

뉴스 빅데이터를 이용한 우리나라 언론의 기록관리 분야 보도 특성 분석: 1999~2018 뉴스를 중심으로*

An Analysis of News Report Characteristics on Archives & Records Management for the Press in Korea: Based on 1999~2018 News Big Data

한승희 (Seunghee Han)**

초 록

이 연구에서는 1999년 1월부터 2018년 6월 현재까지 약 20년 간의 기록관리를 주제로 한 뉴스 빅데이터 4,680 건을 '빅카인즈'에서 추출하여, 이를 대상으로 우리나라 언론의 기록관리 주제에 대해 시계열 기반으로 보도 특성을 분석하고자 하였다. 먼저, 기록관리에 대한 언론 보도량의 차이를 살펴보기 위해 시기별, 주제별, 언론사 유형별 보도량을 분석하였다. 또한 기록관리 주제에 대한 언론 보도 내용의 차이에 대한 특성을 분석하기 위해 단어빈도 기반 내용 분석과 언어 네트워크 분석을 수행하여 언론 보도 내용의 시기별, 주제별, 언론사 유형별 차이를 분석하였다. 분석 결과, 기록관리 분야 뉴스 보도는 보도량과 보도 내용에 있어 시기별, 주제별, 언론사별로 차이가 있는 것으로 나타났다. 뉴스 보도량은 2007년 대통령기록물관리법이 제정된 이후부터 증가하기 시작하여 2013년에 가장 많은 뉴스가 보도된 것으로 나타났으며, 정치와 사회 주제를 중심으로 중앙지와 경제지가 가장 많은 양의 뉴스를 보도한 것으로 나타났다. 또한 뉴스 보도 내용의 분석 결과, 기록관리가 도입된 처음 10년 동안은 기록관리의 현장 적용과 확산 과정에서 발생하는 이슈들을 중심으로 뉴스 주제가 형성되다가, 대통령기록물관리법 제정 이후로 기록관리가 정치적, 사회적 이슈의 주요 요인이 되면서 정치, 사회 분야의 뉴스가 많이 보도된 것으로 나타났다.

ABSTRACT

The purpose of this study is to analyze the characteristics of Korean media on the topic of archives & records management based on time-series analysis. In this study, from January, 1999 to June, 2018, 4,680 news articles on archives & records management topics were extracted from BigKinds. In order to examine the characteristics of the media coverage on the archives & records management topic, this study was analyzed to the difference of the press coverage by period, subject, and type of the media. In addition, this study was conducted word-frequency based content analysis and semantic network analysis to investigate the content characteristics of media on the subject. Based on these results, this study was analyzed to the differences of media coverage by period, subject, and type of media. As a result, the news in the field of records management showed that there was a difference in the amount of news coverage and news contents by period, subject, and type of media. The amount of news coverage began to increase after the Presidential Records Management Act was enacted in 2007, and the largest amount of news was reported in 2013. Daily newspapers and financial newspapers reported the largest amount of news. As a result of analyzing news reports, during the first 10 years after 1999, news topics were formed around the issues arising from the application and diffusion process of the concept of archives & records management. However, since the enactment of the Presidential Records Management Act, archives & records management has become a major factor in political and social issues, and a large amount of political and social news has been reported.

키워드: 뉴스 빅데이터, 시계열 내용 분석법, 언어 네트워크 분석, 내용 분석법, 빅카인즈, 기록관리, 빅데이터 분석, 대통령기록물관리법, 대통령기록물
news big data, time-series content analysis, semantic network analysis, content analysis, BigKinds, archives & records management, big data analysis, Presidential Records Management Act, presidential records

* 이 논문은 2018학년도 서울여자대학교 연구년연구비의 지원을 받았음.

** 서울여자대학교 사회과학대학 문헌정보학과 부교수(hanshee@swu.ac.kr)

■ 논문접수일자: 2018년 8월 18일 ■ 최초심사일자: 2018년 9월 17일 ■ 게재확정일자: 2018년 9월 19일
■ 정보관리학회지, 35(3), 41-75, 2018. [http://dx.doi.org/10.3743/KOSIM.2018.35.3.041]

1. 연구의 목적 및 필요성

정보기술의 발전으로 데이터의 양이 급속도로 증가함에 따라 이를 저장, 처리하는 기술 역시 빠르게 발전하고 있다. 빅데이터(big data)는 현재 단순히 많은 양의 데이터라는 개념을 넘어 데이터를 수집, 저장, 분석함으로써 현상을 이해하고 나아가 미래를 예측하는 기술을 통칭하며, 오늘날 우리 사회 전반에 활용되고 있다. 더구나 초연결사회를 살고 있는 요즘, 연결된 모든 것으로부터 생성되는 데이터는 기하급수적으로 증가하고 있으므로, 이를 처리, 해석하고 적용하기 위한 빅데이터 기술은 앞으로도 계속 사회적으로 관심을 받게 될 것이다.

이러한 빅데이터의 발전으로 학술 연구 분야에서도 빅데이터와 그 처리 기술을 개발하고 적용하는 연구가 증가하고 있다. 빅데이터 활용 연구는 주로 대용량 데이터의 처리를 통해 그간의 표본 기반 연구에서는 얻을 수 없었던 데이터에 숨겨진 의미를 해석하는 것에 의의를 두고 있다. 여러 학술 분야에서 빅데이터를 활용한 연구가 진행되어 왔으나, 최근에는 뉴스 빅데이터를 활용하여 특정 주제에 대한 사회적 이슈를 분석하는 내용 분석 연구가 학계의 관심을 받고 있다. 뉴스 자료는 시계열로 축적되어 온 양질의 데이터 집합이다. 과거에는 처리 시간과 정교함의 문제로 소규모 뉴스 기사 컬렉션을 대상으로만 내용 분석 연구를 수행하였으나, 최근 자연어 처리 기술과 텍스트 마이닝 등 관련 기술의 발전으로 인해 컴퓨터를 이용해 많은 양의 기사를 처리하여 분석하는 일이 과거에 비해 수월해졌다(박대민, 2016). 그리고 이러한 기술의 발전은 뉴스 데이터를 활용

한 학술 연구 분석의 활성화를 가능하게 하고 있다.

한편, 지식정보사회에서 지식, 정보, 데이터, 정보, 도서관, 기록관 등과 같은 문헌정보학과 기록관리학의 핵심 요소는 다양한 사회적 현상에 크고 작은 영향을 주고 받는 개념으로 인식되어 왔다. 그렇기 때문에 이러한 주제와 관련된 이슈를 다루고 있는 뉴스를 심심치 않게 접할 수 있다. 특히 기록관리 분야는 1999년에 처음으로 관련 법이 제정된 이후로 몇 가지 정치, 사회적 이슈의 중심이 되어 언론계에 많은 관심을 받아 왔다. 이러한 현상은 기록관리 분야가 사회적으로 널리 알려지는 데 기여했으나, 한 편으로는 기록관리가 정치적 사안들과 쉽게 독립하기 어려운 구조적 문제를 양산하여 우리나라 기록관리의 독립성 문제를 야기하였다.

