

기록관 웹사이트 검색어의 분석과 활용

- 대통령기록관을 중심으로 -

Analysis and Utilization of Search Terms in Archival Web Sites: A Case Study of Korean Presidential Archives

이 해 영(Hae-Young Rieh)*

목 차

1. 서론	4.1 일반 검색어 분석
2. 이론적 배경	4.2 대통령 이름별 분석
2.1 검색어 분석	4.3 검색어 주제 영역 분석
2.2 대통령 기록관	5. 검색어 분석 결과 활용 방안
2.3 콘텐츠 서비스	5.1 대통령 기록관 콘텐츠 서비스를 위한 적용 방안
3. 연구 설계	5.2 기록정보 서비스를 위한 방법론으로서의 검색어 분석
3.1 검색어 분석 방법	6. 결론 및 제언
3.2 제한점	
4. 대통령기록관 웹사이트 검색어 분석	

<초 록>

본 연구에서는 기록관 웹사이트의 검색어 분석 결과를 이용자에게 대한 기록 정보서비스에 효과적으로 활용할 수 있을지 알아보고자 하였다. 특히 검색어 분석을 콘텐츠서비스 등 기록정보서비스 개발의 한 방법론으로 검토해보고자 하였다. 이를 위해 공기록관이자 주제기록관으로서 일반 이용자들의 폭넓은 관심을 끌만한 특징을 가지고 있는 대통령기록관의 검색어 데이터를 수집하여 분석하였다. 대통령기록관의 웹사이트 검색어 분석 결과는 (1) 일반 검색어 분석; (2) 대통령 이름별 분석; (3) 검색어 주제 영역 분석 등 세 가지로 제시되었다. 검색어 분석 결과는 콘텐츠서비스 및 웹사이트 메뉴 설정, 전시, 교육 등 다양한 서비스에 실질적으로 활용될 수 있을 것으로 판단되었다.

주제어: 대통령 기록관, 웹사이트, 검색어 분석, 기록정보서비스, 콘텐츠서비스

<ABSTRACT>

This study investigated whether search terms analysis of archival Web sites could be utilized effectively for archival information services. The focus was on designing a methodology which brings the search terms analysis and development of archival information services closer, especially for the contents services. The data were collected from the Presidential Archives because it can be characterized as a public archives as well as subject archives. It also tends to draw interests from a broad range of general public. The analysis was conducted with respect to three dimensions: (1) general search terms; (2) names of individual president; (3) subject categories of search terms. The results of search terms analysis have a number of practical implications for developing archival information services including contents services, decision on the menu of the Web sites, exhibition, and education.

Keywords: Presidential Archives, archival Web site, search terms analysis, archival information services, contents services

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

■ 접수일: 2011년 5월 15일 ■ 최초심사일: 2011년 6월 1일 ■ 게재확정일: 2011년 6월 22일

1. 서론

많은 사회구성원들이 웹을 통해 업무를 보고 웹을 통해 정보를 제공받는 것을 당연시하는 사회가 되어가면서, 기록관의 서비스도 웹 기반으로 제공되는 것이 일반화되어가고 있다. 또한 유비쿼터스 환경의 구현에 따라, 앞으로 기록관도 기록정보들을 스마트폰이나 아이패드 등 모바일 환경에서도 활용할 수 있도록 해야 할 것이며, 기록들을 가공하여 콘텐츠로 제공하는 것이 더욱 중요해질 것이다.

기록을 찾는 이용자들의 요구에 부합하면서 이용자들에게 실질적인 도움이 되는 서비스를 기록관에서 제공하는 것이 중요하다는 인식이 확산되면서, 특히 웹을 통해 이용자들이 접근하고 활용하기 원하는 기록은 어떤 종류와 주제가 될지를 이해하고, 이에 대해 적극적으로 대응하는 것은 매우 중요한 일이 되어가고 있다. 추후의 이용자 요구에 대비하기 위해서는 현재까지의 이용자의 활용에 대해 이해하는 것이 바탕이 될 수 있다.

이용자의 요구를 파악하는 방법에는 설문이나 인터뷰를 통하는 방법, 관찰이나 실험을 통하는 방법, 이용자들의 웹에서의 행적 및 검색어 분석 결과를 이용하는 방법 등 여러 가지 다양한 방법이 있다. 각각의 방법은 나름대로 장단점이 있으며 그 결과가 제시하는 내용도 다양하게 서비스에 적용될 수 있을 것이다.

본 연구에서는 이용자들의 기록 활용에 대한 요구에 대처하기 위해 서비스를 어떤 내용으로 어떻게 제공하는 것이 적절할지를 알아보기 위한 방법으로, 특히 웹사이트에 입력된 검색어 로그 분석 결과를 활용하고자 하였다. 검색어

로그는 웹사이트에서 정보를 찾는 이용자와 검색 시스템 사이에서 검색 과정 동안에 진행된 상호작용에 대한 기록이다.

웹사이트의 로그 분석은 다양하게 이용되어 왔다. 박소연, 이준호(2007)는 로그 분석 방법이 이용자의 실제 검색 행위를 사실적으로 반영한다고 하고, 로그 분석을 통하여 대다수 이용자의 전반적인 이용 행태를 분석할 수 있으므로, 이는 이용자 계층이 다양하고 이질적인 웹 검색 분야의 연구에 적합한 방법이라고 하였다. 이들은 이용자들의 전반적인 검색행태 분석, 검색행태 추이분석, 키워드 마케팅 전략구축, 서비스 활용도 평가지표, 개별 서비스 평가, 이용자의 항해경로 조사, 클릭어뷰즈 파악, 클릭 로그 품질 평가, 지역관심사, 개인 이용자의 관심사 파악, 오타 분석 등의 다양한 활용 방법을 제시하였다.

본 연구에서는 특히 기록관 웹사이트를 방문한 이용자들이 입력한 검색어들을 분석해 보고, 그 결과를 기록정보 서비스를 위한 기초자료로 활용할 수 있을지 알아보하고자 하였다. 이를 위해 공공기록물 관리 기관이면서 동시에 주제기록관이라고 볼 수 있고, 또한 많은 일반인들이 관심을 갖고 찾아볼 수 있는 특별한 기록들을 보유하고 있는 대통령기록관의 웹사이트 검색어 분석 결과를 활용한다.

우리나라의 대통령기록관은 2007년 제정된 [대통령기록물 관리에 관한 법률]에 의거해 2007년 개관하였으며, 대통령과 대통령의 보좌기관·자문기관 및 경호업무를 수행하는 기관과 대통령직인수위원회가 생산하고 접수하여 보유했던 국정의 핵심기록인 대통령기록물¹⁾을 수집·정리·보존하고 국민들이 활용할 수 있도록 서비스하는 영구기록물관리기관이다. 대통

령기록물은 국가의 주요정책 결정과정을 보여 주는 증거 자료로, 생산시점의 정치, 사회, 경제, 문화를 파악할 수 있는 정보자료로서의 중요성이 크기 때문에 대통령기록관의 설립은 큰 의미를 가진다(최재혁, 김희섭 2009, 79).

대통령기록관에서는 다양한 측면에서의 업무가 진행되고 서비스가 제공되고 있으며, 특히 2008년 6월 웹사이트의 개설과 함께 역대 대통령의 기록에 대한 정보를 제공하고 있는데, 최근에는 대통령기록포털을 새로 개설하여 서비스를 제공하고 있다. 온라인 콘텐츠도 대통령 온라인 기록관, 취임식, 선물갤러리, 헌법개정사, 연설기록, 행정박물, 어린이 역사교실, 기증기록물, 정책간행물 등에 대한 내용 등 다양하게 제공되고 있다.

이러한 대통령기록관 웹사이트의 이용자들이 특히 어떠한 내용들을 검색해보고자 하였는지 알아보는 것은 그 자체만으로도 의미가 있을 것이며, 또한 어느 기록관 웹사이트에서나 검색어를 분석하여 기록관의 서비스에 활용할 만한 결과를 얻을 수 있다고 판단되면 이 또한 의미 있는 일이 될 것으로 판단하였다.

본 연구에서는 먼저 다양한 로그 분석과 관련된 선행 연구들을 살펴보고 이들의 분석 방법을 반영하여, 본 연구 대상인 대통령 기록관에 적용한 후 검색된 용어들을 분석하여 살펴보고자 한다. 또 검색어 분석 결과가 기록정보 서비스를 위한 유의미한 내용으로 제시될 수 있을지, 그렇다면 그 내용이 제시하는 방향은

어떤 것인지도 검토해보도록 하겠다.

본 연구의 목적은 다음과 같이 제시될 수 있다.

