

공공기록에 대한 접근성 제고 방안에 관한 연구*

- 대통령기록관 주제시소러스 개발 사례를 중심으로 -

A Study on the Improvement of Accessibility to Public Records: Based on the Construction of Subject Thesaurus for Presidential Archives

이 해 영 (Hae-young Rieh)**

권 용 찬 (Yongchan Kwon)***

성 효 주 (Hyojoo Seong)****

유 병 후 (Byonghoo Yoo)*****

목 차

- | | |
|-----------------|-------------------------|
| 1. 서 론 | 4. 대통령기록물 주제 시소러스 구축 절차 |
| 2. 대통령기록물과 시소러스 | 5. 시소러스 활용 방안 및 제언 |
| 3. 현황 및 사례 연구 | |

<초 록>

기능별 분류체계나 출처 중심의 탐색은 이용자들에게 쉽지 않으며 키워드검색도 이용자가 입력한 검색어와 기록물명의 단순 매칭 결과만 제시하여, 이용자의 기대에 부응하지 못하고 있다. 대통령기록관에서는 주제 분류체계를 개발하여 기록검색의 효율성을 높이고자 하였으며, 이와 연계하여 이용자들의 검색 과정을 돕고 기록관리 업무에도 용어통제 등 도움을 줄 주제 시소러스를 개발하고자 하였다. 본 연구에서는 대통령기록물 건명을 형태소 분석하고, 관련 시소러스와 이용자가 웹사이트에 입력한 검색 용어 및 정보공개 청구용어 등의 현황을 면밀히 분석한 후, 주제 분류체계와의 매핑을 통해 주제시소러스를 개발한 과정을 제시한다. 또한 주제 시소러스의 업무 및 온라인 서비스에의 활용 방안을 제시하며, 앞으로의 발전방향을 제시한다.

주제어: 주제시소러스 개발, 대통령기록관, 이용자 검색용어, 정보공개 청구용어, 온라인 서비스

<ABSTRACT>

To search based on the functional classification or provenance is not easy for users, and the key word-based information retrieval presents only simple words matching with the title of the records. The Presidential Archive of Korea developed a subject classification scheme to improve the convenience of searching for various records and came up with a subject thesaurus based on the scheme that utilizes the terms appearing on the title of the records and the terms used by the users who searched the portal or requested information disclosure. This research presents the development process of subject thesaurus. It also presents the utilization methods for records management work and services.

Keywords: subject thesaurus, presidential archives, users' search terms, terms for information disclosure request, online services

* 본 논문은 2014년 대통령기록관 연구용역사업인 <대통령기록물 주제 시소러스 구축>의 내용 일부를 정리한 것임.

** 명지대학교 기록정보과학전문대학원 교수(hyrie@mj.ac.kr)

*** 명지대학교 기록정보과학전문대학원 박사과정(radiyong7@gmail.com)

**** 명지대학교 기록정보과학전문대학원 석사과정(wh4hj@hanmail.net)

***** 명지대학교 기록정보과학전문대학원 졸업(yubumoo@hotmail.com)

■ 접수일: 2014년 10월 24일 ■ 최초심사일: 2014년 11월 9일 ■ 게재확정일: 2014년 11월 19일

■ 한국기록관리학회지 14(4), 127-151, 2014. <<http://dx.doi.org/10.14404/JKSARM.2014.14.4.127>>

1. 서론

1.1 연구 배경과 목적

기록의 기능별 분류체계나 출처를 중심으로 하는 탐색은 이용자들에게 쉽지 않은 일이다. 이용자들은 검색엔진 등의 기능에 익숙해져서, 기록에 대해 원문검색을 원한다든가 혹은 주제어 등에 대한 검색을 원하지만 기록에 대해서는 이러한 검색이 쉽지 않은 상황이다. 사실 기록관 웹사이트에서 제시되는 키워드 검색도 이용자가 입력한 검색어와 기록물명의 단순 매칭 결과만 제시하여, 주제어 검색이나 이용자의 의도를 반영한 결과를 원하는 이용자들의 기대에 부응하지 못하고 있다. 기록정보서비스의 중요성이 부각되어감에 따라 기록관들이 제공하는 기록서비스도 이용자의 편이성에 대한 고려를 확대해가고 있는 추세이고, 기록에 대한 검색의 편이성도 이에 부응해야 할 것이다.

이러한 추세에 맞추어 대통령기록관은 기록에 주제어를 부여하고 이용자들이 주제어 검색을 할 수 있도록 하고자 주제 분류체계를 개발하는 프로젝트를 진행하였다(정광훈, 남영준, 2012; 이해영, 2013). 대통령기록물은 행정적, 정보적 가치뿐만 아니라 그 역사적인 가치가 높아 당대의 정치와 사회를 연구하는데 다양하게 활용될 수 있다. 그러므로 대통령기록물의 주제와 그 맥락을 파악하고 적극적인 활용을 돕는 도구가 필수적이다. 특히 「대통령기록물 관리에 관한 법률」의 시행에 따라 대통령기록물의 생산과 이관이 필수가 됨에 따라 대통령기록물의 양은 크게 증가하였으며, 이들 기록을 이용자들이 쉽게 검색하고, 맥락정보를 파악하도록 하기 위해

서는 검색도구의 역할이 매우 중요하다. 이렇게도 현재 운용 중인 대통령기록관의 기록물 검색도구는 기록정보의 검색 및 맥락정보 파악에는 다소 부족한 형편이다. 현재 대부분의 기록물이 전자적인 형태로 생산됨에 따라 이에 맞게 대통령 기록물을 정밀하게 검색하고 활용 할 수 있도록 하는 다양한 검색도구 및 보조도구의 구축이 절실하다.

대통령기록관 웹사이트에서 이용자들이 입력한 검색 용어를 분석한 결과에 따르면, 이용자들은 개별 대통령의 업적이나 이슈에 관심이 많다(이해영, 2011; 이해영, 2013). 그러나 구체적이고 정확한 검색어를 적절하게 활용하지 못하고 추상적인 단어 입력을 통한 검색 수준에 머물고 있어, 이용자들에게 기록물 검색에 적합한 용어를 쉽게 찾을 수 있도록 하는 적극적인 서비스의 필요성이 확인되었다. 이렇게 이용자들이 원하는 주제어를 쉽게 찾아 검색할 수 있도록 하는 도구가 시소러스이다.

또한 기록물의 등록, 분류, 정리 및 기술 업무와 관련하여, 담당자가 입력하는 내용에 일관된 용어를 사용하도록 통제하거나 참조할 수 있는 도구가 없는 상황에서 적절한 용어사전의 필요성이 지속적으로 제기되어 왔다. 대통령기록물을 기반으로 한 시소러스는, 이러한 기록물의 등록, 정리, 분류, 기술, 평가, 공개 재분류 및 열람 등 기록물관리 전 과정에 있어 유기적으로 활용할 수 있는 용어 통제 사전이 될 수 있으며, 따라서 사용하는 용어의 표준성과 통일성 그리고 계층성을 확보하는 중요한 도구가 될 수 있다. 궁극적으로 대통령기록관 시소러스는 대통령기록물 검색이나 정보공개 청구 시에 이용자들이 많이 사용하는 용어를 주제 분류체계를

기반으로 하여 작성됨으로써, 이용자와 대통령 기록물의 출처별 정리체계 및 낯선 용어들과의 괴리를 최소화할 수 있는 도구가 된다.

본 연구에서는 기록물 건명과 이용자 검색 용어 및 정보공개청구 용어 분석을 통해, 현재와 같은 출처나 기능 중심의 검색도구를 보완할 수 있는 구체적이고 이용자 친화적인 주제별 브라우징 및 주제어 검색을 지원하는 데에 도움을 줄 수 있는 주제 시소러스를 개발하고자 하였다. 또한 용어의 표준성, 통일성, 계층성을 확보하고, 대통령기록물의 관리 업무 프로세스 전반에 유기적으로 활용할 수 있는 용어 통제 사전, 즉 시소러스를 구축하는 데에 본 연구의 목적이 있다. 본 연구는 또한 이러한 시소러스를 기록물관리 업무를 진행하는데 있어 도움을 줄 수 있는 업무측면의 활용방안과 함께, 외부 이용자들의 편의를 증진시킬 수 있는 온라인 서비스 제공방안도 제시하고자 하며, 앞으로 서비스를 고도화할 수 있는 발전 방안도 함께 제시하고자 한다.

1.2 선행연구

기록물 검색을 더 이용자 친화적으로 제공할 수 있도록 하는 다양한 검색도구의 필요성이 대두되고 있으며, 이를 위한 하나의 방편으로 시소러스에 대한 관심이 증가되고 있다. 시소러스에 대한 연구는 전반적으로 많이 진행되어 왔으나, 기록학 분야에서는 시소러스 관련 연구가 많지 않은 편이다.

먼저 국내의 경우, 박부숙(2007)은 호주의 기능시소러스인 Keyword AAA, AGIFT, 미국 NARA의 시소러스, UNESCO 시소러스 등의 해외 사례를 소개하고, 국가기록원의 기록

물 관리·검색용 시소러스 구축 사업 현황 및 개발 계획을 소개하였다. 이해영, 이미영, 이은영, 이혁준, 이현정, 최영실, 박미자(2008)는 대학기록관 기록물에 나타난 용어를 대상으로 시소러스 구축과정을 소개하고, 이를 반영한 지침을 제시하였다. 용어의 선정 방법과 용어들 간의 상호관계 설정과 관련하여 구체적인 사례를 제시하고, 구축과정에서 도출된 이슈들과 그 대응 방안에 대해 소개하였다. 최윤경, 정연경(2014)은 한국 근대 여성 구술기록물의 효율적인 주제 접근을 위한 시소러스를 개발하였다. 이들은 국내외 시소러스 개발 사례와 시소러스에 대한 국제표준인 ISO 25964와 미국표준 ANSI/NISO Z39.19를 참고하여, 구술기록물에 입력된 색인어를 중심으로 일반 주제어를 비롯한 지명, 장소명, 인명, 단체명 등 고유명사가 포함된 시소러스의 구축 과정을 5단계로 제시하였다.