이러한 관점에서, 이 연구에서는 1999년 이후로 약 20년 간의 뉴스 빅데이터를 대상으로 언론에 보도된 기록관리 분야의 보도 특성을 분석하는 것을 목적으로 한다. 분석의 방법은 양적 분석과 질적 분석 모두를 포함한다. 그 동안 문헌정보학과 기록관리학 분야에서 해당 학문 분야의 발전 과정 및 동향 변화를 분석하기 위해 몇몇의 시계열 분석 연구가 있으나, 뉴스 기사를 활용한 시계열 분석한 것은 이 연구가 거의 처음이다.

2. 이론적 배경

2.1 뉴스 빅데이터

빅데이터는 '데이터의 양이 방대하여 기존의

방법이나 도구로 데이터의 수집, 저장, 분석 등이 어려운 정형 및 비정형 데이터'를 말한다(국립중앙과학관, 2018). 지금 이 순간에도 다양한 분야에서 다양한 유형의 데이터가 대규모로 축적되고 있으며, 뉴스 데이터 역시 예외가 아니다. 전통적인 매체와는 별개로 온라인 전용 매체가 증가하면서 축적되는 뉴스 데이터의 양도 기하급수적으로 증가하고 있다. 카카오(2015)는 2015년 9월 현재 140 여 개 매체에서 하루 2~3만 건의 기사가 송고되고 있다고 밝혔다. 또한 뉴스 데이터는 전통적으로 신문 기사를 중심으로 한 텍스트 형식이 주를 이루었으나, 멀티미디어 기술의 발전으로 인해 현재는 한 기사 내에 텍스트, 이미지, 동영상 등 다양한 형식의 데이터가 결합되어 있다.

뉴스 빅데이터는 비정형 데이터를 자연어 처리, 영상 처리, 의미연결망 분석을 통해 인명, 기관, 수치, 인용문 등 정형화된 메타데이터로 추출한 결과물을 말하며(신성현, 2016), 다음과 같은 특성을 가지고 있다(박대민, 2013).

첫째, 뉴스 빅데이터에는 텍스트뿐만 아니라 멀티미디어 데이터가 포함되어 있으나, 현재까지 텍스트 중심의 기사가 차지하는 비중이 더 크고, 멀티미디어 데이터의 경우에도 텍스트로 변환·추출되어 처리하는 것이 일반적이기 때문에 데이터의 규모에 비해 처리해야 할 데이터의 용량은 다른 유형의 빅데이터에 비해 적은 편이다.

둘째, 일반적으로 뉴스 빅데이터는 텍스트 형태의 기사 데이터 자체가 갖고 있는 비정형적 성격뿐만 아니라, 날짜, 매체, 면중, 장르 등 기사 메타데이터가 갖고 있는 정형화된 성격을 모두 갖추고 있어 완전히 비정형적인 데이터보

다는 비교적 처리와 관리가 비교적 용이하다.

셋째, 뉴스 기사는 많은 경우 중복되므로 축약 효과가 크다. 언론사는 기자회견이나 보도자료 등 공유된 정보를 바탕으로 기사화하는 경우가 많으므로, 유사한 내용의 기사가 중복되어 생산되는 경우가 일반적이다. 그러므로 여러 중복 기사들을 하나의 대표 기사로 대신하여 처리할 데이터의 양을 줄일 수도 있고, 다른 한 편으로는 여러 기사들을 중복되지 않은 사실 위주로 종합한 종합 기사를 만들어 서비스를 제공하는 것도 가능하다.

이러한 이유로 관련 업계에서는 뉴스 빅데이터를 활용하여 다양한 서비스를 제공하고 있다. 관련 기사를 묶어서 중요도 순으로 보여주는 구글 뉴스(Google News)나 여러 뉴스를 종합, 요약하여 보여주는 야후 뉴스룸(Yahoo Newsroom) 등과 같은 뉴스 어그리게이션(news aggregation) 서비스 이외에도 클러스터링, 뉴스 큐레이션, 오피니언 마이닝과 감성분석, 기사 시각화 등 다양한 기술이 뉴스 빅데이터에 적용되고 있다.

뉴스 빅데이터 시스템은 비정형 뉴스를 정형화하여 제공함으로써 뉴스 기사를 쉽게 활용할 수 있도록 한다. 우리나라에서는 2013년 서울대 차세대융합기술연구원에서 선보인 '뉴스소스 베타'를 시작으로 하여 현재는 '빅카인즈(Big Kinds, <https://www.bigkinds.or.kr/>)'가 가장 대표적인 뉴스 빅데이터 시스템으로 손꼽히고 있다.

빅카인즈 서비스는 한국언론진흥재단에서 1990년부터 시작한 'KINDS(Korea Integrated News Database System)'에서 진화한 새로운 뉴스 분석 서비스이다. 2018년 6월 현재 종합일간지, 경제지, 지역일간지, 방송사 등 국내 48개의 언

론사로부터 뉴스를 수집하여 통합 데이터베이스를 구축하고 있으며, 여기에 빅데이터 분석 기술을 접목하여 뉴스 분석 서비스를 제공하고 있다(한국언론진흥재단, 2018). 2016년에 첫 선을 보인 이후 2018년 8월 현재 5천 2백만 건이 넘는 데이터를 대상으로 제공하고 있는 이 서비스는 기존의 뉴스 데이터베이스에서 제공했던 단순한 기사 검색에서 벗어나, 뉴스 속 인물 및 기관 간 관계 네트워크 분석, 뉴스 인용문 분석, 뉴스 키워드의 연관어 분석 및 트렌드 분석 등 심층적인 뉴스 데이터 분석 서비스를 제공하고 있다(한국언론진흥재단, 2018). 빅카인즈에서 뉴스 기사를 제공하는 언론사는 <표 1>과 같다.

빅카인즈의 시스템 구성은 <그림 1>과 같다. 먼저 뉴스수집시스템에서는 매일 48개 언론사

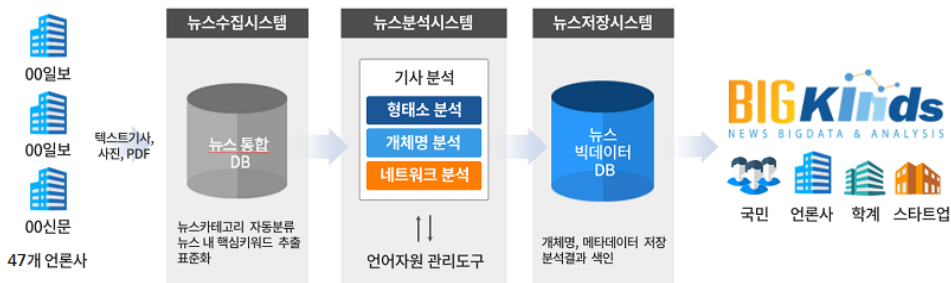
의 뉴스 콘텐츠를 자동으로 수집, 분류하며, 뉴스분석시스템에서는 분류된 콘텐츠에 대해 자연어 처리와 네트워크 분석을 수행하여 이용자에게 제공한다. 그리고 뉴스저장시스템에서는 뉴스 콘텐츠에서 처리된 메타데이터, 색인 등을 저장, 관리한다.