1. 대통령기록관 검색어의 특징은 무엇인지, 검색어 분석으로 나타난 이용자들의 요구를 확인해본다.
2. 검색어 분석 결과가 대통령기록관 기록정보 서비스에 활용될 수 있는 의미 있는 내용을 제시하는지, 특히 콘텐츠 서비스에 적용될 수 있는 내용은 무엇인지 검토해본다.

본 연구는 대통령 기록관을 대상으로 함으로써, 대통령기록이라는 특정 기록에 대한 이용자들의 검색 패턴을 알아본다는 사실에도 의미가 있을 것이며, 이 결과가 유의미하다면 다른 기록관에서의 활용도 검토될 수 있을 것이다. 또한 기록관에서 로그 데이터를 활용하여 검색어 분석을 시행한다면 어떤 사항에 유의해야 할 것인지 등에 대해서 알아본다는 데에도 연구의 의의가 있다고 할 수 있겠다.

2. 이론적 배경

2.1 검색어 분석

웹사이트 로그 분석은 웹서비스가 제공되기 시작하면서 주목을 받아왔다. 여러 연구자가 특히 검색엔진에 입력된 질의어들을 분석하여

1) [대통령기록물 관리에 관한 법률 2조1항] “대통령기록물”이란 대통령의 직무수행과 관련하여 대통령과 대통령의 보좌기관·자문기관 및 경호업무를 수행하는 기관과 대통령직인수위원회가 생산·접수하여 보유하고 있는 기록물 및 물품을 말한다. [2조2항] “대통령기록관”이란 대통령기록물의 영구보존에 필요한 시설 및 장비와 이를 운영하기 위한 전문 인력을 갖추고 대통령기록물을 영구적으로 관리하는 기관을 말한다.

질의의 패턴을 살펴보고 이용자들의 검색 특징을 살펴보았다(Silverstein, et al, 1999; Jansen, Spink, & Saracevic 2000; Ross & Wolfram 2000; Spink et al, 2001; 박소연, 이준호 2002; Wang, Berry, & Yang 2003; Spink & Jansen 2004; Spink & Jansen 2004; Jansen & Spink 2005; Jansen, Spink and Pedersen 2005; Park, Lee, & Bae 2005; 박소연, 이준호 2005; 박소연, 이준호, 김지승 2005; 박소연, 이준호 2007; Du 2010; 박소연 2010). 초기의 웹 검색엔진에 대한 로그 분석 연구들은 웹 검색 이용자들이 적은 수의 검색어로 구성된 단순한 질의를 통해 정보 검색을 수행하는 경향이 있었음을 밝혀내었다(박소연, 이준호 2002, 113).

한편 박소연, 이준호(2002)는 2002년 하루 동안의 네이버 검색 트랜잭션 로그를 분석하였는데, 몇 가지 연구 방법을 검토하여 세션 정의를 위한 방법들을 검토하고 연구에 활용하였다. 이들의 네이버 검색어 로그에 대한 분석 결과, 검색어가 어절 단위로 정의되는 경우 입력 질의별 평균 검색어 수는 1.5841, 형태소 단위로 정의되는 경우 2.5450이었다. 하나의 어절단위 검색어만 포함하는 질의가 67.31%, 형태소 단위로 정의해 하나의 검색어만 포함하는 입력질의는 30.03%였다고 보고하였다. 박소연, 이준호(2005)는 또한 2003년과 2004년 1년 사이의 네이버 검색 질의어를 분석하여, 계절별 질의 주제, 주중과 주말의 질의 주제, 요일별 질의 주제, 날짜별 검색 행태를 분석하였다. 1년간 변화는 거의 없었으나, 겨울엔 엔터테인먼트와 게임 관련 질의가 많고, 봄에는 교육/학문 관련 질의가 많았으며, 지역/여행 관련 질의는 여름에 높고, 컴퓨터 관련 질의는 모든 계절에 다

높았다. 주말은 엔터테인먼트, 게임 및 컴퓨터 관련 질의가 높고, 기업, 경제, 교육, 기관 등과 관련된 주제는 주중에 높은 것을 발견하였다. 이 연구들에서 차용한 방법론은 본 연구의 데이터 분석이 도움이 되었다.

Swanson(1977)은 정보검색이 원하는 문서의 속성에 대해 처음엔 추측으로 시작하여 검색시스템의 반응에 대해 또 다른 시도로 수정하는 상호작용적이며 반복적인 시행착오의 과정이라 하였다. Rieh & Xie(2006)는 정보검색을 상호작용적이며 반복적인 절차로 보고, 검색엔진 활용에 있어서, 복합 웹 질의 재구성의 측면과 패턴을 재구성 차례에 초점을 맞추어 검토하였다. 30분 이내에 후속 질의어가 입력된 경우를 하나의 세션으로 간주하고, 6개 이상의 고유 질의어를 입력한 313개 세션을 분석하였다. 데이터에서 내용(content), 형태(format), 자원(resource)의 세 가지 재구성 측면과 9개 하위측면이 추출되었다. 또한 Chung & Yoon(2009)은 검색엔진의 검색어 로그를 분석하여 이미지 특성 범주와 관련해 질의재구성 패턴을 연구하였다. 질의에서 가장 많은 부분은 특정어(specific)와 지칭어(namable)가, 질의 재구성에 있어서는 평행이동이 가장 많았다. 60-80%의 경우에 동일 범주에서 검색질의가 진행되었다.

Jansen(2009)은 기존의 검색 로그에 의한 연구들을 종합적으로 분석하고, 사용된 방법론들을 제시하였으며, 검색로그의 활용분야와 연구 현황 등을 제시하였다. 특히 본 연구에서는 제시된 로그분석 방법론 중에서 데이터 정리에 관련된 일부 내용을 차용하였다.

Du(2010)는 42명의 대학생에 대한 웹 탐색 과정에 대한 관찰과 인터뷰 등을 통해, 웹 탐색

과정에서 이용자들이 멀티태스킹을 시행하는 것은 복잡한 정보탐색 문제를 다양한 작업의 전환을 통해 진행하는 것이라고 하였다. 웹 탐색 전과 후의 정보문제에 대한 인식과 전체적인 정보 평가에 의한 변화와 탐색과정에서의 초점의 변화에 의해서는 인지 변화가 일어나고, 정보과업 조정 및 전략조정 등의 인식 조정은 멀티태스킹과 인식 변환을 연결한다고 하였다.

최동진 외(2010)는 웹 로그 데이터 중 사용자 질의어를 중심으로 사용자의 성향 및 관심 분야를 파악하고, 어휘사전과 비교분석하여 더욱 정확하게 웹문서를 분석할 수 있는 모델을 설계하여, 사용자의 의도에 적합한 웹문서를 파악하고 사용자의 관심분야를 예측하고자 하였다. 박소연(2010)은 네이버에서 이용자들이 입력한 통합 검색 질의 로그와 검색 결과에서 이용자들이 조회한 문서에 대한 클릭 로그를 분석하여, 이용자들의 멀티미디어 검색 행태를 알아 보았다. 이용자들의 질의에는 음악 관련 질의가 가장 많았으며, 이용자의 조회가 가장 많은 것은 동영상 컬렉션이었고, 엔터테인먼트와 관련된 주제가 가장 조회를 위한 클릭 수가 많은 것을 발견하였다. 선행 로그분석 연구들은 검색엔진 중심으로 많이 진행되어 왔다.

본 연구는 기록관의 검색어들을 분석한 것으로 다른 연구와 차별화를 기하였고, 특히 대통령기록에 대한 검색어를 분석하여 이용자들이 관심 있어 하는 주제들을 파악하여, 실질적인 활용을 염두에 둔 연구로 의의가 있다고 할 수 있겠다.

2.2 대통령 기록관

대통령 기록관은 대통령이 한 국가의 최고 통치자라는 측면에서 소장 기록 자체의 중요성이 매우 큰 기관이다. 특정 시기에 큰 반향을 일으켰던 이슈나 정책에 대해서 대통령기록은 중요한 연구의 대상이 될 수 있을 뿐 아니라, 일반 국민들도 기록에 남겨진 사실들에 대해 많은 관심을 가질 수 있다. 이러한 대통령 기록의 중요성은 기록정보 서비스의 중요성을 시사한다. 즉 기록에 대해, 특히 연구자나 국민들이 관심을 가질만한 이슈에 대해서는, 공개될 수 있는 기록들의 콘텐츠화 등을 통해 쉽게 활용할 수 있도록 적극적으로 서비스 하는 것이 필요한 것이다.

또한 대통령에 대한 기록은 한 시대 통치자의 국가 정책 추진 방향과 특정 정책의 개발 및 시행의 의의를 파악할 수 있을 뿐 아니라, 역대 대통령 개인의 족적을 반영하고 있기 때문에, 여러 나라에서 대통령 기록관은 일반 국민의 관심을 반영하여 관광 명소로 사랑받기도 하고 있으며, 또 온라인상으로 많은 방문을 유도할 수도 있다.