류주희(2005)는 한국 역사용어 사용과 관련하여 이용자들의 검색 재현율과 정확률을 높이기 위한 방안으로 시소러스 개발의 필요성을 언급하고, 국제표준에 근거한 시소러스 구축원칙을 통해 시소러스를 실제 구축하는 과정을 설명하고 있다. 박옥남(2011)은 시소러스 제작과 구축을 위한 지침이 되는 국제 표준 ISO 25964 (Thesauri and interoperability with other vocabularies)에 대해 소개하고, 기존의 시소러스 표준인 ISO 2788: 1986 및 ANSI/NISO Z39.19-2005와의 비교분석을 통해 시소러스의 실질적인 구축 과정을 제시하였다.

외국의 경우, 1990년대 이전부터 국가기록관을 중심으로 하여 일부 기록관에서 시소러스 구축이 진행되어왔다. McNitt(1989)는 PRESNET

(Presidential Libraries Information Network, 대통령 도서관 정보 네트워크)의 주제어 색인 시소러스의 필요성 및 구축 과정을 소개하고 있으며, 시소러스 개발 후의 개선과정, 포드(Gerald R. Ford) 대통령 도서관에서의 적용사례를 통한 이슈를 제시하였다. McNett은 좁고, 전문화된 시소러스는 기록의 특정 영역을 색인하는 도구로서의 역할과 더 넓은 시소러스를 개발하기 위한 잠재적인 기반으로서 역할을 동시에 하고 있다고 설명하고 있다.

Garrod(2000)는 UNESCO 시소러스 두 번째 버전을 기반으로 한 영국의 NDAD(UK National Digital Archive of Datasets, 국가 디지털 데이터셋 아카이브)의 시소러스 개발 과정을 소개하였다. 논문에서는 새로운 시소러스를 개발하기 보다는 계층을 가진 기존의 시소러스를 활용하는 것이 상당히 이점이 있음을 밝히고 있다.

Giménez-Chornet & Escrig-Giménez(2011)는 스페인에서 가장 중요한 역사 아카이브 중의 하나인 Archivo del Reino de Valencia의 역사적 문서를 기반으로 한 시소러스의 구축 과정을 제시하고 있다. 시소러스 구축의 목적은 전문가와 일반 시민 모두를 위해 역사적 문서에 포함된 정보에 대한 가시성을 높이는 데에 두고 있음을 밝히고 있다.

한편, 대통령기록물의 주제 분류체계 개발과 관련하여, 정광훈, 남영준(2013)은 대통령기록물의 사료적 성격과 관리의 필요성을 확인하고, 이를 바탕으로 한 대통령기록물의 관리적 차원의 분류체계를 실제 설계하는 과정을 소개하였다. 국내외 주요지표를 살펴본 결과 정부기능을 기반으로 각각의 기관 특성에 따라 주제어를 설정하여 분류체계를 구성하고, 업무를 세분화하

지 않고 주제개념영역을 특정하는 방식으로 설계하였다. 이해영(2013)은 대통령기록관 주제 분류체계 고도화 사업을 통해 대분류, 중분류, 소분류의 3단계 기존 분류체계를 전반적으로 차용하되, 기록물 건명과 이용자들의 검색 용어 및 정보공개 청구용어를 분석하고 반영하여, 주제 사안에 따라 기록물이 세분되는 경우에는 4단계 세분류를 설정하는 등 세분화되고 구체화된 주제 분류체계를 설계하였다. 이 연구를 통해 기록물 한 건당 3개 이상의 주제어를 부여하는 다중분류체계 방식을 제안하고 실제 적용하였다.

한편 시소러스를 더 효율적으로 활용하고자 하는 다양한 시도를 하는 연구들도 많이 진행되었다. 손대형, 김태수(1998)는 분석합성형 분류체계로부터 개념과 관계를 추출하여 작성한 자모순 시소러스에 패싯(facet)을 결합하여, 색인과 검색의 특정성을 높이고 용어 및 관계의 일관성도 높임으로써 기존의 분류체계나 시소러스가 지니고 있는 한계를 개선하고자 하는 방안을 제시하였다. 황순희, 정한민, 성원경(2006)은 과학기술분야 시소러스 구축에 다양한 개념 패싯(facet)과 범주관계 패싯 및 의미역 관계 패싯을 이용하는 방안을 제시하고, 이가 검색의 효율성을 보장할 수 있다고 하였다. 박지영(2010)은 문학류 문헌의 주제를 효과적으로 표현하기 위해 한국십진분류법(KDC)을 패싯분류표로 재구조화시키고 국립중앙도서관의 주제명표목표의 디스크립터와 연계해 보았는데, 이를 통해 KDC의 주제표현력을 높이고 상세 분류가 가능하다고 하였다. 이승민(2010)은 물리학 분야의 시소러스인 PACS와 PIRA를 선정하여, 상이한 시소러스를 직접 통합하는 것

이 아니라 시소러스 간에 존재하는 관련된 개념들을 상호 연결시켜줄 수 있는 방안으로 패킷 프레임워크를 제시하였다.

이와 함께 시소러스의 한계를 개선하기 위한 방법으로 온톨로지와의 연계를 제안한 연구들도 있는데, 이정희와 김희섭(2007)은 국내 대학의 행정기록물 검색에 키워드 기반검색과 자체적으로 개발한 온톨로지 기반 검색시스템의 성능 평가를 실시하였고, 이혜원, 윤소영(2011)은 한국역사용어 시소러스에 적용할 역사용어 온톨로지 모형을 설계하였다. 이들은 앞으로 기록물 검색 시스템이 지향해야 할 방향을 제시한다.

1.3 연구의 범위와 방법

본 연구에서는 대통령기록관의 주제 분류체계에 기반한 주제 시소러스를 구축하고자 하였으며, 이와 관련하여 다음과 같은 원칙을 정하였다. 먼저 기존 대통령기록물 주제 분류체계의 큰 틀은 가능하면 유지하고, 주제 분류체계에 사용된 용어를 대표어로 채택하고자 하였다. 또한 기록물 건명에서 용어들을 선정하되, 이용자들이 많이 찾는 용어를 최대한 반영하고, 기록물 전반에 사용되는 용어를 대표어로 선정하여 앞으로 생산단계에서도 활용 가능하도록 설계하고자 하였다. 또 직관적으로 이해 가능한 구체적인 용어를 반영하고자 하였다. 전반적으로 국가기록원 시소러스 지침(2009)에 준거하여 대통령기록물의 특성에 맞게 구축 방법을 재설정 하였다.

실질적인 구축 방법을 정립하기 위해 먼저 유관기관의 시소러스 구축 및 활용 현황을 파악하고자 하였다. 백악관 기록관리실(WHORM,

White House Office of Records Management) 주제파일, 국가기록원 일반주제 시소러스, 국사편찬위원회 한국 역사용어 시소러스, 신문기사 종합 시소러스, UNESCO 시소러스 및 UK Archival Thesaurus(UKAT) 등을 분석하여 각각의 특성과 장단점을 살펴보았으며, 일부는 관련 용어 및 비대표어 결정 시에 참고하였다. 특히 서비스 제공 방법도 살펴보고 이용자 편의성을 분석하여, 단점은 개선하고 장점은 반영하고자 하였다.

대통령기록물 건명으로부터 시소러스에 사용할 용어들을 도출하기 위해 1-15대 대통령기록물 25만 건과 16대 대통령기록물 120만 건의 기록물 건명에 대한 형태소 분석을 시행하였다. 이 중 기록물에서 30번 이상 등장한 용어들을 하나하나 기존 주제 분류체계에 매핑하기로 하였다. 대통령기록물 주제 분류체계 대분류 24개 분야 중 '행정지원' 분야는 추후 기능시소러스에서 반영하도록 하고, 나머지 23개 분야를 대상으로 하였다. 23개 분야 중 특히 정치, 경제, 통일, 외교, 국방 분야 및 대통령 고유 업무 활동 분야, 개별 대통령별 업적과 주요 사안 등은 중점분야로 더 세부적으로 확장하고자 하였다. 그러나 대통령 시기별 주요 정책 및 사건, 주요행사 및 회의명, 인명, 국가명, 지명, 기관 및 단체명, 직책명(수석비서관 등), 법률명, 조약명 등의 용어는 확인하여 각 주제 분류체계 영역에 매핑하였으나, 매우 중요하다고 판단된 일부를 제외하고는 시소러스에 용어로 편입하지 않고 따로 정리하였다. 이는 앞으로 대통령기록관에서 패킷 분류체계를 개발할 때 주요한 옵션으로 제시하기로 하였다.

또한 이용자들이 사용하는 용어들을 파악하

고 반영하기 위해, 대통령기록관에서 2008년 6월 개설하여 제공하고 있는 대통령기록관 웹사이트에서 이용자들이 어떠한 용어를 입력하여 기록물을 검색하는지 분석하였다. 2010년 1월부터 11월 및 2011년 1월부터 2013년 5월까지 입력된 용어 약 660만 건 중 3번 이상 입력된 의미 있는 용어를 취합하여 분석하였다. 더불어 대통령기록관에 2010년부터 2014년 6월까지 요청된 정보공개 청구 목록 582건을 분석하여, 기록물 요청에 이용자들이 사용하는 용어들을 반영하고자 하였다.

이렇게 분석된 용어들을 각각 기준에 개발된 주제 분류체계에 중분류 및 소분류 단위에 매핑을 하여 입력하였다. 포괄적 용어는 대분류 단위에 매핑하여 입력하였다. 또한 백악관 내의 기록관리 담당 부서인 백악관기록관리실(WHORM) 및 미국 대통령기록관에서 활용하는 주제파일(Subject File) 체계에 제시된 용어들도 모두 매핑해서 입력하여 용어 선정에 참고하였다.

이렇게 입력된 용어는 본 연구의 저자들과 대통령기록관 연구사가 함께 모여서 용어 하나하나에 대해 대표어 및 비대표어 선택 여부와 관련어 선정 여부에 대해 논의를 계속하였다. 이 과정에서 국회도서관 시소러스, 신문기사 시소러스, 국가기록원 시소러스 등은 비대표어 및 관련어를 선정하는 데에 계속 참고가 되었다. 또한 시소러스의 업무 및 온라인 활용방안을 설정함에 있어, 자문위원들의 의견들을 수렴하고 반영하였으며, 온라인 검색도구 개선 방안을 위한 기록관 이용자들의 포커스그룹 면담에서 개진된 의견도 반영하였다(이해영, 2013; 이윤령, 이해영, 2014).