2.2 뉴스 빅데이터 분석

데이터의 규모가 커지고 데이터 처리 기술이 발전하면서 연구자들은 기존의 소규모 데이터로는 알아내기 어려웠던 여러 현상을 빅데이터 분석을 통해 파악할 수 있게 되었다. 이러한 빅데이터 분석 연구는 표본 분석을 통한 통계적 유추의 한계를 넘어 원 데이터를 직접 분석할

<표 1> 빅카인즈에서 뉴스 기사를 제공하는 유형별 언론사

유형 (수)	언론사 (가나다 순)
중앙지 (8)	경향신문, 국민일보, 내일신문, 문화일보, 서울신문, 세계일보, 한겨레, 한국일보
경제지 (7)	매일경제, 머니투데이, 서울경제, 아시아경제, 파이낸셜뉴스, 한국경제, 헤럴드경제
지역지 (27)	강원도민일보, 경기일보, 경남도민일보, 경남신문, 경상일보, 경인일보, 광주매일신문, 광주일보, 국제신문, 대구일보, 대전일보, 매일신문, 무등일보, 부산일보, 영남일보, 울산매일, 전남일보, 전북도민일보, 전북일보, 제민일보, 중도일보, 중부매일, 중부일보, 충북일보, 충청일보, 충청투데이, 한라일보
방송사 (4)	MBC, OBS, SBS, YTN
전문지 (2)	디지털타임스, 전자신문



<그림 1> 빅카인즈 서비스 시스템 구성도

(출처: 빅카인즈 사용자 매뉴얼, 2018)

수 있고, 알고리즘에 의한 자동 처리를 통해 연구자의 개입으로 인한 데이터 정리 및 해석 과정에서의 오류를 최소화 할 수 있는 장점이 있다. 특히 최근에는 뉴스 자료가 아카이브화 되고 비교적 쉽게 구해서 이용할 수 있는 자연어 처리 및 내용 분석 도구들이 등장하면서 대량의 뉴스 기사를 분석하는 것에 대한 연구자들의 관심이 높아지고 있다. 이러한 관점에서 자연어 처리와 의미연결망 분석이 결합된 자동화된 뉴스 내용 분석 방법을 뉴스 빅데이터 분석이라 한다(박대민, 2016).

2.2.1 뉴스 빅데이터의 시계열 내용 분석

사회 현상과 그를 둘러싼 이슈를 분석하기 위한 다양한 연구방법론 중 시계열 내용 분석(time-series content analysis)은 시간의 흐름에 따라 축적된 시계열 데이터를 대상으로 내용 분석을 수행하는 방법이다. 일반적으로 내용 분석(content analysis)에서는 문헌이나 설문, 인터뷰, 관찰 등을 통해 얻어진 텍스트를 대상으로 이에 출현한 키워드의 빈도를 측정 한 후, 이를 통해 내용을 해석하고, 맥락을 이해하여 상위 차원으로 범주화함으로써 연구 문제에 대한 해답을 얻는다. 시계열 데이터를 대상으로 내용 분석을 수행하는 경우에는 수집된 데이터 전체에 대한 내용 분석과 함께 전체 데이터를 특정 기준 시점으로 나누어 반복적으로 분석하고 그 결과를 서로 비교하거나 다른 시계열 데이터와 비교할 수 있다(박대민, 2016).

일반적으로 시계열 데이터는 오랜 기간 동안 축적된 대규모 데이터의 성격을 갖고 있으므로 수작업에 의존하는 전통적인 내용 분석 방법으로 처리하기에 한계가 있다. 이를 해결하기 위

해 시계열 데이터에 빅데이터 분석 방법을 적용할 수 있다. 특히 뉴스 기사는 당시의 사회적 이슈를 반영하고 있으며 오랜 시간 동안 일 단위로 축적된 방대한 양의 데이터로, 시계열 내용 분석의 대상으로 적합하다. 그러므로 뉴스 빅데이터를 활용하면 어떠한 현상에 대해 시계열을 기반으로 그 담론이나 논조, 동향 등을 비교·분석할 수 있다.

이은별, 전진오, 백지선(2017)은 국내 주요 언론이 서울 내 외국인 마을을 보도하는 방식을 분석하여 특정 외국인 거주 지역에 대한 고정관념을 살피기 위해 1990년부터 2016년까지 보도된 뉴스 중 서래마을, 동부이촌동, 구로동을 주제로 보도한 3천 여 건의 기사를 '빅카인즈'에서 추출한 후, 시간의 흐름에 따른 각 마을의 뉴스 주제별 보도량의 차이를 분석하였다. 분석 결과, 언론이 대중에게 외국인 거주 마을에 대한 특정한 이미지를 강조하거나 고착시킬 가능성이 있다는 점을 확인하였다.

성지영, 김성균(2017)은 언론에 나타난 덕수궁 돌담길의 장소 담론을 해석하기 위해 '빅카인즈'를 이용하여 1966년부터 2017년까지 3천 여 건의 뉴스 기사를 수집하여 시간의 흐름에 따른 주제별 보도량 추이를 분석하고, 덕수궁 돌담길의 장소 담론을 설명하기 위해 '빅카인즈'에서 제공하는 연관어 데이터의 시기별 추이 변화를 분석하고 이들을 범주화하여 해석하였다.

문헌정보학/기록관리 주제를 대상으로 한 시계열 분석 연구는 일반적으로 학술지를 대상으로 한 지적 구조 및 연구 동향의 변화를 분석하기 위한 목적으로 수행되었으나(김희정, 2006; 박준형, 류범모, 오효정, 2018; 서은경, 유소영,

2013; 유사라, 2010; 이재운, 문주영, 김희정, 2007; 최형욱, 최예진, 남소연, 2018), 뉴스 빅데이터를 활용한 시계열 분석 연구는 찾아보기 어려웠다.

2.2.2 뉴스 빅데이터 분석 방법론

일반적으로 빅데이터를 대상으로 한 내용 분석에 많이 쓰이는 기법에는 토픽 모델링(topic modeling)과 언어 네트워크 분석(semantic network analysis)이 있다.

토픽 모델링은 방대한 양의 텍스트에 출현한 단어 중 유사한 의미의 단어를 클러스터링하여 텍스트에 숨어 있는 주제를 추론하고, 이들 주제 간의 관계와 시간에 따른 변화 추이를 분석하는 통계적 방법을 말한다(Blei, 2012). 이러한 특징으로 인해 토픽 모델링은 텍스트를 활용하여 특정 분야의 내용이나 동향을 분석하기 위한 방법론으로 다양한 분야에서 사용되어 왔다(강범일, 송민, 조화순, 2013).

