jansen, B. J. (2002). U.S. versus European Web searching trends. *ACM SIGIR Forum*, 36(2), 32-38. <https://doi.org/10.1145/792550.792555>

Spink, A., Wolfram, D., Jansen, M. B., & Saracevic, T. (2001). Searching the Web: The public and their queries. *Journal of the Association for Information Science and Technology*, 52(3), 226-234. [https://doi.org/10.1002/1097-4571\(2000\)9999:9999::AID-ASII591>3.0.CO;2-R](https://doi.org/10.1002/1097-4571(2000)9999:9999::AID-ASII591>3.0.CO;2-R)

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Cadastral Archive content (2017). National Archives of Korea Portal. 2017, December. Retrieved from <http://theme.archives.go.kr/next/acreage/typeArchive.do?type1=1&type2=1&type3=2>
- Chung, Donghun (2010). Evaluation of the government Web pages according to Web 2.0. *Journal of Cybercommunication Academic Society*, 27(4), 209-255.
- Common Sense Dictionary (2015). Seoul: Parkmungak. Retrieved from <http://terms.naver.com/list.nhn?cid=43667&categoryId=43667>
- Duff, W. (2001). Disclosure of records through Web: Blessing or curse? (Translated by Sang Min Lee). *The Korean Journal of Archival Studies*, 4, 175-196.
- 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.
- Kim, Seo (2013). Big data library use cases: KAIST library Web log analytics system. *KSLA bulletin*, 5, 45-52.
- Lee, Hyoeun, & Yim, Jin Hee (2015). A case study analysing the users of archives through Web analytics. *The Korean Journal of Archival Studies*, 45, 83-120.
- Lee, Jeong-Mee (2013). Understanding big data and utilizing its analysis into library and information services. *Journal of the Korean Biblia Society for Library and Information Science*, 24(4), 53-73. <https://doi.org/10.14699/kbiblia.2013.24.4.053>
- Lee, Joong-Seek (2011). Web search behavior analysis based on the self-bundling query method. *Journal of the Korean Biblia Society for Library and Information Science*, 45(2), 209-228. <https://doi.org/10.4275/KSLIS.2011.45.2.209>
- Lee, Soo-Sang, & Wei, Cheng-Guang (2009). A study on the search behavior of digital library users: Focus on the network analysis of search log data. *Journal of Korean Library and Information Science Society*, 40(4), 139-158.
- Lee, Soyung, & Chung, Young-Mee (2006). Design and evaluation of a personalized search service model based on Web portal user activities. *Journal of the Korean Society for Information Management*, 23(4), 179-196. <https://doi.org/10.3743/KOSIM.2006.23.4.179>

- 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>
- Ministry of the Interior, National Information Society Agency (2016). 2016 Summary report on the survey results of the use status of use of the e-Government service.
- Park, Soyeon (2011). Analysis and evaluation of most clicked documents of Korean search portal. *Journal of Korean Library and Information Science Society*, 42(1), 325-338.
- Park, Soyeon, & Lee, Joon Ho (2002). Investigating Web search behavior via query log analysis. *Journal of the Korean Society for Information Management*, 19(3), 111-122. <https://doi.org/10.3743/KOSIM.2002.19.3.111>
- 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. <https://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. <https://doi.org/10.4275/KSLIS.2007.41.1.231>
- Park, Soyeon, Cho, Kihun, & Choi, Kirin (2015). Information seeking behavior of shopping site users: A log analysis of Popshoes, a Korean shopping search engine. *Journal of the Korean Society for Information Management*, 32(4), 289-305. <https://doi.org/10.3743/KOSIM.2015.32.4.289>
- Park, Soyeon, Cho, Kihun, & Choi, Kirin (2017). A comparison of user search behavior on PC and mobile phone: A log analysis of 1300K site. *Journal of the Korean Society for Information Management*, 34(3), 89-107. <https://doi.org/10.3743/KOSIM.2017.34.3.089>
- Pyo, Soon Hee, Kim, Yun Hyung, Kim, Hye Sun, & Kim, Wan Jong (2015). A study on the developing of big data services in public library. *Journal of the Korean Society for Information Management*, 32(2), 63-86. <https://doi.org/10.3743/KOSIM.2015.32.2.063>
- Republic of Korea. Electronic Gazette (2018). Ministry of the Interior and Safety. Homepage. Retrieved from <http://gwanbo.mois.go.kr/user/intro/definePage.do>
- 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.
- Seol, Moon-Won (2015). A study of online research guides for user support to find and understand archival records in National Archives. *Journal of Korean Society of Archives and Records Management*, 15(1), 53-87. <https://doi.org/10.14404/JKSARM.2015.15.1.053>