그러나 본 연구 결과로 구축된 시소러스는 일

상적인 용어 관계를 제시하는 일반적인 주제어 시소러스라기보다는 대통령기록관의 대통령기록물을 대상으로 하는 주제 분류체계의 용어집 역할을 하는 용어 통제의 도구로 구축하였다는 한계가 있다. 따라서 이 시소러스는 주제 분류 체계에 속하는 주제어를 대표어로 하고, 그에 대한 비대표어와 상·하위어 및 관련어를 보여주는 성격의 시소러스라고 할 수 있으며, 이에 따라 일상적인 시소러스들에 제시되는 용어들의 관계와는 차이가 있을 수 있다. 즉 한 용어가 다양한 상·하위 관계어를 갖는 것이 아니라, 하나의 용어는 주제 분류체계의 한 계층에 할당됨에 따라 한 개의 상위어만 갖게 되는 한계점을 갖게 된다. 이 한계는 다양한 관련어들을 제시함으로써 극복하고자 노력하였다.

2. 대통령기록물과 시소러스

2.1 대통령기록물의 특성 및 현황

대통령기록물은 대통령 직무수행 등과 관련하여 대통령과 대통령의 보좌기관·자문기관 및 경호업무를 수행하는 기관, 대통령인수위원회 등이 생산, 접수하여 보유하고 있는 기록물 및 대통령선물, 대통령상징물 등 물품을 말한다(대통령기록물 관리에 관한 법률). 우리나라에서 대통령은 행정수반으로서 행정에 관한 최고의 지위를 가지며, 정부조직, 국무회의 장의 지위를 가진다. 즉 대통령은 초법적인 지위를 가지고, 행정부, 입법부, 사법부를 통제하는 권력을 부여받고 있다. 이런 대통령 기록물은 생산 당시에는 국정운영에 관련된 공공 기록물로

서의 역할을 하지만, 대통령 퇴임직후에는 차기정권의 국정운영에 참고자료로 활용되기도 하며, 후대에 당시대를 평가할 수 있는 귀중한 사료로 활용될 수 있는 기록물이다. 이렇듯 대통령기록물은 공공기록으로서 다양한 가치를 가지고 있기 때문에, 대통령기록물 관리에 관련된 사항을 「대통령기록물 관리에 관한 법률」에 정하여 다른 일반 기록물과는 달리 특별히 관리하고 있다. 대통령기록물은 당대 대통령의 국정운영의 투명성과 함께 통치업적을 확인할 수 있는 중요한 바로미터의 성격을 가지는 이유로, 그 체계적인 관리제도는 중요하다. 대통령기록관이 소장하고 있는 기록물과 각 대통령도서관, 기념관 등의 외부기관에 소속된 기록물은 각각의 분류체계와 관리기준에 의거 개별적으로 관리되고 있다. 「대통령기록물 관리에 관한 법률」 시행 이전의 대통령기록물은 이관 및 수집이 병행되어 구성되어졌다(정광훈, 남영준, 2013).

대통령기록관에는 2008년 2월, 「대통령기록물 관리에 관한 법률」 시행 후 처음으로 대량의 대통령기록물이 이관되었고, 이어서 국가기록원이 관리해 오던 역대 ‘대통령관련 기록물’ 또한 대통령기록관으로 관리전환 되었다. 이관 기록에는 대통령기록물 생산기관인 대통령비서실 등 34개 기관에서 종이문서류, 간행물류, 시청각기록물 등 기록물과 e지원시스템, 온나라 시스템, 전자문서시스템 등의 표준전자문서시스템에서 생산한 전자기록 및 각종 행정정보시스템, 각 기관 홈페이지 등의 웹기록 및 대통령선물과 행정박물 등이 포함되었다. 국가기록원이 소장하고 있던 대통령기록물을 관리전환받은 기록은 현행 법률의 “대통령기록물”과 범주

를 달리하는 중앙행정기관 등의 재가문서인 “대통령관련 기록물”로서 국가기록원이 구축한 “통치사료시스템”에 등록된 것과, 아직 등록 작업을 대기 중이던 범주로 구분된다(대통령기록관, 2013).

2013년 3월 현재 대통령기록관 소장 기록물은 수집이나 기증기록물을 제외하고, 정리 작업이 완료된 기록물이 총 842만 8천여 건이며 이중 노무현(고건)대통령기록물이 755만여 건, 김대중대통령 기록물이 70만여 건이다. 대통령기록물의 대다수는 대통령의 임기종료에 맞춰 5년 주기로 이관되는 대통령기록물 생산기관의 기록물이다. 이 기록물들은 “대통령별 > 기관별 (RG) > 생산부서별 (Series) > 업무기능·활동 (File) > 기록물철 > 기록물건”순으로 정리된다. 기록물들의 정리 작업은 ① 기록물철 확정 및 목록 입력, ② 분류체계 설계, ③ 1차 정리(분류체계별 기록물 배치-상자편성, 분류체계에 속하는 기록물철 목록정보 매핑), ④ 2차 정리(기록물 세부건 정리 및 목록 입력)의 순서로 진행된다(대통령기록관, 2013).

대통령기록물은 일반 공공기록물보다 더 귀중한 사료가 될 수 있는 특수한 영역이며, 법령 및 관리체계가 다르므로, 더욱 이용자 친화적인 서비스체계 개발이 필수적이다. 대통령기록물 활용에 있어 이용자들이 더 쉽게 접근하도록 하기 위해 주제 분류체계 및 시소러스 등의 검색 도구 개발과 적용은 서비스 영역에 큰 도약이 될 것이다.

2.2 시소러스

시소러스란 동의어·반의어 및 용어의 계층

관계, 종속 관계, 기타 관계 등을 보여주는 통제되고 구조화된 어휘집이다. 이는 자연어를 통제된 언어로 변환하기 위한 어휘 통제 도구로, 이를 통해 검색의 재현율을 향상시키고, 동음이의어 통제를 통해 정확률을 높일 수도 있다(기록학 용어사전, 2008). 시소러스는 주제어 부여 시 적합한 용어 선택으로 일관성을 유지하도록 해주며, 탐색 시 유사 어휘 중 적합한 용어를 선택하도록 하고, 용어에 계층성을 부여하여 범주를 확장 또는 축소하도록 하여 검색 효율 향상에 도움을 줄 수 있다.

Aitchison, Gilchrist, & Bawden(2000)은 다음과 같이 시소러스 구축 과정을 제시하고 있다. 1) 주제 분야 정의, 2) 시소러스의 특성 및 배열 선정 (통제어/자연어 시스템 선택 여부 결정), 3) 용어수집 뒤 상위범주 용어부터 규정하고 할당하는 연역적 방법과 용어 발견 시마다 색인으로 활용하고 범주에 할당하는 귀납적 방법 선택, 4) 용어 선택 및 기록 (용어집, 사전 등 자원, 사용자 질의용어, 전문가 경험/지식), 5) 구조 선택 (용어들을 주요 주제별 범주로 조직화하고 관련어 등 연결), 6) 체계적 표현을 수반하는 자모순 시소러스 개발, 7) 전문가와 최종 검토 (주제분야 전문가와 상의 및 승인 필요), 8) 편집 (용어들 상호 체크, 표기 체크, 철자 체크, 들여쓰기 체크), 9) 테스트 및 생산 (초안 활용하여 선별된 문서에 적용, 수정 후 배열 결정, 출판) 등이다. 한편 Shearer(2004)는 시소러스 구축 절차를 다음과 같이 제안하였다. 1) 문헌과 기타 정보원으로부터 자연어 수집, 2) 유의어는 연결하고 동형의어는 구별, 3) 용어들을 넓은 범주로 그룹핑, 4) 범주와 하위 범주를 순서대로 배열, 5) 중요한 차이를 매우

기 위해 시소러스에 더 많은 용어를 추가, 6) 기호법 추가, 7) 다른 범주에 있는 용어들간의 관계를 확인, 8) 분류된 시소러스에 알파벳순 색인을 생성, 9) 알파벳순으로 표현하는 방법을 개발 등이다.

시소러스 구축 관련 국제 표준 ISO 25964에서는 시소러스 구축 절차를 계획, 구현초기, 구현, 시소러스 공개, 업데이트로 나누고, 실제적인 적용 방안을 제시한다. 예를 들어, 먼저 계획 단계에서 목적과 대상, 업무 및 책임분담, 사용 예정 소프트웨어를 명확히 할 것을 제시하고, 구축에 이용자 및 이해관계자와 시소러스 전문가가 참가해야 함을 제시하였다. 시소러스 공개 시 주제분야, 언어, 참고한 표준, 약어나 특수문자에 대한 설명, 총 용어 수, 용어 선택 및 복합어, 우선어 선택 등 시소러스 구축 전반에 있어 적용된 중요한 규칙 등에 대해 설명해야 함도 권고하였다. 또한 배포를 위해 전자 시스템에서 시소러스를 운용할 것을 제시하고, 시소러스 변경이 발생할 수 있는 상황으로 새로운 개념의 추가, 비우선어의 추가, 우선어나 비우선어의 수정, 우선어 삭제, 우선어와 비우선어의 교체, 사용범위 주기 및 정의의 변경 등을 제시하였다 (박옥남, 2011).

한편 시소러스 구축 시 용어 수집의 소스는 매우 다양하다. 즉 일반적인 또는 전문적인 시소러스 용어리스트, 범용적 십진 분류, 전문사전, 어휘집, 용어사전, 용어집 등, 저널 색인 및 초록지와 해당 분야 출판물 색인, 문서의 제목, 초록 및 전문에서의 빈도수 분석 또는 자동 용어추출, 시스템에 나타난 질의어, 사용자 및 전문가의 경험과 지식, 편집자의 경험과 지식 등이 있다 (Aitchison, Gilchrist, & Bawden, 2000).