언어 네트워크 분석은 의미 연결망 분석이라고도 하는데, 사회 네트워크 분석(social network analysis)을 응용한 내용 분석의 한 방법(이수상, 2012)으로, 앞에서 언급한 전통적인 내용 분석 기법의 한계점을 극복하기 위해 제안되었다(박한우, Leydesdorff, 2004). 텍스트 내 출현한 단어들은 함께 출현한 다른 단어들과 조합되어 특별한 의미 체계를 만드는 경향이 있다는 가정 하에, 전통적인 내용 분석에서 개별 단어의 출현 빈도에만 초점을 맞추는 것과 달리 언어 네트워크 분석에서는 단어 간의 동시 출현빈도(co-occurrence)를 바탕으로 의미 네트워크 행렬을 만들어서 텍스트 내에서 차지하는 단어의 강도와 단어 간의 관계를 시각화하

여 표현한다(김수연, 김대욱, 최명일, 2013; 장정우, 최경호, 2012).

강범일, 송민, 조화순(2013)은 토픽 모델링 기법을 이용하여 18대 대선 직전 2년 동안 총 4종의 진보, 보수매체를 대상으로 3명의 대통령 후보가 실린 기사에서 10개의 토픽이 시기별로 어떻게 분포, 변화하는지를 분석하였다. 이수상(2017) 역시 토픽 모델링 기법을 이용하여 신문기사에 실린 2016년 경주 지진 사건에 대해 관련 토픽의 유형, 사건 발생 이후 약 3개월 간 토픽의 변화 양상 등을 분석하였다. 박승정 외(2017)도 토픽 모델링 기법을 이용하여 박근혜 정부 4년간 박근혜 관련 신문 사설을 분석하여 23개의 토픽으로 분류하고, 토픽별 보도량을 살펴봄으로써 진보 및 보수 매체의 뉴스 이슈에 대한 소유권을 비교 분석하는 연구를 수행하였다. 이들 연구들은 토픽 모델링 기법을 이용하였고, 분석 기간이 상대적으로 길지 않다는 공통점을 가지고 있다.

최윤정, 권상희(2014)는 2011년부터 2013년까지 4개 매체를 대상으로 용어 '빅데이터'를 포함하고 있는 신문 기사를 수집하여 언어 네트워크 분석을 실시하였는데, 분석 결과, '빅데이터'라는 용어가 처음 출현한 2011년에는 보도의 의미 구조가 중립적 성격이었으나, 2012년부터 긍정적인 방향으로 자리매김하고 있으며, 시간이 지날수록 빅데이터의 활용 가치에 관심이 많아지고 있음을 확인하였다. 김봉제(2018)는 의미 네트워크 분석을 통해 2000년부터 2016년까지 뉴스 보도에 나타난 '도덕'과 '윤리' 용어의 사용 특성을 분석하였는데, 뉴스 텍스트에 대한 의미 네트워크 분석을 통해 뉴스의 의도와 흐름을 파악할 수 있음을 확인하였다. 또한 박

대민(2016)은 1990년부터 2015년까지 8개 중앙지의 정치와 사회면 기사 약 100만 건에 대해 기사 정보원과 인용문 주제를 중심으로 언어 네트워크 분석을 실시하여 매체별로 정보원과 주제의 시계열적 변화를 살펴보았다.

토픽 모델링은 일반적으로 수학적으로 계산 복잡도가 높기 때문에 복잡한 프로그래밍 기법이 요구된다. 그러나 언어 네트워크 분석은 토픽 모델링에 비해 상대적으로 쉬운 수학적 개념을 바탕으로 하고 있으며, 온라인에 공개된 무료 소프트웨어를 활용하여 쉽게 처리할 수 있다. 또한 언어 네트워크 분석에서는 단어 간의 관계를 시각적으로 표현함으로써 직관적으로 단어 간의 의미를 파악할 수 있는데, 이는 토픽 모델링 기법이 갖지 못하는 장점이라 할 수 있다(박대민, 2016).

3. 연구 방법

3.1 연구 문제

이 연구에서는 기록관리 주제에 대한 언론 보도의 특성을 양적 분석과 내용 분석을 통해 살펴보고자 한다. 이를 위해 설정된 연구문제는 다음과 같다.

- 연구문제 1: 기록관리에 대한 언론 보도량은 차이가 있는가?
- 1-1: 기록관리에 대한 언론 보도량은 시기별로 차이가 있는가?
- 1-2: 기록관리에 대한 언론 보도량은 주제별로 차이가 있는가?

- 1-3: 기록관리에 대한 언론 보도량은 언론사 유형별로 차이가 있는가?

이 연구에서는 먼저 1999년 기록관리 관련 법령이 처음 제정된 이후로 생산된 기록관리 주제 뉴스 기사의 양적 분석을 통해 기록관리에 대한 보도가 시기별로, 주제별로, 언론사 유형별로 어떠한 차이를 보이는지를 분석하고자 한다.

- 연구문제 2: 기록물 관리에 대한 언론 보도 내용은 차이가 있는가?
- 2-1: 기록관리에 대한 언론 보도 내용은 시기별로 차이가 있는가?
- 2-2: 기록관리에 대한 언론 보도 내용은 주제별로 차이가 있는가?
- 2-3: 기록관리에 대한 언론 보도 내용은 언론사 유형별로 차이가 있는가?

또한 이 연구에서는 기록관리 주제를 다루는 뉴스 기사의 내용 분석을 통해 뉴스 기사의 내용이 시기별, 주제별, 언론사 유형별로 어떠한 차이를 나타내는지를 분석하고자 한다.

3.2 분석 대상

기록관리에 대한 언론 보도 분석을 위해서는 일정 기간 동안의 뉴스 기사가 필요하다. 이 연구에서는 '빅카인즈'를 활용하여 국내에서 생산된 기록관리 주제의 뉴스 기사를 수집하였다. 이 연구에서는 언론사 유형별 보도 특성 분석을 위해 중앙지를 포함한 모든 유형의 언론사 뉴스를 모두 수집하였다. 빅카인즈에 기사를 제공하지 않고 독자적인 데이터베이스

를 구축하고 있는 3대 언론(조선일보, 중앙일보, 동아일보)은 분석 대상에 포함되지 않았다. 그러나 방송사를 포함해 더 많은 수의 뉴스 데이터를 수집하여 분석함으로써 데이터 타당도를 높이고자 하였다.