우리나라의 대통령 기록관은 신생 조직으로, 아직 많은 기록물을 소장하고 있지는 못하나 이러한 기록의 중요성은 간과할 만한 것이 아니다. 그래서 대통령기록에 대한 연구들도 여러 가지로 진행되어 왔고 발표된 바 있다. 대부분은 대통령기록관의 기능, 역할, 대통령기록 관리의 제도, 대통령기록 관리 법체계, 대통령기록물 보존 및 관리 등에 대한 내용들이었다(주희경 1999; 박건홍 2001; 이상민 2001; 김성수, 서혜란 2002; 이승휘 2002; 김관석, 김관

보, 권영주 2003; 남태우, 오지영, 유보현 2007; 이선정 2007; 김향란 2008; 조영삼 2009; 최성렬 2009). 또한 대통령기록의 수집, 분류, 기술 등에 대한 연구도 진행되어 왔다(홍원기 2004; 강경희 2008; 박성경 2008; 손성호 2008; 정상화 2008; 구분윤 2009; 정진화 2010).

이 외에 대통령기록관의 서비스와 관련하여서는 조민지(2006)와 전용호(2008)가 연구를 진행한 바 있으며, 최재혁, 김희섭(2009)은 '역대 대통령 웹기록 서비스'의 이용자 만족도를 조사, 분석하여 이용 만족도에 영향을 미치는 요인을 분석하고, 검증과정에서 나타난 웹기록 서비스의 문제점 지적 및 개선 방안을 모색하였다. 분석결과 정보품질과 시스템품질 만족도가 각각 이용자의 전반적인 만족도에 통계적으로 유의한 영향을 미치는 것으로 나타났고, 이용자들의 지속적인 이용과 관련하여서는 정보품질과 서비스품질, 이용만족도가 유의한 영향을 미치는 것으로 나타났다.

또한 김연희(2010)는 대통령기록관 웹사이트 사용성 평가 연구를 통해 웹사이트의 효율성을 연구하였고, 남건우(2008)는 대통령 기록관을 중심으로 하여 웹기록 이관의 문제점과 개선방안에 대한 연구도 살펴보아, 대통령 기록의 웹사이트와 웹기록에 대한 관심도 상당함을 알 수 있었다.

Drake(2007)에 의하면, 1978년 PRA(Presidential Records Act of 1978) 이전에는 미국의 대통령들도 대통령 기록물을 개인 소유물로

생각했었다고 하며, PRA에서 처음으로 대통령과 부통령의 기록이 정부소유로 미국 국민에 속한다고 선언하였다. PRA에 의하면 대통령의 임기 후 5년간은 기록에 대한 접근이 제한된다.

현재 미국의 대통령기록은 NARA(U.S. National Archives and Records Administration)의 Office of Presidential Libraries에서 관할하는 13개의 대통령도서관(Presidential Libraries)들이 관리하고 있는데, 이들 도서관은 기록관과 박물관을 겸한 성격을 가진다.²⁾ 13개 대통령도서관은 통합검색 또는 개별 도서관 검색이 가능하다.

한편 Conway(1986)는 미국 4개의 대통령도서관 이용자들에 대한 설문조사 연구를 진행하여 114명의 응답을 분석하였다(대학생 11, 대학원생 19, 교수 58, 전문가 21, 개인 5). 개별적인 이용 경험과 사전 준비가 없던 그룹은 오히려 결과에 놀라고 더 만족했다고 나타났는데, 이는 기대 이상의 결과를 얻었기 때문으로 해석되며, 기록에 대한 오리엔테이션 인터뷰와 정보서비스의 확실성이 이용자 만족도에 영향을 준 것으로 나타났다. 이는 대통령기록관에서도 필요한 정보서비스의 중요성을 다시 한 번 일깨워준 예라고 할 수 있겠다.

2.3 콘텐츠 서비스

문화산업진흥 기본법에 의하면(2조) '콘텐츠'란 부호·문자·도형·색채·음성·음향·이

2) 13개 Presidential Libraries는 다음과 같다. Herbert Hoover Library, Franklin D. Roosevelt Library, Harry S. Truman Library, Dwight D. Eisenhower Library, John F. Kennedy Library, Lyndon B. Johnson Library, Richard Nixon Library, Gerald R. Ford Library, Jimmy Carter Library, Ronald Reagan Library, George H. W. Bush Library, William J. Clinton Library, George W. Bush Library.
출처: <<http://www.archives.gov/presidential-libraries/>>.

미지 및 영상 등(이들의 복합체를 포함한다)의 자료 또는 정보를 말하며, '디지털콘텐츠'란 부호·문자·도형·색채·음성·음향·이미지 및 영상 등(이들의 복합체를 포함한다)의 자료 또는 정보로서 그 보존 및 이용의 효용을 높일 수 있도록 디지털 형태로 제작하거나 처리한 것을 말한다.

기록학분야에서의 콘텐츠는 역사적 사건이나 인물, 행위 등에 관한 것이 대부분이므로, 사회적, 역사적 맥락을 염두에 두고서 주제를 다루어야 한다고 제시되었고(한국국가기록연구원 2007), 전수진(2008)은 기록정보콘텐츠는 기록을 원자료로 하여 제작 또는 가공된 것으로, 관련된 기록과 함께 구조화된 것으로, 기록과 맥락정보를 연관관계에 따라 구조화하고 해석을 덧붙임으로써 이용자에게 제공되는 정보라고 정의하였다. 설문원, 김익한(2006)은 기록의 활용범위를 넓히고 점차 다양해지는 이용자 수요에 대응하기 위해서는 기록 안에 담긴 풍부한 정보를 디지털 콘텐츠로 가공하여 지적 접근성(intellectual accessibility)을 높일 필요가 있다고 주장하며, 기록을 토대로 한 역사콘텐츠는 연속적인 흐름을 보여주어야 한다고 하였다. 심성보(2007)는 기록정보콘텐츠에 대해, 고객(이용자)의 요구에 기반하여 고객의 기록정보에 대한 접근성을 제고하는데 필요함을 역설하고, 기록정보 서비스를 추진할 때에 이용자(고객)를 세분화하여 이용자(고객)별로 차별화된 서비스를 제공하기 위한 노력이 경주되고 있음에 주목하였다. 이러한 기존 연구들은 기록콘텐츠 서비스의 의미와 중요성을 주장하면서 동시에,

콘텐츠 서비스가 이용자를 고려하여 맥락정보와 함께 제공되는 것이 중요함을 제시하였다.

본 연구에서는 서비스 제공에 있어, 특히 이렇게 이용자를 고려한 콘텐츠 서비스를 제공하는 데에, 기존의 웹사이트 이용자가 검색한 결과를 활용하는 것이 콘텐츠 주제와 내용 구성에 적절할 것이라고 보고 연구를 진행하였다.

3. 연구 설계

3.1 검색어 분석 방법

본 연구를 위해 대통령 기록관 웹사이트에서 2010년 1월부터 11월 사이에 검색 상자에 입력된 검색어 데이터를 입수하였다.³⁾ 여타 검색 로그 연구에서 활용한 것과 같이, 입력된 검색어와 검색어가 입력된 IP주소와 입력날짜 및 입력시간이 포함된 형태의 데이터를 분석하고자 하였으나, 그러한 형태로 데이터를 입수하는 데에는 한계가 있어서, 다음과 같은 두 가지 조건을 충족하는 데이터를 요청하여 입수하였다.

첫째, 같은 IP주소에서 101개 이상 연속해서 질의어가 입력된 경우는 제외하는 것으로 하였다. Jansen(2009)은 검색 로그 분석 연구들을 종합적으로 정리하면서, 시스템에서 자동으로 생성되는 세션과 사람이 검색어를 입력한 세션을 구분하여 사람이 입력한 데이터만 활용하고자 할 때에는, 대부분의 로그 분석에서 101개 이상 연속해서 같은 IP주소에서 검색어가 입력

3) 대통령기록관에 정보공개 청구 활용.

된 경우는 분석 과정에서 제외하는 것으로 해왔다고 하였다. 이에 의해, 본 연구에서도 그러한 세션은 검색어 분석에서 제외하는 것으로 하였다. 둘째, 같은 IP주소에서 동일한 검색어가 연속해서 바로 이어 입력된 경우는, 박소연, 이준호(2002) 및 Jansen(2009)이 기술한 바에 의하면, 새로 고침 등을 통해 검색어가 자동으로 재입력된 경우로 판단하여 분석 대상에서 제외하는 것이 적절하다고 하였다.