3. 현황 및 사례 연구

3.1 대통령기록관 주제 분류체계

대통령기록관 주제 분류체계는 2013년에 1-15대 대통령기록물 건명의 형태소 분석 결과와 대통령기록관 이용자 검색 용어 분석 결과 및 대통령기록관 정보공개청구 용어 분석 결과를 활용하여 이용자 친화적인 주제 분류체계로 구축되었다. 이 주제 분류체계는 대분류-중분류-소분류로 나뉘며, 대분류 주제어는 총 24개, 중분류 주제어 135개, 소분류 주제어 394개로 구성되었고, 본표 외 '개별 대통령', '국제기구', '지역'에 해당하는 3개의 별표는 대분류 주제어 총 3개, 중분류 36개, 소분류 289개로 구성되었다. 가장 상위의 대분류는 <표 1>과 같이 24개의 항목으로 구성되었다. 각 대분류 항목의 활용에서 식별성을 높이고자 영어 주제어를 차용하여 영어 대문자 두 문자를 활용한 코드를 부여하였고, 각각의 대분류 안에는 중분류와 소분

류의 주제어 항목을 배치하였으며, 대분류어를 제외한 중분류 주제어와 소분류 주제어는 바로 주제어로 채택하도록 하였다. 기록물 건별로 다수의 주제어를 중복 부여하는 다중분류체계를 채택하여 복합적 내용의 기록물에 대한 유기적인 이용자 접근을 제고하고자 하였다(이혜영, 2013). 본 연구는 기존 24개 영역 중 앞으로 기능시소러스를 구축하는 데에 사용될 '행정지원' 분야를 제외하고 23개 영역을 확장하고 세분화하여 시소러스에 사용될 용어들을 추출하는 데 사용하였다.

3.2 관련 시소러스

여러 시소러스들이 본 연구를 진행하는 데에 벤치마킹의 대상이 되기도 하였고, 또 피해야할 단점을 확인한 대상이 되기도 하여, 시소러스 구축에 영향을 주었다. 백악관 기록관리실 주제파일(WHORM Subject File)도 기록물에서 도출된 용어들과 비교하여 매핑하는 대상이 되었

<표 1> 고도화된 대통령기록물 주제 분류체계 대분류 항목

대분류	코드의 준거 용어	대분류	코드의 준거 용어
1. GV. 국정운영	Government	13. IN. 산업	Industry
2. CT. 건설교통	Construction and Transportation	14. WF. 여성가족	Women and Family
3. ST. 과학기술	Science and Technology	15. DP. 외교	Diplomacy
4. ED. 교육	Education	16. DS. 재난안전	Disaster and Safety
5. DF. 국방	Defence	17. FE. 재정경제	Finance and Economy
6. LB. 노동	Labor	18. PL. 정치	Politics
7. AF. 농림수산	Agriculture and Fishery	19. IT. 정보통신	Information Technology
8. PT. 대통령	President	20. LA. 지방행정	Local Administration
9. CS. 문화관광체육	Culture and Sports	21. UF. 통일	Unification
10. JT. 법무	Justice	22. AD. 행정지원	Administration
11. HM. 보건의료	Health and Medicine	23. PB. 홍보	Publicity
12. SW. 사회복지	Social Welfare	24. EV. 환경	Environment

으며, 국가기록원 시소러스는 대통령기록관의 상위기관이라는 점에서 시소러스 작성 지침과 함께 직접적으로 참고할 수 있는 대상이 되었다. 국회도서관 시소러스도 폭넓은 용어와 다양한 관련 용어의 제시로 수시로 참고할 대상이었으며, 국사편찬위원회 한국역사용어 시소러스도 시소러스 자체 및 온라인 제시 방법 등에서 주요한 참고원이 되었다. 신문기사 종합시소러스는 좀 오래 되었으나, 본 연구의 대상이 된 기록물도 초대 대통령부터의 기록물이 분석 대상이어 폭넓은 대표어와 관련어 등의 선정에 중요한 참고자료로 사용되었다. UKAT(UK Archival Thesaurus)는 UNESCO 시소러스와 유사하여 구체성은 떨어지나 용어의 제시 방안 등에 참고가 될 수 있었다. 이들 중 시소러스에 대한 선행 연구들에서 다루진 백악관 기록관리실 주제파일(정광훈, 남영준, 2013), 국가기록원 시소러스(박부숙, 2007; 최윤경, 정연경, 2014), 국사편찬위원회 한국역사용어시소러스(최윤경, 정연경, 2014), UKAT(최윤경, 정연경, 2014) 등은 제외하고, 국회도서관 시소러스, 신문기사 종합시소러스만 간략히 살펴본다.

3.2.1 국회도서관 시소러스

국회도서관의 표준 시소러스는 1998년부터 추진해온 국회전자도서관 구축사업의 연장선상으로 2000년도부터 개발하여 2003년도 완료하였으며, 웹 시소러스는 계속 업데이트 되고 있다. 국회도서관 표준 시소러스는 2000년도부터 2002년도까지 3개년 사업으로 기 구축한 14개 주제 분야의 시소러스를 통합한 것으로, 각 주제 분야별 시소러스 간의 중복용어들을 통제하고 용어관계를 재조정하는 동시에 기술방법

을 통일하는 등 일관성 있고 잘 정의된 단일체계를 지닌 시소러스이다. 수록 대상 주제로 정치, 경제, 법률, 행정, 사회학, 사회복지, 여성, 교육, 심리, 신문방송, 문헌정보, 정보통신, 교통, 역사지리, 인류민속 등이 망라되었으며, 각 색인어 사이에는 동의관계, 계층관계, 연관관계 등이 정의되어 있고, 계층관계가 없는 고립어도 있다(김남희, 2005). 온라인 검색 결과에 있어서는 국가기록원 시소러스와 마찬가지로 복합용어의 특성을 잘 반영하지 못한다.

3.2.2 신문기사 종합시소러스

한국언론연구원(현 한국언론진흥재단)은 1991년 1월 1일부터 온라인으로 각 언론사 기사자료를 KINDS(Korean Integrated Newspapers Database System)라는 이름으로 제공해왔으며, 이의 효율적인 검색을 위해 신문기사 종합시소러스 개발에 착수하여 1991년 말 1차로 완성된 후 수정·보완작업을 거쳐 1992년 출판하였다. 신문기사의 주제가 되는 중요한 어휘를 모아 동의관계, 계층관계, 연관관계 등으로 체계화하였으며, 자모순 시소러스와 분류별 계층색인, 분류별 자모색인으로 구성하였다. 우선어(대표어) 7천 2백여 개, 비우선어(비대표어) 3천 5백여 개 등 모두 1만 7백여 개의 색인어를 담고 있다. 색인어의 구성은 일반주제명, 기관단체명, 지역명으로 이루어지며 우선어를 표현하는 비우선어의 표현이 많아 포괄성이 크다(한국언론연구원, 1992). 이후 업데이트가 진행되지 않아 새로운 용어들이 추가되지 않고 온라인에 활용이 되고 있지 않아 시대성이 떨어지기는 하나 다양한 관련 용어 확인에 유용하다.

4. 대통령기록물 주제 시소러스 구축 절차

일반적으로 시소러스 구축의 절차는 해당 분야의 자료에서 주제어를 추출하고 시소러스를 구축하는 귀납적 방법을 사용하지만, 본 연구에서는 2013년도 기 수행된 대통령기록물 주제 분류체계에 사용된 용어를 중심으로 시소러스를 구축하는 연역적 방법을 사용하였다. 즉, 앞서 구축된 대통령기록물 주제 분류체계의 주제 영역 명을 최상위어로 규정하고, 기록물 건명에 사용된 용어들을 분석하여 많이 사용된 용어들과 이용자들이 사용하는 용어들을 추출하고 주제 분류체계의 관련 범주에 매핑하여 용어들을 배치하였다. 즉 이용자들의 대통령기록관 웹사이트 검색 용어, 정보공개청구 용어 및 미국 백악관 주제파일의 용어 및 국회도서관 시소러스와 신문기사 시소러스 등을 활용하여 시소러스를 구축한 것이다.

대통령기록관의 특성을 반영하여 대통령 주제 분야는 중점적으로 세분화하여 정리하였다. 대통령기록물 제목에서 확인된 고유명사인 인명, 지명, 법률명, 기관명, 국제기구명, 행사명, 사건명, 직책명 등에 대해서는, 일부 주요 용어는 시소러스에 반영하였지만 이들은 모두 따로 정리하여, 후에 패시 검색이 가능한 옵션으로 제공하기로 결정하였다. 전반적인 개발 절차 및 원칙은 『국가기록원 시소러스 지침』(2009)을 따랐다. 이 과정을 좀 더 상세히 제시하면 다음과 같다.

4.1 기록물 건명 및 이용자 사용 용어 분석

4.1.1 대통령기록물 제목 형태소 분석

먼저 시소러스에 편입될 용어 도출을 위해 제1대-제15대 대통령기록물 25만 건, 제16대 대통령 기록물 120만 건의 제목을 대상으로 형태소 분석을 실시하였다. 형태소 분석은 기록물 건명 전체를 텍스트 파일로 만든 후, 형태소 사전, 전문용어 사전, 사용자정의 사전, 기 분석 사전 등 다수의 사전을 기반으로 한 형태소 분석기 소프트웨어를 통하여 명사, 복합명사, 추정 명사를 추출하는 것이다. 예를 들면 기록물 ‘인선문제에 관한 건’은 ‘인선문제’, ‘인선’, ‘문제’ 등으로 추출되는 것이다. 이렇게 기록물 제목에서 추출된 명사는 순서 없이 나열되어 있으므로 단어의 빈도를 계산하기 위해서 가나다순으로 정렬하고, 가나다순으로 정렬된 단어에 대해서 빈도수를 계산하게 된다. 즉, 단어1, 단어1, 단어1, 단어2, 단어2, 단어3이 추출되었다면, 단어1:3(회), 단어2:2(회), 단어3:1(회)로 빈도수가 나타나게 되는 것이다. 추출된 단어 중에서 빈도수를 기준으로 주제어를 추출하고, 빈도수가 아주 높은 조사를 비롯한 불용어들을 제외하고 나머지 단어들 중에서 의미있는 단어를 추출하게 된다(이해영, 2013).¹⁾

제1대~제15대 대통령 기록물 25만 건의 분석 결과, 빈도수가 1회 이상인 단어의 수는 약 14만 건, 빈도수가 30회 이상 단어의 수는 4,387건으로 확인되었다. 또한 제16대 대통령 기록물 120만 건의 분석 결과, 빈도수가 1회 이상인 단어의 수는 약 46만 건, 빈도수가 30회 이상인 단

1) 이 과정은 전문 형태소 분석기를 활용하여 이산TS라는 회사에서 진행하였다.