뉴스 기사를 수집하기 위해서는 검색 질의어와 수집 기간을 결정해야 한다. 그 동안 기록관리와 관련된 다양한 세부 이슈가 언론에 노출되어 왔으나, 이 연구에서는 기록관리에 대한 일반적 보도 특성을 분석하기 위해 검색 주제를 '기록관리'로 한정하였다. 일반적으로 학계에서는 '기록관리'가 학문 분야 및 연구 주제를 나타내는 표준적 용어로 수용되어 있다. 그러나 빅카인즈에서 '기록 관리'로 검색하는 것보다는 '기록물 관리'로 검색했을 때 더 많은 검색 결과를 얻을 수 있었는데, 이것은 언론 보도에서 '기록물 관리'라는 표현을 더욱 보편적으로 사용하고 있다는 것을 의미하므로, 최종적으로 질의어를 '기록물 관리'로 수정하여 검색하였다.

데이터 수집 기간은 우리나라 최초의 본격적인 기록관리 관련 법이라 할 수 있는 '공공기관의 기록물 관리에 관한 법률'이 공포된 1999년 1월에서부터 2018년 6월 30일까지 약 20년 간으로 하였다. 이 연구에서는 오랜 기간을 대상으로 많은 수의 뉴스 기사를 분석함으로써 기존의 신문 기사를 대상으로 한 내용 분석 연구와 차별화하고자 하였다. 이러한 연구를 통해 뉴스 보도 성향의 거시적 지형을 보는 새로운 시도가 가능하다(이은별, 전진오, 백지선, 2017).

앞에서 언급한 바와 같이 검색 질의어와 수집 기간을 결정하여 이 연구에서는 최종적으로 4,680건의 뉴스 기사를 수집, 분석 대상으로 하였다. '빅카인즈'에서는 뉴스 검색결과와 메타

정보를 다운로드하여 빅데이터 분석에 이용할 수 있도록 하고 있다. 메타정보는 뉴스식별자, 일자, 언론사, 기고자, 제목, 주제 분류, 사건/사고 분류, 인물, 위치, 기관, 키워드, 특성추출, 본문, URL 등을 포함하며, 주제 분류는 3개까지, 사건/사고 분류는 2개까지 중복 분류를 허용하고 있다. 이 연구에서는 검색 결과를 통해 얻은 메타정보에서 키워드, 주제, 언론사, 일자, 제목 등을 추출하여 보도 특성 분석에 활용하였다.

3.3 분석 절차 및 방법

이 연구에서 진행된 전체적인 데이터 분석 절차는 <그림 2>와 같다. 먼저, 3.1에서 제시한 연구 문제를 해결하기 위해 이 연구에서는 빅카인즈에서 수집된 뉴스 데이터를 대상으로 두 가지 방법으로 연구 데이터를 분석한다. 하나는 보도량에 대한 시기별, 주제별, 언론사 유형별 통계적 분석이고 다른 하나는 내용 분석이다. 보도 내용 분석은 다시 두 가지 방법으로 나누어 진행되는데, 하나는 전통적인 내용 분석 연구에서 자주 사용되는 기사에 출현한 키워드에 대한 빈도 분석이고, 다른 하나는 언어 네트워크 분석이다.

먼저 출현 단어의 빈도 분석에 기반한 내용 분석은 '빅카인즈'의 검색 결과의 메타정보에서 제공하는 키워드를 추출한 후 전체 기사를 대상으로 하여 단어의 출현이 나타내는 특징을 먼저 분석한 후, 세부적으로 시기별, 언론사 유형별로 단어의 출현 빈도를 분석하여 내용 분석을 실시하였다. 또한 언어 네트워크 분석에서는 전체 뉴스 기사 데이터 중 제목 데이터만 추출하여 시기별, 주제별로 나누어 언어 네트워크를

자동 생성하고, 이를 대상으로 내용 분석을 실시하였다. 키워드 전체를 대상으로 하지 않고 기사 제목만을 대상으로 언어 네트워크를 분석한 이유는 뉴스 기사에서 제목이 주는 정보량이 매우 크기 때문이다. 이 연구에서 언어 네트워크 분석은 다음과 같은 5단계로 진행된다.

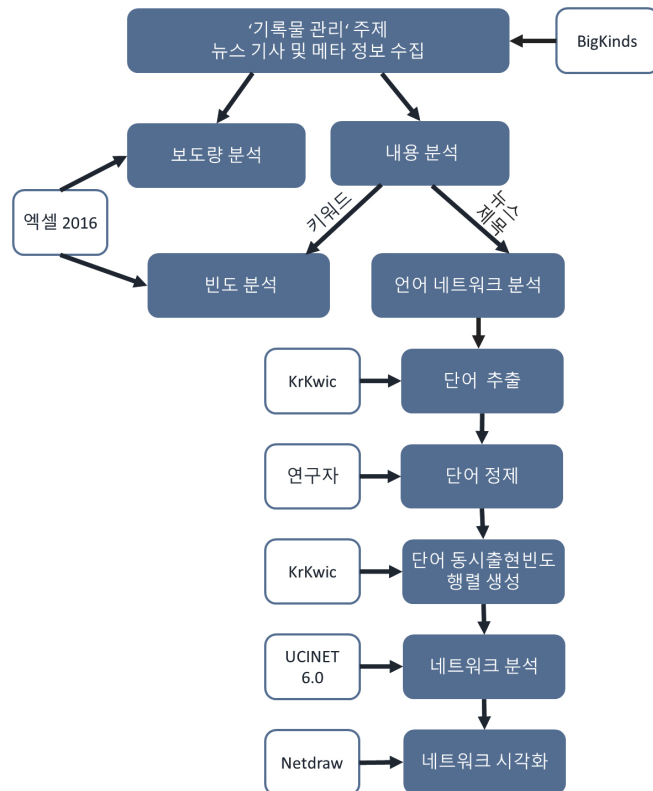
먼저, 추출된 뉴스 제목을 대상으로 단어와 빈도를 추출하는데, 이 때 KrKwic을 사용한다. KrKwic은 한국어 텍스트의 내용 분석을 위해 박한우, Leydesdorff(2004)에 의해 개발된 소프트웨어인데, KrText, KrTitle, KrWords 등의 하위 프로그램으로 구성되어 있다. 제목 텍스트로부터 단어와 빈도의 추출은 KrText 프

로그램을 통해 수행한다.

둘째, 추출된 단어와 빈도를 이용하여 동시출현빈도 행렬을 만들기 전에 의미 있는 네트워크 분석 결과를 얻기 위해 불용어 제거, 유사어 처리, 일정 빈도 이상의 핵심어 추출 등 단어 정제 작업이 연구자에 의해 진행된다.

셋째, KrKwic의 하위 프로그램인 KrTitle을 이용하여 일정 빈도 이상의 정제된 핵심어를 대상으로 단어 동시출현빈도 행렬을 작성한다.

넷째, UCINET 6.0(Borgatti, Everett, & Freeman, 2002)을 이용하여 앞서 생성된 단어 동시출현 빈도 행렬을 대상으로 네트워크 분석을 수행한다.



<그림 2> '기록물 관리' 주제 뉴스 기사의 분석 절차와 활용 소프트웨어

마지막으로, Netdraw(Borgatti, 2002)를 이용하여 네트워크 분석 결과를 시각화한다.