본 연구에서도 앞의 두 가지 조건에 해당되는 데이터는 제외된 데이터를 입수하였다. 즉 2010년 1월부터 11월 사이에 입력된 검색어와 입력이 이루어진 IP주소 및 입력 날짜와 시간이 제시된 데이터에 대해, 먼저 IP주소 순으로 일차 정렬하고, 같은 IP주소에 대해서는 날짜 순과 시간 순으로 정렬한 후, 같은 IP주소에서 101개 이상 연속 입력된 데이터와 같은 용어의 재입력에 의한 중복 검색어를 데이터 세트에서 제외시켜줄 것을 요청하였다. 이러한 과정을 거쳐 정렬된 상태의 검색어 데이터를 엑셀 파일로 입수하였다. 입수된 데이터에서 깨진 글자나 부호(da ? / ~) 및 어휘로 성립되지 않는 영어, 예를 들면 shangus(노무현의 영어 자판 타자어) 등을 제거한 후 6,790건의 검색어들이 최종적으로 분석되었다.

검색어 빈도수는 검색어를 가나다순으로 정렬한 후 엑셀의 [부분합] 기능으로 검색어가 나타난 빈도수를 계산하였으며, 복합어에 대해서는 빈도수가 높은 검색어를 대상으로 [모두 찾기] 기능을 통해 갯수를 계산하였다. '대통령', '대통령' 또는 '대통령' 등 '대통령'을 입력한 것이 확실하다고 판단되는 경우는 빈도수 분석에 포함하였다.

3.2 제한점

바로 앞에서 기술했듯이 IP주소 숫자가 같은 곳에서 검색어 입력이 101개 이상 연속적으로 계속된 경우는 시스템에서 자동 생성되었을 것으로 보고 이들을 제외한 데이터를 입수하였는데, 이 과정에서 의미 있는 검색어들이 제외되었을 가능성이 있으나, 이들 중 일부를 선택하여 적용하는 것은 더 바람직하지 않다고 판단하여 제외하는 것으로 결정하였다.

그리고 대통령기록관 웹사이트의 통합검색창 아래에는 추천키워드가 8개 제시되어 있었다(<그림 1> 참조). 가장 많이 입력된 검색어 1-8위는 모두 이 검색창 아래에 제시된 추천키워드가 차지했다. 따라서 실제 모든 검색어가 이용자들의 기록 검색의 필요성이나 관심사를 반영하였다기보다 이용자들이 웹사이트에 들어왔다가 추천키워드 중에 하나를 입력한 경우가 많았던 것으로 판단된다. 따라서 검색어 빈도수가 제시하는 의미는 한정된 것으로 보인다. 그러나 이 키워드 중에도 선호도가 있었고, 또 그 외에 다른 검색어들도 많이 입력되었으므로, 이용자들의 관심사를 반영하지 않는다고 볼 수는 없다.

또한 2010년 1월부터 11월 사이에 입력된 검색어들에 대한 분석을 진행했는데, 이 기간 동안에는 노무현 대통령과 김대중 대통령 서거 1주기가 포함되어 있어 관련 검색어가 증가하였을 수 있으나, 전체적인 빈도 분석에 있어서 이러한 영향을 반영할 수 없었다. 이와 같은 제한점이 있을 것이나, 본 연구 결과를 객관화하는데에 영향을 미칠 정도는 아니라고 판단되어 연구를 진행하였다.



〈그림 1〉 대통령기록관 웹사이트

4. 대통령기록관 웹사이트 검색어 분석

4.1 일반 검색어 분석

대통령 기록관 웹사이트에서 2010년 1월부터 11월 사이에 입력된 검색어를 정렬한 6,790건의 용어들을 분석한 결과, 앞의 연구 제한점에서 설명하였듯이, 가장 많이 입력된 용어 1-8위는 대통령기록관 웹사이트의 통합검색상자 아래에 제시된 추천키워드들이 차지하였다. 추천키워드는 대통령 임기, 대통령의 꿈, 대통령님, 대통령선거, 대통령, 대통령선물, 대통령연설, 대통령기록의 차례로 제시되었다. 이러한 추천키워드들의 적절성에 대한 것은 논외로 한다.

검색어 중에 가장 많이 입력된 용어는 '대통령연설'(496회, 전체 검색어 6,790개 중 7.3%)이었고, 그 다음이 '대통령기록'(447회, 6.6%)이었

으며, '대통령'이라고 단독 입력한 경우가 429회(6.3%)로 나타났다(〈표 1〉 참조). 그런데 전체적으로 보면 '대통령연설'이라는 용어가 가장 많기도 했지만, 이 외에도 '연설문'이 포함된 경우가 모두 83회였으며, 그 외에, '연두연설', '시정연설', '국정연설', '신년연설' 및 대통령 이름과 '연설'이라는 용어를 결합하여 검색한 경우가 78회, '취임사'가 단독 32회를 포함하여 복합어 59회, '담화' 또는 '담화문'이 포함된 경우가 20회, '경축사'가 포함된 경우가 14회, '회견'이 포함된 경우가 6회, '신년사' 포함이 6회, '치사' 포함이 5회 등으로 나타나, 매우 많은 사람들이 대통령의 연설과 회견 및 담화문 등에 대해 많은 관심을 갖고 검색을 하는 것으로 나타났다. 다음으로 '대통령기록'과 '대통령'이라는 용어가 많은 이유는 추천키워드로 제시된 바도 있지만, 또한 대통령기록관에서 검색을 한 만큼 이와 관련된 용어로서 가장 보편적인 용어라고 볼 수 있겠다.

〈표 1〉 일반 검색어 입력 현황

입력 검색어	빈도	왼쪽 빈도수에 포함되지 않은 관련 용어 및 복합어
대통령연설	496(7.3%)	그 외 '대통령연설' 포함 16, 3.1절, 4.19 등 기념일 연설문, (*대통령)연설문 등 '연설문' 포함 83, 연두연설, 시정연설, 국정연설, 신년연설 등 '연설' 포함 78, '취임사'(32) 포함 59, '답화(문)' 포함 20, '경축사' 포함 14, '회견' 포함 6, '신년사' 포함 6, '치사' 포함 5
대통령기록	447(6.6%)	'대통령기록관' 포함 24, 그 외 대통령기록관리, 대통령기록물 등 '대통령 기록' 포함 16
대통령	429(6.3%)	대통령님 112, '역대대통령' 포함 108, 'n대 대통령' 포함 56
대통령선물	242(3.6%)	'G20, 선물' 포함 7, 그 외 '선물' 포함 15
대통령임기	212(3.1%)	재임기간 등 '재임' 포함 12
대통령선거	203(3.0%)	선거공약(3), 선거유세(2) 등 '선거' 포함 22
대통령의 꿈	137(2.0%)	
대통령님	112(1.6%)	
역대대통령	82(1.2%)	
청와대	51(0.8%)	청와대브리핑 8, 그 외 '청와대' 포함 19, '청남대' 포함 16
연설문	40(0.6%)	
취임사	32(0.5%)	
혈액형	26(0.4%)	'대통령 혈액형' 포함 12, 박정희, 혈액형 포함 6
국가기록원	23(0.3%)	
정책	22(0.3%)	
대통령이 하는 일	21(0.3%)	'대통령, 일과' 포함 5, '일정' 포함 8

※ 빈도의 %는 전체 검색어 6,790개에 대한 비율; 포함 다음의 숫자는 앞의 용어를 포함한 복합어 빈도수.

다음으로, '대통령선물'도 242회(3.6%)로 이 용자들이 많이 입력하였으며, 또 많이 검색된 일반 용어는 대통령임기(212회, 3.1%), 대통령 선거(203회, 3.0%), 대통령의 꿈(137회, 2.0%) 과 대통령님(112회, 1.6%) 등이 있으며, 이 단어들은 모두 검색창 아래에 추천 검색어로 제시된 용어들이다.

그 외에 일반 용어로 이용자들이 직접 단독으로 많이 입력한 용어는 역대대통령(82회), 청와대(51회), 연설문(40회), 취임사(32회), 혈액형(26회), 국가기록원(23회), 정책(22회), 대통령이 하는 일(21회) 등으로, 각각 20회 이상 입력되었으며, 그 외에 참여정부(18), 기록관리(16), 참여정부 정책보고서(15), 비전 2030(15), 청남대(13), 농협상품권(13), 사진(11), 취임식(11)

등도 단독으로 여러 번 입력되었다.