〈표 2〉 1~16대 대통령 기록물 건명에 나타난 빈도수 높은 용어

1-15대 대통령 기록물		16대 대통령 기록물	
단어	빈도수	단어	빈도수
자문위원신상기록표	21,342	회의	105,360
회의	21,326	대통령	86,590
계획	15,404	관련	76,777
결과	13,116	위원회	68,118
통일	11,218	보고	56,817
보고	11,176	계획	55,343
위원회	10,852	국정	53,538
대통령	10,602	과제	52,037
년도	10,580	업무	51,989
제출	10,312	요청	50,505
개최	10,153	관리	48,809
통보	9,166	자료	46,108
업무	8,958	사항	45,739
지출	8,501	정책	43,022
자료	8,109	결과	40,669
지방	7,814	의견	37,995
협의회	7,687	추진	37,517
지역	6,934	개최	36,385
공무원	6,893	국정과제	36,365
요청	6,827	대통령님	35,816

어의 수는 20,054건으로 확인되었다. 참고로 대통령기록물 건명에서 가장 많이 추출된 빈도수 상위의 용어는 〈표 2〉와 같다. 이렇게 추출된 단어 중, 빈도 수 30회 이상인 4,387건과 20,054건의 단어를 시소러스 구축을 위한 대상 용어로 활용하기로 결정하였다.

4.1.2 대통령기록관 웹사이트 이용자 검색 용어 분석

대통령기록관은 2008년 6월 웹사이트를 개설하여 역대 대통령의 기록물과 그에 대한 정보를 검색하도록 하고 있다. 본 연구에서는 대통령기록관 웹사이트 시스템에 남겨진 이용자들의 검색 용어에 대한 로그 분석결과를 분석하여 활용함으로써, 가능하면 이용자들이 검색에 사

용하는 용어를 주제 시소러스에 반영하고자 하였다. 2010년 1월부터 2010년 11월 및 2011년 1월부터 2013년 5월까지를 포괄한 검색 용어 약 660만 건 중 의미있는 것으로 판단된 3번 이상 입력된 용어들을 취합하여 활용하였다. 이용자 검색 용어 분석은 이해영(2011) 논문에서 제시된 방법을 활용하였다.

3회 이상 검색된 용어 중 가장 많이 추출된 용어 15개는 〈표 3〉과 같다. 이 중 가장 많이 입력된 검색용어인 대통령입기, 대통령선물, 대통령선거, 대통령의 꿈 등의 용어는 대통령기록관 웹사이트의 추천 키워드였다. 종합적으로 개별 대통령에 대한 검색 경향을 살펴보았을 때, 아직까지 이용자들은 구체적이고 정확한 사안의 검색어를 적절하게 활용하지 않는 것으로 파악되

〈표 3〉 대통령기록관 포털 검색 빈도수 높은 용어 ('포함'은 복합어로 검색된 경우를 포함)

단어	빈도수
대통령입기	2,601
대통령기록	2,364
대통령선물	2,232
대통령연설	1,946
대통령	1,867
박정희(포함)	1,552
대통령선거	1,334
노무현(포함)	1,330
대통령님	1,069
대통령의 꿈	984
이승만(포함)	846
김대중(포함)	563
대통령꿈	538
전두환(포함)	427
이명박(포함)	413
연설문	401
청와대	375

므로, 시소러스 등의 추가적인 검색 도구를 제공하여 원하는 정보에 좀 더 쉽게 접근할 수 있는 안내가 필요함을 확인할 수 있었다.

4.1.3 대통령기록관 정보공개 청구 용어 분석
 대통령기록관의 이용자 정보공개 청구는 웹사이트 검색에 비해 시간과 수고가 많이 걸리는 작업이므로 이용자들이 기록물을 활용하려는 특정 목적을 위해 용어 선정에 신중을 기했을 것으로 판단하였다. 따라서 정보공개 청구 용어를 분석하는 작업은 검색용어 분석에 비해 보다 더 비중 있는 이용자의 요구를 파악할 수 있다고 보아, 대통령기록관에 요청되었던 2010년부터 2014년 6월까지의 정보공개 청구에 대한 처리 목록 582건을 분석 대상으로 삼았고 모든 용어를 다 분석하였다. 정보공개 청구 처리 목록은 접수일자, 청구사항(내용), 대통령별로 작성되었는데, 인

명, 직급, 기관명, 일시 등을 제외하고 유의미한 정보공개 청구 용어를 분석한 결과 유의미한 용어 627개를 추출하였다. 추출된 용어 중 가장 많이 청구된 용어는 〈표 4〉와 같다. 분석 결과, 대통령 개개인의 이력 확인, 여러 대통령이나 정부인사와 관련된 정보 확인(역대 대통령 자문위원회, 역대 대통령 홍보수석과 대변인 임명자 목록 등), 대통령별로 규정할 수 없는 다양한 정보, 예를 들면 인수위원회, 대통령기록물, 대통령(존영)사진 등이 요구됨을 확인하였다.

〈표 4〉 대통령기록관 정보공개 청구 빈도수 높은 용어(대통령별 및 일반 용어별)

	용어	빈도수
대통령	박정희	186
	노무현	61
	전두환	45
	이승만	33
	김대중	31
	김영삼	27
	이명박	19
	최규하	8
	노태우	5
	윤보선	3
	박근혜	2
일반용어	기록	82
	사진	20
	취임	11
	기록물목록	10
	광주대단지	9
	서울	9
	지시사항	9
	의전일지	8
	2007남북정상회담	8
	대통령실	7
	인수위원회	7
	정보공개	7
	연설(문)	6
	남북회담	5
	명단	5
	미국	5
	연두순시	5

4.2 용어 매핑

대통령기록관은 2013년도 연구사업을 통해 대통령기록물 주제 분류체계 24개를 확정하였다. 대통령기록물 주제 분류체계 구조는 대분류-중분류-소분류 및 일부 세분류로 구성되었으며, 각각 24개-135개-394개-20개의 대표어로 구성되어 있다(이해영, 2013).²⁾ 이 분류체계 표에 제1대~제16대 대통령기록물 건명의 형태소 분석 결과 용어를 중분류, 소분류 단위까지 상세 매핑하여 입력하였다. 매핑 과정에서는 주제 분류체계에 같은 용어가 있는지 먼저 살펴보고, 해당 용어가 주제 분류체계에 사용된 용어와 동일할 경우 우선 매핑하여 입력하였다. 용어가

어떤 맥락에서 사용되었는지 확인이 필요한 경우도 많아, 용어가 추출된 기록물 제목을 수시로 확인하고 사용 맥락을 파악하여 매핑을 확정하였다. 매핑과정에서 기존의 대분류-중분류-소분류 및 세분류를 확장하여 정리하였다. 앞서 설명한대로 그 중 인명, 지명, 법률명, 기관명, 국제기구명, 행사명, 사건명, 직책명 등은 표에서 따로 구분하였다(〈그림 1〉 참조). 또 다른 주제 분류체계 표에는 앞서 분석한 대통령기록관 정보공개청구 용어, 대통령기록관 웹사이트 이용자 검색용어를 주제 분야에 맞추어 나란히 정리하였고, 추가적으로 백악관 기록관리실(WHORM) 주제 파일을 분석하여 모든 용어를 표에 매핑하여 입력하였다(〈그림 2〉 참조).

8. PT. 대통령 UE: 대통령 명남	청와대 시설				<16대> 민정, 민정비서관, 민정비서관실, 민정비서관, 민정수석, 민정수석실, 민주행정, 민주화, 민주화운동, 본관의전, 본위원회. 대통령관련, 대통령국경, 대통령기록관리역, 신T/F, 대통령보고서, 대통령보좌, 대통령보좌시스템, 대통령비서, 대통령비서관, 대통령비서관장, 대통령실, 대통령자문, 대통령자문위원회, 대통령직속, 대통령집무, 대변, 대리, PPR(비서관실?)	김 대통령(2705), 김영삼(261),	집무실(104), 집무환경(49), 접견실, 접견장, 영빈관 본관관리, 본관식당
	청와대 방문		건축(146)				청와대Q&A, 청와대 게시판, 청와대블로그, 콘서트, 온라인국민, 온라인대학,
	청와대 홈페이지			국민제안(99)			이철민원, 정부민원, 정부민원내화센터, 이지원, 이지원시스템, 이지원운영, 이지원 운영지원
	청와대 민원						비서관, 비서관급, 비서관님, 비서관님방, 비서관실, 비서관실과, 비서관회, 비서관회의, 비서관회의자료, 비서관, 비서관, 비서관개인, 비서관개인전사장비, 비서관관련, 비서관내, 비서관운영, 비서관운영책, 비서관장, 비서관장님, 비서관장수, 비서관장수보, 비서관장수보회, 비서관장수보회의, 비서관장실, 비서관장주관, 비서관장주재, 비서관장직속, 비서관장직속실, 비서관장직속실회, 비서관장직속실회의, 비서관장직속회의, 비서관장직속, 비서관장직원, 비서관회, 비서관회의, 비서관업무
대통령 경호실		경호실				노동비서관, 농어촌비서관,	
대통령 행사	시찰	순시, 민정현장 방문	순시(199), 지역 순방(94), 정시찰(58)	지역 관문	생일(90), 격려금(101), 개막식(185), 개회식(157), 후원(142)		
	기념일	식목일, 국군의날	신년 인사 회(100), 창설(107), 식목일				

〈그림 1〉 주제 분류체계에 대통령기록물 제목 추출 용어 매핑

2) 본표 외 '개별 대통령', '국제기구', '지역'에 해당하는 3개의 별표는 대분류 주제어 총 3개, 중분류 36개, 소분류 289개로 구성되었다.