4. 연구 결과

4.1 보도량 분석 결과

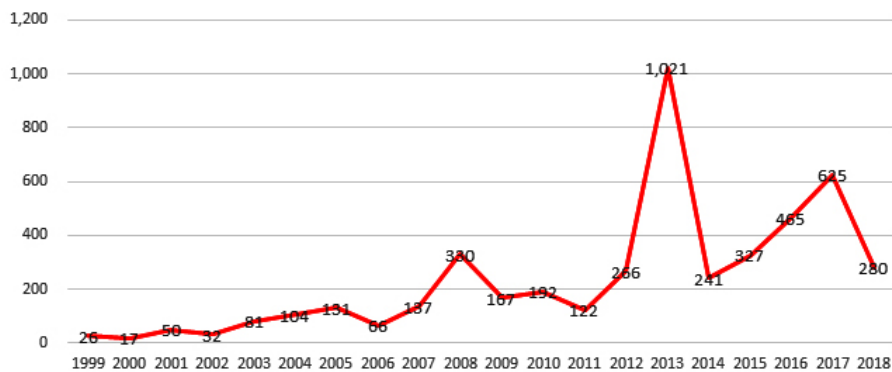
‘기록물 관리’에 대한 보도량 추이를 분석하기 위해 연도별 기사 건 수를 조사한 결과는 <그림 3>과 같다. 최초로 기록관리 관련 법이 공포된 1999년부터 시작하여 2006년까지는 기록물 관리와 관련된 기사 건 수가 매우 적었으나, ‘대통령기록물 관리에 관한 법률’이 제정된 2007년을 시작으로 보도량이 증가하기 시작하였다. 노무현 전 대통령의 봉하 이지원(e-지원) 반출 논란이 있었던 2008년에도 보도량이 증가한 후, 참여정부 시절 작성된 남북정상회담 회의록 유출 논란이 있었던 2013년에 보도량이 폭발적으로 증가한 것으로 나타났다. 그리고 이후 2014년에 감소한 뉴스 건 수는 이후 차츰 증가세를 보이고 있다. 2018년의 경우 2017년에 비해 보

도량이 감소한 것으로 보이나, 6월까지의 뉴스만을 대상으로 하였으므로 증가의 여지가 있다고 할 수 있다. 이를 통해 대통령기록물과 관련된 이슈가 기록물 관리 관련 뉴스의 증가에 영향을 미친 것으로 유추할 수 있다.

4.1.1 시기별 보도량

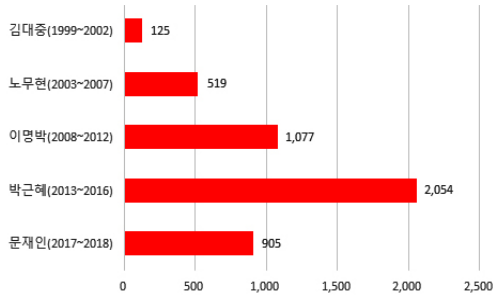
전체 데이터 수집 기간을 임의의 시기로 나누어 시계열 분석을 시도할 수 있다. 이 연구에서는 대통령의 임기를 기준으로 전체 데이터 수집 기간인 20년을 5개의 시기로 나누어 시기별 뉴스 보도량을 분석하였다. 탄핵으로 인해 파면을 맞은 박근혜 전 대통령과 조기 대선으로 대통령에 선출된 문재인 대통령을 제외하고 일반적으로 새 대통령의 임기는 2월 말부터 시작하기 때문에, 편의상 연 단위로 구분하여 대통령 임기를 나누고 이를 기준으로 시기별 뉴스 보도량을 분석하였다.

<그림 4>에서 보는 바와 같이, 기록관리 도입 초기인 김대중 전 대통령 임기에 가장 적은 뉴스가 보도되었으나, 이후 지속적으로 보도량이 증가하여, 박근혜 전 대통령 임기인 2013년~



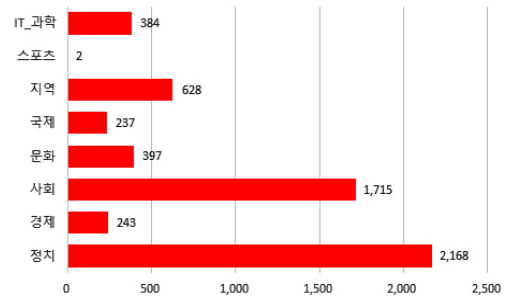
<그림 3> ‘기록물 관리’ 주제의 연간 보도량 변화 추이

2016년에 가장 많은 뉴스가 보도되었다.



〈그림 4〉 시기별 뉴스 보도량

이어 '사회'와 '지역'이 높은 비중을 차지하는 것을 알 수 있다. '스포츠' 주제에서는 단 2건의 뉴스를 보도하였다.

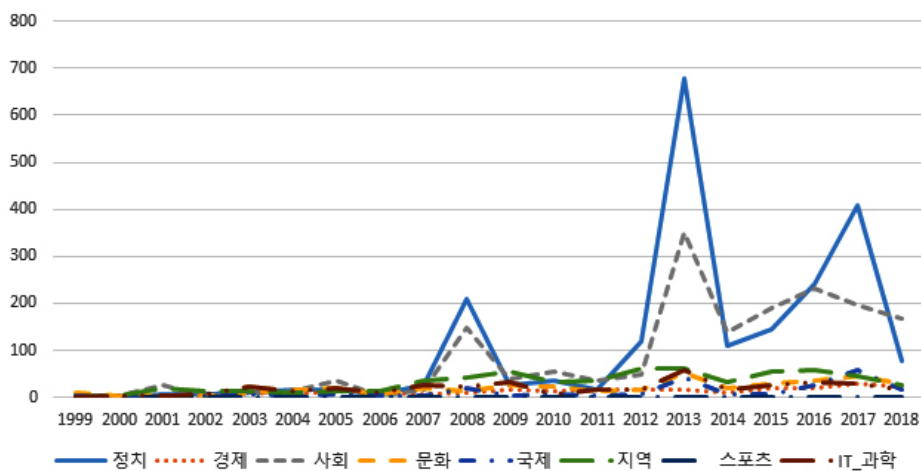


〈그림 5〉 주제별 뉴스 보도량

4.1.2 주제별 보도량

'빅카인즈'에서는 뉴스 기사를 8개의 주제로 분류하고 있는데, 한 기사 당 최대 3개의 주제까지 중복 분류를 허용하고 있다. 이 연구에서는 '빅카인즈'에서 제공하고 있는 뉴스 분류 메타정보를 이용하여 주제별 보도량을 분석하고자 하였다. 〈그림 5〉에서 보는 바와 같이 기록관리 관련 뉴스에서 가장 많은 비중을 차지하는 주제는 '정치'인 것으로 나타났으며, 그 뒤를

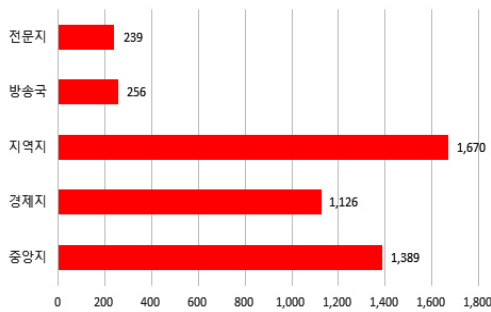
주제별 보도량에 대한 시계열 분석을 위해 연간 주제별 뉴스 보도량 추이를 분석하였으며, 그 결과는 〈그림 6〉과 같다. 전체적인 추이는 〈그림 3〉의 연간 보도량 변화 추이와 유사한 것으로 나타났다. 특히 주제 분류 비중이 높은 '정치'와 '사회' 주제의 보도량 변화 추이가 연간 보도량의 변화 추이와 유사한 것을 알 수 있다.