4.2 대통령 이름별 분석

앞에서 살펴본 일반 용어 외에 이용자들은 각각의 대통령 이름들을 많이 입력하여 검색을 수행하였다. 먼저 가장 많이 검색된 노무현 대통령은 모두 239회(전체 검색어 6,790개의 3.5%) 입력되었는데, '노무현' 단독 입력된 경우도 132회(1.9%)였고, '노무현대통령'이 7회였다(〈표 2〉 참조). 노무현 대통령에 대해서는 연설에 관심이 많아, '연설'을 포함한 복합어 입력이 18회였고, '정부(조직)' 포함 8회, '정책'을 포함한 경우가 5회, '취임' 포함 4회, '자이툰' 포함 3회 등으로 나타났다. 박정희 대통령이 그 다음으로

〈표 2〉 역대 대통령 이름 포함 검색어 입력 현황

대통령 이름	빈도	단독 및 복합어 입력 빈도
노무현	239(3.5%)	'노무현' 단독 132, 노무현대통령 7, '연설' 포함 18, '정부(조직)' 포함 8, '정책' 포함 5, 취임 포함 4, 자이툰 포함 3, 그 외 '노무현' 포함 56
박정희	206(3.0%)	'박정희' 단독 118, 박정희대통령 33, 박정희 자주국방 5, '혈액형' 포함 5, '연설' 포함 4, '핵(개발)' 포함 3, 그 외 '박정희' 포함 38
이승만	165(2.4%)	'이승만' 단독 94, 이승만대통령 21, '취임' 포함 6, 이승만업적 4, '미국' 포함 4, '연설' 포함 4 그 외 '이승만' 포함 32
김대중	144(2.1%)	'김대중' 단독 61, 김대중대통령 13, '연설' 포함 18, '취임 연설' 7회 취임 연설 제외한 '취임' 포함 8, '노벨상' 포함 3, (노벨(평화)상 모두 9), 그 외 '김대중' 포함 31
이명박	68(1.0%)	'이명박' 단독 35, 이명박대통령 8, '연설' 포함 4, 그 외 '이명박' 포함 21
김영삼	63(0.5%)	'김영삼' 단독 39, 김영삼대통령 8, 그 외 '김영삼' 포함 16
전두환	56(0.5%)	'전두환' 단독 37, 전두환대통령 5, 그 외 '전두환' 포함 14
노태우	51(0.5%)	'노태우' 단독 28, 노태우대통령 9, 그 외 '노태우' 포함 14
윤보선	33(0.5%)	'윤보선' 단독 23, '업적' 포함 6, 그 외 '윤보선' 포함 4
최규하	25(0.5%)	'최규하' 단독 15, 최규하대통령 5, 그 외 '최규하' 포함 5

※ 빈도의 %는 전체 검색어 6,790개에 대한 비율

많아, 복합어와 단독 입력 경우를 모두 합치면 206회(3.0%) 입력되었는데, '박정희' 단독 입력이 118회(1.7%), '박정희대통령' 33회, '자주국방'과 '혈액형'을 포함한 경우가 각각 5회, '연설'을 포함한 경우가 4회, '핵(개발)' 포함이 3회 등으로 나타났다.

이승만 대통령도 모두 165회(2.4%) 입력되었는데, '이승만' 단독 입력된 경우도 94회(1.4%)였고, '이승만대통령' 입력이 21회, '취임'을 포함한 경우가 6회, '업적' 및 '미국'과 '연설'을 포함한 경우가 각각 4회로 나타났다.

김대중 대통령도 복합어와 단독 입력 경우를 모두 합치면 144회(2.1%) 입력되었는데, '김대중' 단독 입력된 경우가 61회, '김대중대통령'이 입력된 경우가 13회, '연설'을 포함한 경우가 18회, 그 중 '취임 연설'이 7회, '취임'만 따로 포함된 경우가 8회, '노벨상'과 '김대중'이 같이 입력된 경우는 3회였고, '노벨(평화)상'은 모두 9회가 입력되었다.

이명박 대통령은 복합어와 단독 입력 경우를 모두 합치면 68회로, '이명박' 단독 입력이 35회, '이명박 대통령' 입력이 8회, 그리고 '연설' 포함이 4회 등으로 나타났다. 김영삼 대통령도 모두 63회 입력되었고, 전두환 대통령은 모두 56회, 노태우 대통령 51회, 윤보선 대통령 33회, 최규하 대통령 25회 등이 입력되었다.

대통령 이름이 입력된 경우, 역시 연설이라는 용어가 같이 입력된 경우가 많았으며, 그 외에 정부조직, 정책, 취임 등이 여러 번 포함되었고, 각 대통령들의 업적과 업무 수행관련 용어들이 고루 입력되었다.

4.3 검색어 주제 영역 분석

전체 검색된 용어들 중에서 3회 이상 입력된 검색어들을 범주화하여 구분하였더니, 약 8개 주제영역으로 나누어 볼 수 있었다(〈표 3〉 참조). 단독어로 여러 번 입력된 용어에 대해서는

그 용어가 복합어의 일부로 입력된 경우도 빈도를 확인하여 주제 영역에 할당해 보았다. 임의로 8개 영역 정도로 구분하다보니 애매하거나 적합하지 않다고 판단되는 부분도 있었으나, 편의상 검색어들을 살펴보기 위해 영역을 더 넓히지는 않았다. 이렇게 나누어 본 주제 분

야는 1) 정책, 업적 및 업무수행 분야, 2) 대통령 및 주변인 이름과 개인사, 3) 기록관리, 4) 행정박물 및 매체, 5) 조직도, 정부부서 및 기관, 단체, 6) 선거, 7) 외교 및 남북관계, 8) 행사 등이었다.

〈표 3〉 주제 분야 별 검색어 입력 현황

주제분야	입력 용어	빈도
정책, 업적 및 업무수행	대통령연설 495, 대통령입기(212) 등 '입기' 포함 229, 대통령의 꿈 138, 정책(22) 포함 98, 연설(문) 50, '교육' 포함 37, (대통령 이름, 대통령)취임사 59, '새마을(운동)' 포함 33, 대통령업적(8) 등 '업적' 포함 28, (국가)비전2030 27, '복지' 포함 24, '국무회의' 포함 20, 대통령(이) 하는 일 18, '노무현, 연설' 포함 18, '부동산(정책, 대책)' 포함 16, '독도' 포함 16, 참여정부 정책보고서 15, '세종시' 포함 13, '노벨(평화)상' 포함 9, '상생협력' 포함 9, '국정과제' 포함 8, (노무현)탄핵 8, '일정포' 포함 8, '지주국방' 포함 8, '제임기간' 포함 8, 대통령령 7, (대통령)보고서 7, '신년(국정)연설' 7, '위기관리' 포함 7, (5공)청문회 포함 7, 경부고속도로 6, 새만금 6, '언론' 포함 6, 저출산(고령화) 6, 담화문 5, '리더십' 포함 5, '호주제' 포함 5, 문화재 4, 연설문집 4, 인수위백서 4, 청와대브리핑 4, 채용 4, 해외건설(수출) 4, 결과보고 3, '국정목표' 포함 3, 관보 3, 민속촌 3, 자원봉사 3, 자전기 3, 집단소송(제도) 3, 핵융합 3	1,514 (22.3%)
대통령 및 주변인 이름과 개인사	대통령(님) 539, 박정희(대통령) 151, 노무현(대통령) 139, 이승만(대통령) 115, 역대대통령 39, 김대중(대통령) 74, 김영삼(대통령) 47, 대통령혈액형(10) 등 '혈액형' 포함 44, 이명박(대통령) 43, 전두환(대통령) 59, 노태우(대통령) 53, 윤보선(대통령) 25, 최규하(대통령) 20, (제)1대대통령 포함 16, (제)2대대통령 8, 영부인 6, 초대대통령 5, 가족(관계) 5, (제)15대대통령 4, 김재규 4, 우리나라역대대통령 4, 전직대통령 4, 권양숙 3, 유영수(여사) 3, 프란체스카 3	1,413 (20.8%)
기록관리	대통령기록 447, 국가기록원 23, 기록관리 16, 대통령기록관 8, 나라기록관 6, 표어 5, 관장 4, 대통령기록관리 4, 참여정부 5년의 기록 4, 공개재분류 3, 기록물관리법 3, e-知園 3	526 (7.7%)
행정박물 및 매체	(대통령)선물 246, 사진 11, 대통령사진(기록) 8, (대통령이름)(친필)휘호 8, 'G20(특별)선물' 포함 7, (대통령)초상(화) 7, 동영상 4, 행정박물 4, (대통령)서명 3, 싸인 3, '훈장' 포함 3	304 (4.5%)
조직도, 정부부서 및 기관, 단체	'청와대' 포함 77, '참여정부' 포함 62, 인수위(위원, 위원회) 포함 28, 그 외 *위원회(교육개혁위원회, 혁신위원회, 정책기획위원회 저출산고령사회위원회 등) 27, (정부)조직(도) 17, '청남대' 포함 16, (대통령)비서실 8, '국방부' 포함 6, '비서관' 포함 5, '비서실장' 포함 5, 기관 4, 노사모 4, 대통령실 4, 법원주사보 4, 사단장 3	270 (4.0%)
선거	대통령선거 203, '선거'(4) 포함 14, 공약 포함 20, 대선 포함 7, 당선(자) 포함 6, (대통령)후보 포함 5, 유세 포함 3	258 (3.8%)
외교 및 남북관계	'외교'(3) 포함 24, 'G20' 포함 20, 한일관계담화(4), 한일공동선언(3) 등 한일 포함 19, '미국'(3) 포함 17, 'FTA'(8) 포함 17, '통일'(4) 포함 15, '일본'(3) 포함 14, '북한'(5) 포함 13, '남북정상회담' 포함 12, 그 외 정상회담 포함 10, 대북(정책) 5, '대일' 포함 5, 오키나와 5, '친서' 포함 5, 캐나다(캐나다) 5, 동북아(시대) 4, 러시아 4, 오바마 4, 외국인 4, 청구권(자금) 4, 한인(회) 4, 교민 3, 남북군사실무회담(결과) 3, 도라산역 3, 동아시아공동체 3, '방북' 포함 3, 북괴 3	228 (3.4%)
행사	'취임식', '대통령취임' 등 '취임' 포함 116, '국군의 날'(4) '어버이날' 등 '날' 포함 17, (광복절)(8.15)경축사 14, (어린이)(현장)보도대회 11, '방문' 포함 11, 노벨(평화)상 포함 9, 전국새마을지도자(대회) 8, 공모(전) 6, '전시(회)' 포함 5	197 (2.9%)