	주제 분류체계	정보공개청구	2010-2013년 로그분석	백악관	
8. PT 대통령	청와대	청와대 시설	-개전절 공식 행사 2, 식목일 행사 1, 연설(문) 5, 담화문 2, 취임(사 1, 식 4) 6, 신년사 1, 신년인사회 1, 국산전전자교환기1, 천만회선개통기념식 참석 1, 우수발명인초청 다과회 1, 크리스마스 메시지 1, 격려 2, 편지 1, 서신재발급 1, 친서철 1, 의전일지(일지 포함) 8, 국민과의 대화 1,	-취임식 '대통령취임' 등 '취임' 포함 116, -'어버이날' 등 '날' 포함 17, -(광복절)(8.15)경축사 14, -(어린이)(현장)보도대회 11, -'방문' 포함 11, 전국세마을 지도자(대회) 8, -공모(전) 6, -'전시(회)' 포함 5	기념일(법적으로 승인 권고된 휴일 및 행사) -국군의 날 -권리장전-인권의 날, 주, 년 -보이-걸 스카우즈 주간 -설날 -크리스마스-새해 -크리스마스 인사(국가 원수 송/수신) -시민의 날- 헌법 주간 -콜럼버스의 날 -국가 범죄 예방 주간 -국방 주간 -약물 남용 예방 주간 -부활절 -교육 주간 -아버지의 날 -어머니의 날 -화재 예방 주간 -취임 기념일 -독립기념일 -독립기념일 인사 -토머스 제퍼슨 탄생일 -유대일 휴일 -노동절 -법의 날 -링컨 탄생일 -적십자의 날/ 주간 -성 패트릭의 날 -추수감사절
		청와대 방문			
		청와대 홈페이지			
		청와대 민원			
		대통령 비서실			
		대통령 경호실			
	대통령 행사	시찰			
		기념일			
		국경일			
		조찬			
		오찬			
		만찬			
	영부인	간담회			
		리셉션			
		접견			
		기자회견			
		영부인 행사			
		영부인 말씀자료			
	연설	영부인 연설			
		영부인 서한			
국회연설					
메시지					
말씀자료	취임사				
	신년사				
취임·퇴임	취임				
	퇴임				
대통령 기록	대통령 기록				
	대통령 기록				

〈그림 2〉 정보공개청구 용어, 웹사이트 이용자 사용 용어, 백악관 주제어 매핑

4.3 용어 선정 및 관계 설정

앞의 단계에서 매핑이 완성된 두 개의 표를 나란히 하고 시소러스에 편입할 대표어와 비대표어를 선정하였다. 이 과정에서는 다음과 같은 원칙을 설정하였다.

우선, 2013년 대통령기록관에서 기 구축한 기존 주제 분류체계를 반영하여 용어를 선정하고, 필요시 조정하는 것이다. 또한 본 시소러스 구축에서는 기존 주제 분류체계의 상위 항목 조정을 최소화하고 중분류 이하 분류체계는 필요시 조정하고, 소분류, 세분류 용어는 신설하는 것으로 하였다. 가능하면 이미 주제 분류체계에 항목명으로 채택된 용어를 대표어로 선정하는 걸 원칙으

로 하였으나, 일부 기록물 건명에서 도출되지 않는 용어 등은 배제하거나 비대표어로 선정하였다. 또 주제 시소러스에서 대표어로 사용이 적절하지 않은 경우 분류체계도 조정하여 분류항목에 재배치가 가능하면 그렇게 하거나 선정된 대표어의 비대표어(UF)로 조정도 하였다. 예를 들면, 건설교통 대분류 항목 하의 재해복구는, '재난안전' 대분류의 중분류어인 '재난' 항목 아래 소분류 용어인 '재난관리'에 '재난복구'를 새로 만들어 대표어로 선정하고, 재해복구를 비대표어로 조정하였다.

전체적으로 대표어로 선정된 용어는 먼저 약 145만 건에 달하는 제1대~16대 기록물 건명의 형태소 분석 결과, 빈도수가 높은 용어이다. 또 대통령기록관 열람실에 요청된 정보공개청구

서에 수록된 형태소에서 의미있는 용어와 대통령기록관 웹사이트 검색창에 입력된 이용자 검색 용어 로그분석을 통해 상위에 랭킹된 용어들이 대표어로 선정되었다.

비대표어는 대표어와 유사동의어 관계에 있는 용어로서, 대표어에 비해 출현율이 떨어지지만, 검색시 유용하게 사용될 만한 용어로, 맞춤법을 준수한 한글 대표어를 기준으로 하였다. 그러나 기록물 제목이나 이용자 사용 용어에 빈번하게 출현하는 영어나 축약어의 경우는 예외적으로 대표어로 설정하였다. 예를 들면 DMZ, 방폐장 등이 이에 속한다.

추가적으로 비대표어와 관련어(RT)를 선정하기 위해서는 국내 유관 기관의 시소러스를 참고할 필요가 있었다. 우선 『국가기록원 시소러스』를 참고하였고, 『국회 시소러스』와 『신문기사 시소러스』를 활용하여 용어들을 확인하고 비대표어와 관련어들을 확인하고 활용하였다. 그밖에 국어사전 및 백과사전, 그리고 관련 부서의 홈페이지에 제시된 직제, 기능, 업무 설명 등을 참고하여 실제 용어의 사용을 검토하였다.

4.4 시소러스 작성

이상의 분석과 매핑 작업 및 용어의 관계 설정을 마치고 그 결과를 활용하여 계층별 시소러스와 자모순 시소러스를 작성하였다. 완성된 시소러스에 수록된 용어는 모두 3,463개로서, 대표어(USE)는 1,454개, 비대표어(USE, EN)는 2,009개이다. “08대통령” 주제 용어는 모두 473개로서, 전체 시소러스의 14%에 달한다.

먼저 계층별 시소러스의 경우, 23개 주제 분야의 최상위 주제어(TT)로부터 계층깊이의

정도를 ‘·’으로 나타낸 것으로, 최대 계층깊이는 3단계(...)로 표현하였으며, “01국정운영~23환경”까지 23개 주제 분야 순으로 정렬하였다. 각 분야는 높은 계층에서 낮은 계층의 대표어만 수록하였고, 같은 계층의 경우는 한글 자모순으로 정렬하였다. 자모순 시소러스는 대표어의 자모순으로 구성하였으며, 한글(ㄱ~ㅎ), 숫자(0~9), 영어(A~Z) 순으로 정렬하고, 비대표어는 대표어와 함께 자모순으로 정렬하여 ‘→’를 사이에 두고 대표어를 진하게 표기하였다.

사용된 관계 지시기호는 국가기록원 시소러스에도 사용되는 지시기호를 사용하여, SN(Scope Note, 범위주기), UF(Used For, 비대표어), BT(Broader Term, 상위어), NT(Narrower Term, 하위어), NTI(Narrower Term Instance, 하위어(사례관계)), RT(Related Term, 관련어), EN(English, 영어 비대표어) 등이 사용되었다. 관계지시기호의 순서는 SN → UF → BT → NT → RT → EN 순으로 배열하였다.

완성된 시소러스는 다음과 같은 원칙에 의해 작성되었다. 먼저 한글 표기를 우선했으며, 자모순 시소러스에서는 한자와 영어 표기를 부가하여, 한글 표기만으로는 용어의 의미를 명확히 전달하기 어려울 경우 대표어 옆에 한자를 병기한다(예: 추곡수매[秋穀收買]). 둘째, 개념상 띄어쓰기가 원칙이나, 전문용어나 고유명사일 경우, 혹은 띄어쓰기 없이 사용해야 의미가 잘 전달될 경우는 띄어쓰기 없이 표기한다(예: 국정감사). 셋째, 동음이의어인 경우 대표어 끝에 한정어 기호(@)를 사용하며, 한정어는 대표어를 쉽게 인식할 수 있는 한글 용어를 사용한다(예: 포장@상훈). 넷째, 대표어 및 비대표어에 사용된 특수 기호는 온점(.)과 붙임표(-)이다

(예: 3.1절). 다섯째, 대표어는 기본적으로 한글이지만, 영어로 된 용어를 일상적으로 쓸 경우 대표어로 쓴다(예: DMZ). 한글로 된 대표어의 영어 표기가 자주 쓰일 경우 풀네임과 축약된 형태를 모두 비대표어로 표기한다. 여섯째, 정식 명칭을 우선하되, 약칭이 일반적으로 더욱 알려진 경우에는 그 역으로 하였으며, 이 경우 반드시 비대표어로 참조하였다(예: 방폐장[放廢場] UF 방사성폐기물처리장 외). 완성된 시소러스의 일부를 보여주는 <그림 3>은 계층별 시소러스의 일부 예이고, <그림 4>는 자모순 시소러스의 일부 예를 보여주고 있다.

04 교육	
· 교육문제	· 교육정책
· · 공교육	· · 교과서
· · 교권침해	· · · 교과서 제도
· · 교육양극화	· · 교육개혁
· 사교육	· · 지방교육혁신
· · · 과외	· · 교육지치
· · 사교육비	· · 지방교육행정
· · · 학원	· · 국제교육
· · 학교폭력	· · · 유학생
· · · 왕따	· · 임시제도
· · 학생인권	· · 대학입시
· · · 두발자유	· · · 수확능력시험
· · · 체벌	· · · 평준화

<그림 3> 계층별 시소러스의 예

교육문제	교육정책
BT 교육	BT 교육
NT 공교육	NT 교과서
교권침해	교육개혁
교육양극화	교육지치
사교육	국제교육
학교폭력	임시제도
학생인권	
교육복지	교육지원
BT 교육지원	BT 교육
NT 결식아동(除食兒童)	NT 교육복지
방과후학교	교육경보화
학교급식	교직원
	정학금
	체험학습
	학자금

<그림 4> 자모순 시소러스의 예

5. 시소러스 활용 방안 및 제언

이렇게 구축된 시소러스는 업무와 이용자 서비스에 여러 가지 방법으로 활용이 가능하다. 특히 대통령기록관에서 기록관리 업무에서의 활용방안과 온라인 서비스에의 활용방안을 알아본다. 또한 앞으로 시소러스의 발전 방향에 대해서도 제시해 본다.

5.1 시소러스의 업무에의 활용 방안

시소러스는 첫째, 정리업무 시 색인어 작업에 활용하는 것이 가능하다. 즉 기록물 기술요소 중 색인어 요소에 시소러스를 이용하여 통제된 주제어를 부여함으로써 통일성 있고 일관성 있는 색인어의 부여가 가능하며, 이렇게 시소러스를 이용한 색인 작업은 검색의 재현율과 정확률을 높이는 데에 기여할 것이다.

둘째, 주제 가이드 개발 작업을 지원할 수 있을 것이다. 일반 이용자들에게는 계층적 기술체계가 복잡하고 어려우므로, 기록의 맥락을 이해할 수 있도록 도와주는 안내도구로서 다양한 주제 가이드(research guide)가 필요하다. 국가기록원의 '주제별 검색' 서비스와 같은 내용해설 위주의 가이드가 아니라, 철저하게 소장하고 있는 기록에 기반하여 만들어지는 가이드가 필요하며, 주제 분류체계에 기반한 시소러스가 이의 바탕이 될 수 있을 것이다. 또 시소러스의 용어들이 기록물에 부여되고 유지된다면, 기록물 가이드를 만드는 데에 상당한 기여를 할 수 있을 것이다.