〈그림 6〉 연간 주제별 뉴스 보도량 변화 추이

4.1.3 언론사 유형별 보도량

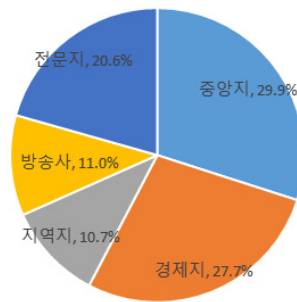
‘빅카인즈’에서는 언론사의 유형을 크게 5가지로 나누고 있다. 이 연구에서는 언론사 유형별 기록관리 분야 뉴스 보도량을 분석하고자 하였다. <그림 7>에서 보는 바와 같이, 가장 많은 뉴스를 보도한 언론사 유형은 지역지이며, 그 다음으로 중앙지가 많은 양의 뉴스를 보도한 것으로 나타났다. 그러나 <표 1>에서 보는 바와 같이, 유형별 언론사의 수가 달라 절대적인 뉴스 보도량으로는 실제 언론사 유형별 보도량을 알 수 없다.



<그림 7> 언론사 유형별 뉴스 보도량

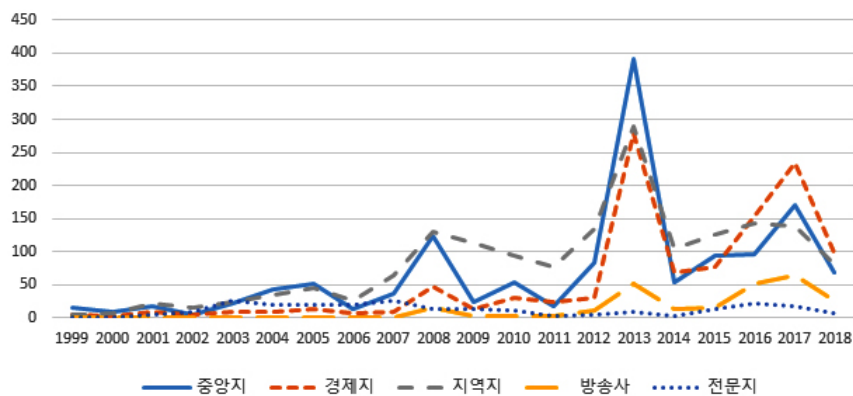
언론사 유형별로 보도된 뉴스 기사 건 수를

해당 유형의 언론사 수로 나눈 후 이를 백분율로 나타내면 언론사 유형별 뉴스 보도 점유율을 계산할 수 있다. <그림 8>에서 보는 바와 같이, 중앙지와 경제지가 기록관리 분야의 뉴스를 가장 많이 보도한 반면, <그림 7>에서 가장 많은 뉴스를 보도한 지역지가 오히려 뉴스 보도 비율이 가장 낮은 것으로 나타났다. 이러한 이유는 지역지의 종 수가 27종으로 가장 많아 지역지 1종 당 평균 보도량이 적어졌기 때문인 것으로 볼 수 있다.



<그림 8> 언론사 유형별 뉴스 보도 점유율

<그림 9>는 언론사 유형별 보도량에 대한 시계열 분석을 위해 연간 언론사 유형별 뉴스 보



<그림 9> 연간 언론사 유형별 보도량 추이

도량 추이를 분석한 결과이다. 이 중 방송사의 경우 2007년까지 뉴스 보도가 거의 전무하다. 2008년부터 적은 양이지만 기사의 수가 증가한 것으로 나타났다. 전체적인 추이는 <그림 3>의 연간 보도량 변화 추이와 유사한 것으로 나타났다.

<그림 10>에서는 시기별 언론사 유형별 뉴스 보도량을 나타내고 있다. 문재인 대통령 임기를 제외한 나머지 시기에는 지역지와 중앙지가 많은 양의 기사를 보도하였으며, 전문지는 노무현 전 대통령 임기에 상대적으로 많은 양

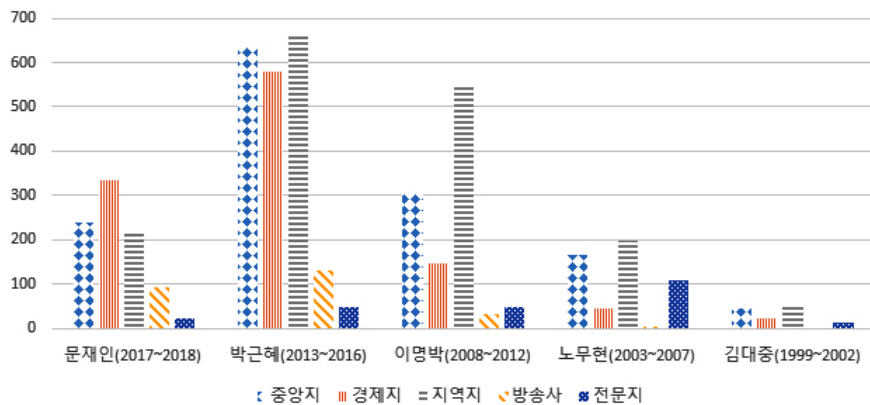
의 기사를 보도한 것으로 나타났다.

4.2 빈도 분석 기반 보도 내용 분석 결과

4.2.1 뉴스 전체 분석

수집된 전체 뉴스를 대상으로 '빅카인즈'의 메타정보에서 제공하는 키워드를 추출하여 그 빈도를 계산하였으며, 이 중 상위 30위에 해당하는 단어는 <표 2>와 같다.