※ *은 여러 가지 용어 생략 표시,
()안의 숫자는 해당용어만 단독으로 입력된 횟수,
빈도의 %는 전체 검색어 6,790개에 대한 비율

가장 입력이 많고 이용자들이 다양하게 많은 관심을 가진 분야는 '정책, 업적 및 업무수행 영역'으로 묶을 수 있는 주제 분야로, 총 1,514개 (전체 6,790개 검색어의 22.3%)의 검색어를 묶을 수 있었으며, 이에 포함된 '대통령연설', '대통령 임기', '대통령의 꿈'은 대통령기록관 웹사이트 검색창 아래 추천키워드로 제시되어 많이 입력된 용어들이다. 그 외에 '정책'이 단일 검색어로 22회 입력된 것을 포함 복합어로 입력이 98회, 대통령이름 또는 '대통령'이라는 용어와 함께 많이 검색된 '취임사'가 59회, 단일 검색어인 '연설' 또는 '연설문'을 입력한 경우가 50회, '교육'을 포함한 복합어 입력이 37회, '새마을(운동)'을 포함한 복합어 입력이 33회, '대통령업적' 8회를 포함하여 '업적'의 복합어 입력이 28회 등으로 나타났다. 이러한 경향과 함께 1위인 검색어가 대통령연설임을 고려할 때, 대통령이 어떤 비전과 목표를 가지고 어떻게 정책을 세우고 업무를 수행하는지에 대해 많은 사람들이 가장 크게 관심을 가지고 있는 것을 알 수 있었다.

다음으로 '대통령' 또는 대통령님, 그리고 역대 대통령 이름, 이름 다음에 대통령을 입력한 경우가 많아(1,413회, 20.8%), 대통령 개인에 대한 관심이 높음을 알 수 있었다. 또한 '영부인' 6회 및 몇몇 영부인 이름들도 3회씩 입력되었고, '가족' 또는 '가족관계'라는 용어도 5번 입력되었으며, '혈액형'이라는 용어가 44회 입력된 것은 대통령 개인과 개인사에 대한 관심이 높음을 보여준 것이다.

기록관리와 관련된 용어도, 대통령기록, 국가기록원, 대통령기록관 등의 용어를 포함하여 많이 검색되어 중요한 주제 분야로 볼 수 있었다

(526회, 7.7%). 이들은 대통령 기록관에서 검색한 것을 고려할 때, 가장 일반적으로 입력될 수 있는 용어로 판단되었다.

대통령선물을 포함하여, 사진, 휘호, 싸인, 서명, 초상화 등도 많이 검색된 분야였다(304회, 4.5%). 이 범주는 형태나 매체를 중심으로 검색을 한 것으로 보고, 선물과 사진과 동영상 등을 같은 영역으로 묶어보았다. 이러한 선물 등 행정 박물과 매체에 대한 관심이 큰 것은, 이용자들이 기록의 매체 범주를 다양하게 보고 있음을 보여준다.

청와대, 참여정부, 인수위원회 및 각종 국가위원회, 정부 조직 및 조직도, 정부부서 이름 및 기관, 단체 등에 대한 검색도 꽤 많이 있어서 하나의 주제 범주로 묶을 수 있었다(270회, 4.0%). 다양한 위원회에 대한 관심도 많다는 사실은, 대통령 기록의 범주가 폭넓음을 알고 있는 이용자들이 대통령기록관을 상당히 많이 이용하고 있음을 대변해 준다.

그 외에 대통령선거 및 공약, 당선 또는 당선자 및 유세 등 선거 관련 용어들을 한 주제 분야로 묶을 수 있었다(258회, 3.8%). 앞서 정책 등의 범주에 넣었던, 많이 검색된 바 있는 '임기'라는 용어도 선거에 대한 관심과 연결해 볼 수 있을 것이다.

다양한 외교 관련 용어 및 미국, 한일관계담화 및 일본 등 한일 문제, G20, FTA 등 외교 현안들에 대해서도 많이 검색되었으며, 통일, 북한, 남북정상회담, 대북(정책) 등 남북관계 관련되어서도 많은 검색이 이루어졌다(228회, 3.4%).

취임식 등 취임을 포함한 용어들과, 국군의 날, 어버이날, 어린이날, 광복절 및 8.15 경축사 등 다양한 행사 관련 용어들도 많아 행사와 관

련된 대통령의 행적과 연설에도 관심이 많음을 알 수 있었다. 따라서 이렇게 행사관련 내용을 하나의 주제 분야로 묶어볼 수 있었다. 그런데, 광복절 경축사가 14회 검색된 것으로 미루어, 중요한 제안이 나온 연설이나 행사에 더 큰 관심을 보였음을 알 수 있으며, 역시 대통령에 대한 관심은, 연설 등을 통해 나타나는 대통령의 사안에 대한 견해나 정책방향 등에 대해 가장 높다는 것을 다시 한 번 확인할 수 있었다.

이렇게 주제 분야별로 검색어들을 묶어보면, 이용자들의 관심 영역에 대해 좀 더 잘 이해할 수 있어서, 콘텐츠 서비스나 전시, 교육 등 일반 서비스 영역의 주제를 정할 때나 홈페이지의 메뉴를 결정할 때에 도움을 받을 수 있을 것이며, 기관 내외부에 홍보를 할 때에도 유용하게 사용할 수 있을 것이다. 다음에서는 좀 더 구체적으로 검색어 분석 결과의 활용 방안을 검토해 본다.

5. 검색어 분석 결과 활용 방안

5.1 대통령 기록관 콘텐츠 서비스를 위한 적용 방안

앞에서 살펴본 것과 같이 대통령기록관 웹사이트에 접속한 이용자들은 다양한 검색어들을 입력하여 기록을 검색하였다. 가장 많이 검색된 용어인 '대통령연설'은 많이 검색된 용어군인 연두연설, 국경연설, 신년연설, 연설문, 취임사, 경축사, 담화문 등등과 함께 이용자들이 많은 관심을 가진 내용이었다. 앞의 <표 3>에서도 제시한 바와 같이 대통령연설에 대한 관심은

가장 검색이 많았던 주제 영역인 대통령의 정책, 업적 및 업무수행과 관련되어 있는 것으로 판단되며, 이에 따라 역대 대통령의 연설문 자체를 제시하는 것도 의미 있을 뿐 아니라, 연설문의 핵심 내용을 그 당시 사회의 주요 화두와 대통령들의 비전 및 핵심 정책과 연결해서 콘텐츠화하여 제시한다면 이용자들의 관심사에 좋은 대응이 될 것으로 판단된다. 또한 동시에 두 대통령의 이름이 같이 검색된 경우도 많아, 앞뒤로 임기가 연결되거나, 특히 관심이 높은 각 대통령들의 업적과 정책을 분야별로 비교하여 제시하는 내용도 많은 호응을 얻을 것으로 생각된다.

다음으로 개별 대통령 이름이 많이 검색되었으며, 영부인 및 가족과 개인사에 관련된 내용도 많이 검색되어, 개별 대통령 각각의 연보, 일상사, 일정 정보, 사진, 초상화 등과 함께 가족들과 개인정보 등에 관련된 콘텐츠를 다양한 내용으로 폭넓게 제공한다면 많은 이용자들의 관심을 끌 것임이 시사되었다.

다음으로 대통령선물도, G20 선물을 포함하여, 많은 이용자들이 검색한 것으로 나타났다. 대통령들이 임기 중에 외국국민들로부터 받은 선물에 대한 온라인·오프라인 전시도 진행하고 관련 정보도 제공하면 많은 이용자들에게 환영받을 것이다.