셋째, 기술(description) 작업 시 활용이 가능하다. 즉 시소러스의 대표어는 이관된 기록물에

대한 기술 작업에서도 표준화된 용어를 활용하는 기술을 작성하도록 하여 기록물 관리에 통일성과 일관성을 높이며 검색 개선에 도움을 줄 수 있다.

넷째, 기록물 생산단계부터 사용하도록 한다면 기록물에 사용되는 용어 표준화가 가능할 것이다. 즉 시소러스의 대표어를 생산단계에 기록 진명이나 색인어에 사용하도록 하면 용어 표준화와 통일성을 유지하는 것이 가능하다. 이렇게 기록의 생산 단계에서 시소러스 용어를 최대한 활용토록 대통령비서실, 대통령경호실 및 대통령 자문위원회 등 관련 기관과 협의가 필요하다.

다섯째, 후속 기능시소러스의 개발과 전거레코드가 구축되어 함께 활용되면 기록관리 업무의 효율성이 향상될 것이다. 기록의 기술 업무에 있어, 주제어를 부여하거나 연혁 정보나 기능의 기술 등에 같은 내용을 반복하는 일이 줄어들게 될 것이다. 국제 표준기술표준의 동향을 보면, 기록물 기술레코드, 전거레코드에 기능레코드를 별도로 구축하여 연계하는 다중개체 모형을 지향하고 있다. ISDF를 준용한 '기능' 레코드 구축을 하면 시소러스의 대표기능어를 표제어로 적용할 수 있을 것이다. 특히 대기능과 같은 범주구분은 기록물 집합체나 기록생산기관별 브라우징 서비스에 활용할 수 있을 것이다.

5.2 시소러스의 온라인 서비스 활용 방안

시소러스는 첫째, 주제어 브라우징에의 활용이 가능하다. 본 연구에서 개발된 시소러스는 주제 분류체계에 기반한 것으로, 23개 대주제 아래 중분류, 소분류 및 세분류 용어 관계가 설정되어 있으므로, 이 계층체계를 온라인에서 이

용자들이 확인할 수 있도록 하면 편리할 것이다. 이 기능은 이용자들이 정확하게 기록물에서 사용된 용어를 잘 모를 경우에 쉽게 기록을 확인하도록 할 수 있다. 주제 분류체계는 대통령기록관 웹사이트의 기록검색 메뉴에서 바로 선택해서 활용할 수도 있으며, 현재 대통령기록관 기록검색 웹사이트의 '빠른 검색' 화면을 따로 활용할 수도 있다(<그림 5> 참조).

둘째, 검색어 입력 관련하여 활용이 가능하다. 즉 이용자가 원하는 용어를 입력하면 주제 분류체계에서 해당 용어가 들어간 분류체계의 중분류 및 소분류와 세분류 체계가 펼쳐지고, 이용자들이 상하위어 등을 확인 가능하도록 하는 방식을 제안하고자 한다(<그림 6> 참조). 특히 해당 분류체계에 입력 용어가 포함된 기록의 건수도 같이 제시되도록 하면, 이용자들에게 더 쉽게 용어의 선택 여부를 판단할 수 있다.

또한 이용자가 원하는 용어를 입력한 상자 아래쪽에 입력 용어가 포함된 시소러스의 대표어들이 제시되고, 제시된 용어를 펼치기 하면, 각각의 용어가 속한 주제 카테고리 목록이 제시되도록 하여 이용자들이 쉽게 용어 및 관련 용어를 선택할 수 있도록 하면 편리할 것이다. 또 용어입력 상자 아래에 제시된 용어들에는 이용자가 마우스를 가져가면 해당 용어의 하위어와 관련어가 제시 되도록 하면, 이용자들이 본인이 입력한 용어와 관련되거나 하위 계층에 속하는 용어를 다양하게 선택할 수 있어 기록 검색이 매우 쉬워지게 될 것이다.

셋째, 상세검색 옵션으로 시소러스의 용어 계층 구조를 활용하도록 할 수 있다. 국가기록원이나 국회도서관 등에서 사용하는 시소러스는 이용자들에게 소장자료 검색에서 '관련용어



〈그림 5〉 대통령기록관 웹사이트 ‘빠른 검색’ 기능의 활용 방안



〈그림 6〉 시소러스를 활용한 온라인 서비스 방안

검색'이나 '연관어 검색' 기능을 선택하면 시소러스의 계층구조를 바로 보여주는 방식으로 시소러스에서 사용된 대표어, 동의어(비대표어), 상위어, 하위어, 관련어 등 관련 용어를 제시하고 있다. 대통령기록관에서도 이러한 시소러스 계층구조 그대로를 제시하는 방법도 시소러스 사용이 익숙한 이용자들을 위해 제시할 수 있다. 또 이용자들이 용어를 입력하면 이 용어가

주제어로 부여된 기록이 검색결과 상위에 나오도록 하는 것도 좋은 방법일 것이다. 주제어가 부여되지 않은 기록에 대해서는, 대표어나 그 동의어인 비대표어가 기록건명에 나타나는 기록물들이 검색 결과에 제시되도록 할 수 있을 것이다.

넷째, 실제 이 연구를 통해 도출되었으나 주제어 시소러스에는 포함되지 않은 다양한 고유

명사를 활용하여 카테고리별 옵션을 활용하면, 대통령기록 검색의 효율성을 극대화 할 수 있다. 개별 대통령 시기 별 주요 정책 및 사건 (예: 3.15부정선거, 4.19혁명, 새마을 운동, 이산가족 찾기 등), 주요 행사 및 회의 (예: ASEM, 정부 혁신토론회 등), 기관 및 단체명 (예: 21세기 위원회, 교육정책자문회의, 전국교직원노동조합, 한국도로공사 등), 직책 (예: 사단장, 사령관, 수석비서관, 재무장관 등), 조약 (예 한미 FTA 등), 법률명 (예: 『정부출연연구기관 등의 설립·운영 및 육성에 관한 법률』 등), 지역 및 국가 (예: 아프리카, 유럽, 미국, 일본 등) 등이 가능할 것이다. 이는 기록유형과 이용자 요구 등을 분석하여 실질적으로 가장 유용한 카테고리 선택하는 것이 필요하다.

5.3 발전방향 및 제언

이용자들에게 기록물 검색은 큰 난제이다. 출처 중심의 접근 방식은 기록에 담긴 내용에 대한 다양한 주제 접근을 제공하지 못하는 한계가 있다(설문원, 2010). 이러한 상황에서 시소러스의 활용은 이용자들이 쉽게 생각하는 주제 접근을 할 수 있도록 하는 효율적인 도구가 될 수 있다. 이렇게 주제로 접근하게 하는 것에 덧붙여, 타 논문에서도 제안하였듯이(이윤령, 이해영, 2014), 기록물 검색은 가능하면 다양한 옵션을 제시하여 이용자들이 원하는 대로 선택할 수 있도록 하되, 특정 형태나 종류의 기록물은 결과에서 이용자들이 다양하게 제외(exclude)할 수 있는 옵션이 제공되도록 하는 것이 바람직하다. 또한 검색결과가 유형별, 시기별, 기록물 포맷별 등으로 다양하게 제시되도록 해야

할 것이며, 또한 검색 결과에 기록물 간략 내용을 제시하여 따로 기록을 살펴보지 않더라도 이용자가 원하는 기록인지 쉽게 파악되도록 해주는 것이 필요하다.

앞으로 대통령기록물 검색도 좀 더 발전시켜 나아가면 의미기반의 이용자 의도를 반영한 검색을 제시하는 방향으로 가야할 것이다. 의미기반 검색은 '사용자의 명시적인 의도뿐만 아니라 묵시적인 의도를 이해하기 위해 온톨로지와 기타 방법들을 이용하여 사용자가 제시한 키워드 검색어에 대하여 원하는 결과들을 검색하는 과정'으로(박상언, 임재익, 강주영, 2013), 좀 더 사용자의 검색 의도에 맞는 결과를 주고자 하는 정보검색 분야의 공통의 고민을 해결하기 위한 방법이다. 팀 버너스 리 등이(Berners-Lee, Hendler, & Lassila, 2001) 시맨틱 웹에 대해 제안한 이후 많은 연구들이 진행되고 있으며, 기록물 검색과 관련하여 이병길, 김희섭(2013)은 새마을운동과 관련된 기록의 메타데이터 정보를 기반으로 메타데이터 요소로 기록물의 주제(분야)·생산년도·물리적 형태·내용을 선정하여, 이용자가 보다 편리하고 보다 정확하게 새마을운동 기록물을 검색할 수 있는 온톨로지 기반 검색 시스템의 개발하고 성능을 평가하였다.

대통령기록관의 기록검색도 앞으로 이용자들이 주제어 및 다양한 기록의 유형이나 기록이 메타데이터에 포괄하고 있는 다양한 기록의 속성을 활용하여, 이용자들이 실질적으로 원하는 기록을 쉽게 찾아갈 수 있는 의미기반의 기록검색이 가능한 방향으로 발전시켜나가도록 해야 할 것이다. 주제 시소러스는 앞으로 개발될 기능시소러스와 함께 이를 위한 주요한 기반이 될 수 있을 것이다. 예를 들면 이용자들이

입력한 용어와 관련하여 시소러스에서 상하위어 및 관련 용어들을 제시해주되, 함께 출현하는 빈도가 높은 주제어를 포함하는 기록을 검색 결과에서 앞에 제시해주어 이용자들이 기록을 살펴보도록 하면 더 적합한 결과를 얻을 수 있을 것이다.

이와 함께 이용자가 입력한 용어가 포함된 기록의 종류, 기관명, 지명, 행사명, 직책명, 인명 등 고유명사 등의 패시를 활용하여 기록의 숫자가 많은 순서대로 제시하여, 이용자가 그 중에서 선택을 하도록 하면 원하는 기록을 쉽게 찾아가도록 해줄 수 있다. 그 외에 입력되거나 주제 분류체계에서 선택된 용어를 포함하는 기록의 메타데이터 정보를 기반으로 기록물의 형태나 포맷 등 옵션들을 다양하게 제시하여 이용자들이 원하는 방향으로 결과를 좁혀나가면 더 적합한 결과를 제시할 수 있을 것이다.

시소러스 구축은 쉬운 일은 아니나 용어사전

으로부터 확장해나가면 어느 기록관이나 개발이 가능할 것이다. 특히 이용자의 사용용어를 반영하여 시소러스가 구축되고 검색시스템에 편입된다면, 기록물 검색의 어려움 해결에 많은 도움이 될 것이다. 앞으로 다양한 영역에서의 다양한 시소러스 개발을 기대한다.

그러나 구축 기관에서 잊지 말아야 할 것은 용어는 끊임없이 변화하고 기록문 생산환경도 끊임없이 변화한다는 사실이다. 즉 기록용어에 사용되고 이용자 검색으로 입력되는 용어도 계속 변화해갈 것이다. 그러므로 시소러스 구축은 한번 만들어 놓는다고 끝나는 것이 아니다. 지속적으로 이러한 기록명과 이용자 검색어의 변화를 반영하여 동적인 시소러스가 되도록 해야 할 것이다. 지속적인 업데이트가 시소러스의 기능을 100% 활용하도록 하는 데에 필수적임을 잊지 말아야 할 것이다.

참 고 문 헌

- 국가기록원 (2009). 국가기록원 시소러스 지침. NAK/A 8:2009(v1.1). 대전: 국가기록원.
- 김남희 (2005). 국회도서관 시소러스 구축과 활용 그리고 유지관리. 국회도서관보, 11, 36-49.
- 대통령기록관 (2013). 대통령기록물('08년 이관) 정리·기술 백서. 성남: 대통령기록관.
- 류주희 (2004). 문화콘텐츠 기반으로서의 한국역사용어 시소러스. 인문콘텐츠, 4, 199-211.
- 박부숙 (2007). 기록물 검색 서비스를 위한 시소러스의 개발과 활용. 기록인, 1, 56-57.
- 박상연, 임재익, 강주영 (2013). 공간정보사업의 중복사업 검색을 위한 의미기반검색 시스템의 설계. 한국IT서비스학회지, 12(3), 389-404.
- 박옥남 (2011). ISO25964를 통해 살펴본 시소러스 구축 표준에 관한 연구. 정보관리연구, 42(2), 43-65.
- 박지영 (2010). 시소러스를 연계한 문학류 패시 분류체계 개발. 한국비블리아학회지, 21(3), 77-89.
- 설문원 (2010). 기록 검색도구의 발전과 전망. 기록학연구, 23, 3-43.
- 손대형, 김태수 (1998). 패시분류체계를 이용한 시소러스작성에 관한 연구. 한국정보관리학회 제5회

학술대회논문집, 235-238.

- 유영준 (2008). 신학 용어 분류를 위한 패시 설계에 관한 연구. 한국문헌정보학회지, 42(3), 259-279.
- 이병길, 김희섭 (2013). 새마을운동 기록물의 개체기반 온톨로지 검색시스템 설계 및 평가. 한국기록관리학회지, 13(3): 67-97.
- 이승민 (2010). 시소러스 통합을 위한 개념기반 패시 프레임워크 구축. 한국도서관·정보학회지, 41(3), 269-290.
- 이윤령, 이해영 (2014). 온라인 검색도구의 개선 방안에 관한 연구. 한국기록관리학회지, 14(1), 75-100.
- 이정희, 김희섭 (2007). 대학 전자기록물을 위한 온톨로지 기반 검색시스템 설계 및 구현. 한국정보관리학회지, 24(3): 343-362.
- 이해영 (2011). 기록관 웹사이트 검색어의 분석과 활용. 한국기록관리학회지, 11(1), 93-112.
- 이해영, 이미영, 이은영, 이혁준, 이현정, 최영실, 박미자 (2008). 대학기록관 시소러스 구축 지침의 개발 연구. 한국기록관리학회지, 8(1), 189-210.
- 이해영 (2013). 대통령 주제분류체계 고도화 연구용역 최종보고서. 성남: 국가기록원 대통령기록관.
- 이혜원, 윤소영 (2011). 역사용어 온톨로지 모형 적용 방안 연구. 한국비블리아학회지, 22(1), 263-280.
- 정광훈, 남영준 (2013). 우리나라 대통령기록물의 분류기준에 관한 연구. 한국문헌정보학회지, 47(1), 419-448.
- 최윤경, 정연경 (2014). 한국 근대 여성 구술 기록물을 통한 시소러스 개발에 관한 연구. 한국기록관리학회지, 14(1), 7-24.
- 한국언론연구원 (1992). 신문기사 종합시소러스. 서울: 한국언론연구원.
- 황순희, 정한민, 성원경 (2006). 패시(Facet)을 이용한 과학기술분야 시소러스 구축과 활용방안. 정보관리연구, 37(3), 61-84.
- Aitchison, J., Gilchrist, A., & Bawden, D. (2000). Thesaurus construction and use: A practical manual. 4th ed. Psychology Press.
- Berners-Lee, T., Hendler, J., & Lassila, O. (2001). The semantic web. Scientific American, 284(5), 28-37.
- Garrod, P. (2000). Use of the UNESCO thesaurus for archival subject indexing at UK NDAD. Journal of the Society of Archivists, 21(1), 37-54.
- Giménez-Chornet, V. & Escrig-Giménez, M. (2011). Designing a thesaurus to give visibility to the historical archives in the Archivo del Reino in Valencia. Knowledge Organization, 38(2), 154-166.
- McNitt, W. H. (1989). Development of the PRESNET subject descriptor thesaurus. The American Archivist, 52(3), 358-364.
- Shearer, J. R. (2004). A practical exercise in building a thesaurus. Cataloging & Classification

Quarterly, 37(3-4), 35-56.

[관련법령]

「대통령기록물 관리에 관한 법률」 [법률 제10009호, 2010.2.4., 일부개정].

[웹사이트]

대통령기록관 (2014). 검색일자: 2014. 10. 2. <http://www.pa.go.kr/>

Ronald Reagan Presidential Library & Museum, White House Office Of Records Management (WHORM) Subject File, 1981-1989. Retrieved October 20, 2014, from <http://www.reagan.utexas.edu/archives/textual/whormsub/whormsub.html>

George Bush Presidential Library & Museum, White House Office of Records Management. Retrieved October 20, 2014, from http://bushlibrary.tamu.edu/research/finding_aids/whorm/

• 국문 참고자료의 영어 표기

(English translation / romanization of references originally written in Korean)

Choi, Yoon-Kyung & Chung, Yeon-Kyoung (2014). A Study on thesaurus development based on women's oral history records in modern korea. *Journal of Korean Society of Archives and Records Management*, 14(1), 7-24.

Hwang, Soonhee, Jung, Hanmin, & Sung, Won-Kyung (2006). Building thesaurus for science & technology domain using facets and its application to inference services. *Journal of information management*, 37(3), 61-84.

Jung, Kwang-Hun & Nam, Young-Joon (2013). Study on the classification guideline for the Korean presidential records. *Journal of the Korean Library and Information Science Society*, 47(1), 419-448.

Kim, Namhee (2005). National Assembly Library thesaurus construction, utilization, and maintenance. *National Assembly Library*, 11, 36-49.

Korea Newspaper Research Institute (1992). *Newspaper Article Comprehensive Thesaurus*. Seoul: Korea Newspaper Research Institute.

Lee, Byung Gil, & Kim, Heesop (2013). Design and evaluation of an individual instance-based ontology retrieval system for archival records of the "Saemaul Movement". *Journal of Korean Society of Archives and Records Management*, 13(3): 67-97.

- Lee, Hyewon & Yoon, So-Young (2011). A study on the model of history ontology: A focus on korean modern historical person. *Journal of the Korean Biblia Society for Library and Information Science*, 22(1), 263-280.
- Lee, Junghee & Kim, Heesop (2007). A design and implementation of ontology-based retrieval system for the electronic records of universities. *Journal of the Korean Society for information Management*, 24(3): 343-362.
- Lee, Seungmin (2010). Construction of the concept-based faceted framework for thesaurus integration. *Journal of the Korean Library and Information Science Society*, 41(3), 269-290.
- Lee, Yoon-Ryeong & Rieh, Hae-young (2014). A study on the improvement direction for online finding aids: based on the assessment of National Archives. *Journal of Korean Society of Archives and Records Management*, 14(1), 75-100.
- National Archives of Korea (2009). *Guide for thesauri of National Archives of Korea (1.1)*. Daejeon: National Archives of Korea.
- Park, Boo Sook (2007). Development and usage of thesauri for archive searching services. *Girok-in*, 1, 56-57.
- Park, Ok-Nam (2011). The study on the standards of thesaurus construction based on ISO 25964. *Journal of information management*, 42(2), 43-65.
- Park, Sangun, Lim, Jay Ick, & Kang, Juyoung (2013). Design of semantic search system for the search of duplicated geospatial projects. *Korea Society of IT Services Journal*, 12(3), 389-404.
- Park, Ziyoung (2010). Developing a faceted classification scheme integrated with a thesaurus for literature. *Journal of the Korean Biblia Society for Library and Information Science*, 21(3), 77-89.
- Presidential Archives (2013). *Presidential records (transferred '08) arrangement · description white paper*. Sungnam: Presidential Archives.
- 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.
- Rieh, Hae-young et al. (2013). *Final report of research project on the advancement of subject classification scheme for Presidential records*, Sungnam: National Archives of Korea, Presidential Archives.
- Rieh, Hae-young, Lee, Mi Young, Lee, Eun Young, Lee, Hyuk Joon, Lee, Hyun Jeoung, Choi, Young Sil, & Park, Mi Ja (2008). *Study on the development of guidelines for thesaurus*

- construction at university archives: Case study of Myongji University Archives Center. *Journal of Korean Society of Archives and Records Management*, 8(1), 189-210.
- Ryu, Ju-hee (2004). Korean history thesaurus as a basis of cultural content. *Humanities Content*, 4, 199-211.
- Seol, Moon-Won (2010). A study on development and prospects of archival finding aids. *The Korean Journal of Archival Studies*, 23, 3-43.
- Son, Dae-Hyung & Kim, Tae-Soo (1998). A study on thesaurus construction using facet classification. *Korea Society for Information Management*, 5, 235-238.
- Yoo, Yeong-Jun (2008). Establishing facet for classifying theological terms. *Journal of the Korean Library and Information Science Society*, 42(3), 259-279.