그리고 각 대통령 임기 당시의 정부의 조직도나 정부부서의 역할 및 청와대와 비서실의 추이 등도 관심의 대상이었음을 알 수 있었는데, 각 대통령 당시의 정부조직 및 청와대 조직과 책임자 및 역할 등의 정보가 일목요연하게 제시된다면 다양한 이용자들에게 도움이 될 것이다. 또한 취임 및 기념일 행사 등과 관련된

내용의 검색도 꽤 많이 있었음을 감안하여, 재직 시의 주요 행사에 대한 소개와 사진, 연설 및 관련 기록 등을 복합적으로 제시하면 흥미를 끌 것이다. 이와 더불어 대통령의 외교 관련 업적이나 방문국과의 관계, 특히 미국과 일본과의 외교 관계와 관련된 기록들, 그리고 남북 관계에서 중요한 정책이나 사진, 관련된 업무 결정사항과 관련 기록 등도 콘텐츠화하여 제시하면 호응이 클 것이다.

또한 대통령 임기 및 대통령 선거라는 용어도 많이 검색된 것으로 보아, 각 대통령의 임기 정보 및 선거과정에 대한 내용도, 공약, 유세 등을 포함하여 제공하면 좋을 것으로 판단된다. 그 외에 대통령기록관에서 소장하고 있는 기록들에 대한 전반적인 소개와 기록관리 과정도 대통령기록관에서 콘텐츠화하여 제공하면, 기록관 자체에 대한 소개로 뿐 아니라 기록관의 업무 영역에 대한 소개로도 의미 있을 것이다. 이러한 콘텐츠 서비스 외에 다양한 다른 기록 정보 서비스에도 검색어 통계와 주제 범주 분석이 다양하게 활용될 수 있을 것이다.

5.2 기록정보 서비스를 위한 방법론으로서의 검색어 분석

앞에서는 검색어를 분석하여 이용자들에게 더 다가설 수 있을만한 주제영역들을 살펴보고 콘텐츠화하면 좋을 내용들을 살펴보았다. 그 외에도 이용자들이 많이 찾는 관심주제는 다른 기록관에서도, 현재 대통령기록관 웹사이트의 검색창 아래에 제시하고 있는 추천키워드와 같은 방식으로 활용될 수 있을 것이며, 기록관의 아웃리치 서비스 영역인 전시, 교육, 출판 등에

도 효율적으로 활용될 수 있을 것이다.

또한 이용자들이 자주 이용하는 메뉴와 주제에 관련된 내용은 홈페이지 상에서도 전진 배치하여 눈에 가장 잘 띄는 영역에 쉽게 찾을 수 있게 하는 것도 중요할 것이며, 콘텐츠 서비스에 있어서도 이용자들이 더 관심있어 하는 영역의 정보를 더 보강하고 확충하여 심도 있게 제공하는 것이 필요할 것이다. 또한 용어 몇 개만 추천 검색어로 제시할 것이 아니라, 많이 검색되는 용어를 활용한 태그 클라우드를 제시할 수도 있겠고, 제시된 용어들의 관련어도 함께 제공하여 더 폭넓게 검색어를 선택하게 하는 것도 의미있을 것이다.

이용자들의 관심 주제 영역을 더 잘 파악하고 활용하기 위해서는, 검색어 분석결과를 활용하는 것도 필요하지만, 그 외에 이용자들이 자주 클릭하여 이동하는 관련 링크 정보를 제공하여, 다른 이용자들의 이동경로를 참고로 하여 해당 웹사이트의 다양한 다른 정보에 대한 안내를 하는 방안도 검토될 수 있을 것이다.

이렇게 검색어 분석은 실제 이용자들이 원하는 정보를 파악하도록 하여주며, 관심 있는 분야도 파악하도록 하여준다. 따라서 다양한 주제영역의 기록관에서는 본 연구에서 살펴본 것과 같이 검색어도 분석하고, 또 이용자들의 홈페이지 메뉴 활용과 관련된 각종 데이터를 분석하여 보는 것이 필요할 것이다. 이러한 결과를 적용한 서비스를 제공한다면, 홈페이지에 들어오는 이용자들의 만족도도 높아질 것이고, 이를 통해 이용자들이 늘어날 수도 있을 것이다.

6. 결론 및 제언

기록정보서비스는 몇 년 전까지도 기록관리 제도와 보존 영역에 비해 관심을 받지 못하던 주제 분야이다. 이제 그 경향은 바뀌고 있고, 이제 기록관들이 더 많은 의지를 갖고 노력을 경주해야 할 영역이다. 기록은 언젠가는 누군가에게 의미 있게 잘 활용되기 위해 제도와 규준을 만들어 관리하고 보존하고 있는 것이다. 그러나 먼 훗날을 위해서가 아니라 현재 시대를 살고 있는 기록관의 소속 기관 및 외부의 이용자들이 의미 있게 기록을 활용하도록 하는 것도 매우 중요하다. 현재 이용자들이 많이 활용하고 관심 있어 하는 내용을 이용자들이 좀 더 쉽고 편하게 이용하도록 하는 것은 기록관의 중요한 서비스이다.

본 연구에서 검색어들을 분석해 본 대통령기록관은 특별한 기록을 소장하고 있어서, 일반 이용자들의 관심을 충분히 끌 수 있고 반영할 수 있는 곳이다. 그런 의미에서 미리 이용자들이 관심을 갖고 검색을 많이 한 주제 영역에 대해 콘텐츠를 제작하는 등, 다양한 서비스에 활용하는 것은 매우 의미 있는 일이라고 할 수 있을 것이다.

그러나 이러한 서비스는 대통령기록관 같은 특수한 곳에서만 진행되어야 할 일은 아니다. 대부분의 기록관은 모기관에 소속되어 있다. 모기관은 기관 나름대로의 목표와 비전과 업무에 있어 특정한 영역을 다루고 있다. 기록관에서 이러한 모기관의 비전 및 목표와 관련된 주제들과 업무수행자나 이용자들이 관심을 가지고 있는 주제에 대해 미리 파악을 하고, 소장 기록과 내외부의 관련 정보 등을 복합적으로

활용하여 콘텐츠화하여 기관의 홈페이지 등에 띄워 놓으면, 기록관의 정보소스로서의 역할에 대해서도 주목을 받을 수 있을 것이다. 이러한 서비스는 실제 업무 수행에 있어서도, 과거의 유사 사안에 대한 결정 과정에 제시되었던 의견들을 참조할 수 있는 등, 미처 업무수행자들이 파악하지 못했을 수도 있는 정보를 제공하는 기회가 될 수도 있을 것이다. 또한 주요 현안이 있는 경우, 일반 이용자들도 그 기관 홈페이지에 들어와서 관련 정보를 찾아보고자 할 수도 있을 것이며, 그런 경우 미리 관련 정보를 콘텐츠화하여 홈페이지에 올린다면 이용자들의 요구에 미리 대처하는 적절한 서비스가 될 수 있을 것이다.

이러한 기관 내·외부 이용자들의 요구를 파악하는 다양한 방법 중에 검색어 분석을 활용하는 것도 매우 의미가 있음을 본 논문에서는 제시하고자 하였다. 또한 기록관에서 검색어 분석을 활용하기 위해서는 어떠한 방법으로 이를 정련하고 활용하는 것이 좋을지도 간략히 제시하고자 하였다. 정련된 검색어 분석을 통해, 이용자들을 위한 추천 검색어 제공, 이를 반영한 홈페이지 메뉴 설계, 그리고 체계적이고 복합적인 디지털 기술을 활용한 콘텐츠 제공 등은 이용자들이 기록관에서 기록들을 더 쉽고 더 흥미 있고 체계적으로 파악하여 활용하도록 하여 줄 것이다.

그러나 검색어 분석만으로는 이용자의 요구를 파악하는데 완전하지 않기 때문에, 이용자들이 대한 인터뷰를 진행하여 이를 보강하는 방법도 병행한다면 더욱 수준 높은 서비스가 제공될 수 있을 것이다. 또한 검색어 분석 결과를 콘텐츠 서비스나 기타 관련 서비스에 반영하기 위해

적절한 주제와 내용을 선정하는 과정에 이용자와 전문가들이 참여한다면 더 의미 있는 서비스가 제공될 수 있을 것이다. 향후에 이러한 내용들이 심도 있게 연구되기 바란다.

앞으로 지역기록관리기관이나 혹은 특정 주제를 다루고 있는 국가기관, 단체, 기업 등에서

도 기존의 검색어 데이터를 활용하고 이용자의 요구를 반영하여 기록정보서비스에 적극 활용하는 것이 필요할 것이다. 많은 기록관들의 적극적인 마인드를 기반으로 한 진일보된 서비스들을 기대해본다.

참 고 문 헌

- 강경희. 2008. 대통령기록물의 분류에 관한 연구. 한남대학교 석사학위논문.
- 곽건홍. 2001. 대통령기록 관리기구의 기능과 역할. 『기록학연구』, 4: 3-30.
- 구분윤. 2009. 대통령기록관의 수집정책에 관한 연구. 한국외국어대학교 석사학위논문.
- 김성수, 서혜란. 2002. 대통령기록관의 설립 및 정부기록보존소의 위상에 관한 연구. 『한국기록관리학회지』, 2(1): 41-66.
- 김연희. 2010. 대통령기록관 웹사이트 사용성 평가 연구. 한남대학교 석사학위논문.
- 김관석, 김관보, 권영주. 2003. 대통령기록물 보존·관리제도 실태분석과 선진화 방안. 『한국행정학보』, 37(4): 303-328.
- 김향란. 2008. 대통령기록물의 관리 방안에 관한 연구. 명지대학교 석사학위논문.
- 남건우. 2008. 웹기록 이관의 문제점과 개선방안에 대한 연구: 대통령 기록관의 이관사례를 중심으로. 한남대학교 석사학위논문.
- 남태우, 오지영, 유보현. 2007. 대통령기록물관리법에 관한 연구. 『한국기록관리학회지』, 7(2): 165-188.
- 박성경. 2008. 대통령 행정박물의 유형별 기술규칙에 관한 연구. 명지대학교 석사학위논문.
- 박소연. 2010. 국내 포털 이용자들의 멀티미디어 검색 행태 분석. 『한국문헌정보학회지』, 44(1): 101-115.
- 박소연, 이준호, 김지승. 2005. 클릭 로그에 근거한 네이버 검색 질의의 형태 및 주제 분석. 『한국문헌정보학회지』, 39(1): 265-278.
- 박소연, 이준호. 2002. 로그 분석을 통한 이용자의 웹 문서 검색 행태에 관한 연구 『정보관리학회지』, 19(3): 111-122.
- 박소연, 이준호. 2005. 국내 웹 이용자의 검색 행태 추이 분석. 『한국문헌정보학회지』, 39(2): 147-160.
- 박소연, 이준호. 2007. 웹 검색 분야에서의 로그 분석 방법론의 활용도. 『한국문헌정보학회지』, 41(1): 231-242.
- 설문원, 김익한. 2006. 이승만시기 국무회의록과 정부부처 기록의 연관구조 분석에 기반한 역사 콘텐츠 설계 방안. 『한국비블리학회지』, 17(2): 115-136.
- 손성호. 2008. 민간소장 대통령기록물의 수집 방

- 안. 한국외국어대학교 석사학위논문.
- 심성보. 2007. 교수·학습자료용 기록정보 콘텐츠 서비스의 구성 및 개발. 『기록학연구』, 16: 201-256.
- 이상민. 2001. 대통령기록관의 설립과 운영 방향: 미국 대통령기록관의 사례와 교훈. 『한국기록관리학회지』, 1(2): 31-56.
- 이선정. 2007. 대통령기록관의 운영 연구: 미국 대통령기록관의 홍보기능 분석을 중심으로. 한남대학교 석사학위논문.
- 이승휘. 2002. 대통령기록물의 보존과 그 과제. 『기록학연구』, 6: 279-292.
- 전수진. 2008. 보존기록관리기관의 온라인 콘텐츠 유형과 특징. 『한국기록관리학회지』, 8(2): 127-152.
- 전용호. 2008. 대통령기록관의 기록정보 서비스 체계 연구: 한국의 대통령기록관을 중심으로. 명지대학교 석사학위논문.
- 정상화. 2008. 대통령 기록물의 수집 전략: 이승만 대통령 기록물의 사례를 중심으로. 충남대학교 석사학위논문.
- 정진화. 2010. 국내 기록물관리기관의 대통령관련기록물 수집에 관한 연구. 연세대학교 석사학위논문.
- 조민지. 2006. 대통령 기록관의 서비스 프로그램 사례 연구. 『한국기록관리학회지』, 6(2): 157-184.
- 조영삼. 2009. 대통령기록관리의 현황과 전망. 『기록학연구』, 21: 283-322.
- 주희경. 1999. 대통령기록물의 관리와 제도 확립. 부산대학교 석사학위논문.
- 최동진, 황명권, 김관구. 2010. 웹 로그 데이터를 이용한 사용자 관심분야 분석 모델 설계. 『한국정보기술학회 하계종합학술발표논문집』, 333-336.
- 최성렬. 2009. 대통령지정기록물 관리 방안에 관한 연구. 한국외국어대학교 석사학위논문.
- 최재혁, 김희섭. 2009. '역대 대통령 웹기록 서비스'의 이용자 만족도 분석. 『한국기록관리학회지』, 9(1): 77-98.
- 한국국가기록연구원. 2007. 『민주화운동기념사업회 사료정보콘텐츠 구축 방안』. 민주화운동기념사업회.
- 홍원기. 2004. 대통령기록물의 수집·이관 제도에 관한 연구. 한국외국어대학교 석사학위논문.
- Chung, E. K. & J. W. Yoon. 2010. Examining Categorical Transition and Query Reformulation Patterns in Image Search Process. 『정보관리학회지』, 27(2): 37-60.
- Conway, P. 1986. Research in Presidential Libraries: A User Survey. *The Midwestern Archivist*, 11(1): 35-56.
- Drake, M. A. 2007. Presidential Archives: Hype, Reality, and Limits to Access. *Information Today*, 24(6): 1-3.
- Du, J. 2010. Multitasking, Cognitive Coordination and Cognitive Shifts during Web Searching. Thesis. Queensland University of Technology. [cited 2011. 1. 22]. <http://eprints.qut.edu.au/35717/1/Jia_Du_Thesis.pdf>.
- Jansen, B. J. 2009. Understanding User: Web Interactions via Web Analytics. (Synthesis Lectures on Information Con-

- cepts, Retrieval, and Services) Plymouth: Morgan and Claypool Publishers.
- Jansen, B. J. and A. Spink. 2005. How are We Searching the World Wide Web? An Analysis of Nine Search Engine Transaction Logs. *Information Processing & Management*, 42(1): 248-263.
- Jansen, B. J. and Spink, A. 2005. An Analysis of Web Searching by European Alltheweb.com Users. *Information Processing and Management*, 41(2): 361-381.
- Jansen, B. J., A. Spink, and 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.
- Jansen, B. J., A. Spink, and J. Pedersen. 2005. A Temporal Comparison of Alta Vista Web Searching. *Journal of the American Society for information Science and Technology*, 56(6): 559-570.
- Park, S., J. H. Lee and H. J. Bae, 2005. End User Searching: A Web Log Analysis of NAVER, a Korean Web Search Engine. *Library & Information Science Research*, 27(2): 203-221.
- Rieh, S. Y. & H. Xie. 2006. Analysis of Multiple Query Reformulations on the Web: The Interactive Information Retrieval Context. *Information Processing and Management*, 42: 751-768.
- Ross, N. C. M., and D. Wolfram. 2000. End User Searching on the Internet. An Analysis of Term Pair Topics Submitted to the Excite Search Engine. *Journal of the American Society for Information Science and Technology*, 51(10): 949-958.
- Silverstein, C., M. Henzinger, H. Marais, and M. Moricz, 1999. Analysis of a Very Large Web Search Engine Query Log. *SIGIR Forum*, 33(1): 6-12.
- Spink, A., and B. J. Jansen. 2004. A study of Web Search Trends. *Webology*, 1(2), Article 4. [cited 2011. 2. 20]. <<http://www.webology.ir/2004/v1n2/a4.html>>.
- Spink, A., D. Wolfram, B. J. Jansen, and T. Saracevic. 2001. Searching the Web: The Public and their Queries. *Journal of the American Society for Information Science and Technology*, 52(3): 226-234.
- Swanson, D. R. 1977. Information Retrieval as a Trial-and-Error Process. *Library Quarterly*, 47(2): 128-148.
- Wang, P., M. W. Berry, and Y. Yang. 2003. Mining Longitudinal Web Queries: Trends and Patterns. *Journal of the American Society for Information Science and Technology*, 54(8): 743-758.

[관련 법령]

대통령기록물 관리에 관한 법률

[관련 웹사이트]

대통령기록관 웹사이트, 대통령기록관소개 비
전/미션.

[〈http://www.pa.go.kr/PMS/vision/vision.html〉](http://www.pa.go.kr/PMS/vision/vision.html).

U.S. National Archives and Records Administration, Presidential Libraries.

[〈http://www.archives.gov/presidential-libraries/〉](http://www.archives.gov/presidential-libraries/).