빈도가 가장 높은 단어는 '기록물'이며, 그 다음으로 '대통령', '관리', '기록', '검찰', '청와대'



<그림 10> 시기별 언론사 유형별 보도량

<표 2> '기록물 관리' 관련 전체 뉴스 기사에 대한 상위 30위 빈도 단어 리스트

순위	단어	빈도	순위	단어	빈도	순위	단어	빈도
1	기록물	17,323	11	자료	4,000	21	협의	2,574
2	대통령	15,622	12	의원	3,863	22	보존	2,557
3	관리	12,116	13	수사	3,782	23	조사	2,485
4	기록	7,482	14	대통령기록물	3,443	24	의혹	2,454
5	검찰	6,913	15	문서	3,257	25	국가	2,409
6	청와대	6,417	16	열람	3,113	26	폐기	2,400
7	공개	4,926	17	문건	3,027	27	주장	2,398
8	정부	4,589	18	회의록	2,966	28	법률	2,391
9	국가기록원	4,387	19	이관	2,747	29	업무	2,347
10	대화록	4,073	20	유출	2,713	30	사건	2,334

등의 단어가 그 뒤를 이었다. 상위 빈도를 차지하는 단어를 통해 유추해보면, 기록관리 관련 뉴스 기사의 주요 주제가 대통령 기록물과 깊은 관계가 있다는 것을 알 수 있다. 그리고 이는 2007년 대통령기록물관리법이 제정된 이후 뉴스 기사가 증가하기 시작한 <그림 3>의 분석 결과와 일맥상통한다.

가장 보도량이 많았던 2013년에는 참여정부 시절 작성된 남북정상회담 회의록 유출 논란이 있었다. 이를 반영하여 '회의록', '대화록', '유출' 등과 같은 단어가 높은 빈도 순위를 차지하였다. 기록관리 행위를 나타내는 '이관', '폐기', '공개', '열람', '보존' 등의 단어 역시 뉴스 기사에 자주 출현한 것을 확인할 수 있었으며, '수사', '조사', '혐의', '의혹', '사건' 등과 같은 단어가 고빈도어인 것을 볼 때, 기록관리와 관련된 일련의 이슈들이 있었고 이들이 기사화되었음을 유추할 수 있다.

상위 30위에는 포함되지 않았으나, 전·현직 대통령 중 노무현 대통령(2,005, 41위)이 가장 높은 빈도를 나타냈으며, 이어서 박근혜(1,569, 54위), 문재인(1,105, 104위), 이명박(1,095, 106위), 김대중(133, 1,025위) 대통령의 순으로 나타났다.

4.2.2 시기별 분석

대통령 임기별로 상위 빈도 30위에 포함된 단어는 <표 3>과 같다. 이를 통해 시기별 기록관리 뉴스가 다루었던 주요 이슈와 시기별 주요 현상을 파악할 수 있다.

1) 김대중 대통령 시기(1999~2002)

이 시기에는 기록의 유사 개념이라 할 수 있

는 '정보', '자료'와 같은 단어들이 높은 빈도를 나타냈는데, 이 시기는 우리나라에서 기록관리 개념을 도입한 초기 단계였기 때문에 개념이 정립되는 과정에서 이를 표현하는 단어들이 혼재되어 사용되었기 때문인 것으로 볼 수 있다. 이러한 현상은 2000년대 초반까지 이어지는데, 2000년에 상위 빈도어로 나타났던 '도서관', '공공도서관', 2002년의 '지식'과 같은 단어가 2000년대 중반에는 상위 빈도에서 사라졌다([부록 1] 참조). 또한 '솔루션', '구축', '계획', '추진'과 같은 단어를 통해 이 시기에는 기록관리의 체계적 도입을 위해 시스템적 접근이 요구되었고, 이를 위한 다양한 추진 계획이 수립되었음을 유추할 수 있다. 이러한 현상을 통해 김대중 대통령 시기는 기록관리 개념이 도입되고 관련 법 시행을 위한 체계화 시기였다는 것을 알 수 있다.

또한, 이 시기에는 현재는 사라진 개념을 지칭하는 용어들이 상위 빈도에 나타났는데, '자료관'과 '정부기록보존소'가 그 예이다. '공공기관의 기록물 관리에 관한 법률'이 2000년에 시행되면서 자료관의 개념이 도입되고 2003년부터는 자료관 시스템의 보급과 확산이 본격적으로 이루어지기 시작했다. 현재의 기록관으로 대체되는 이 단어는 2002년 이후로 상위 빈도에서 사라졌다. 또한 현재 국가기록원의 전신이라 할 수 있는 '정부기록보존소'가 상대적으로 높은 빈도 순위를 나타냈는데, 2004년에 국가기록원으로 명칭이 변경되면서 2004년 이후에는 상위 빈도에서 사라졌다([부록 1] 참조).

2) 노무현 대통령 시기(2003~2007)

이 시기는 김대중 대통령 시기에 이어 기록관리 제도가 체계화되면서 확장 발전하는 시기라

〈표 3〉 시기별 상위 30위 빈도 단어 리스트

시기 순위	김대중 (1999~2002)		노무현 (2003~2007)		이명박 (2008~2012)		박근혜 (2013~2016)		문재인 (2017~2018)	
	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도
1	기록물	352	기록물	2,476	기록물	4,581	대통령	7,082	대통령	4,154
2	관리	313	관리	2,003	대통령	3,306	기록물	6,179	기록물	3,735
3	대통령	258	기록	1,306	관리	3,067	관리	4,763	청와대	2,025
4	기록	246	대통령	822	기록	1,956	검찰	4,683	관리	1,970
5	정보	143	정부	606	국가기록원	1,102	대화록	3,578	검찰	1,628
6	보존	137	보존	574	자료	1,024	청와대	3,236	정부	1,402
7	문서	134	문서	548	청와대	973	기록	2,768	문건	1,266
8	구축	124	구축	544	정부	932	의원	2,569	기록	1,206
9	문건	123	국가기록원	532	공개	785	공개	2,511	공개	1,112
10	정부	114	기관	490	유출	721	회의록	2,474	수사	1,085
11	공공기관	113	자료관	487	국가	633	수사	2,319	후보	1,002
12	경찰	109	국가	471	사업	600	열람	2,087	자료	982
13	자료	109	공개	462	열람	584	대통령 기록물	2,077	이관	928
14	한나라당	105	업무	424	보존	580	국가기록원	1,951	장관	906
15	법률	104	사업	420	의원	574	새누리당	1,816	문서	890
16	사업	97	추진	378	업무	555	국정원	1,797	대통령 기록물	873
17	작성	93	자료	372	폐기	555	의혹	1,703	국가기록원	802
18	의원	91	정보	350	평가	534	혐의	1,651	박근혜	799
19	유출	86	폐기	304	기관	527	민주당	1,595	파기	719
20	회의록	74	공공기관	296	법률	523	문건	1,553	혐의	686
21	정부기록 보존소	69	계획	290	검찰	517	정부	1,535	앵커	668
22	기관	68	솔루션	259	노무현	508	유출	1,520	위반	668
23	자료관	66	보관	257	추진	463	자료	1,513	지정	639
24	국가	65	생산	249	관계자	455	조사	1,402	주장	627
25	추진	65	홈페이지	241	대통령 기록물	443	사건	1,394	법률	601
26	공공	64	공공	241	고발	438	국회	1,359	위안부	599
27	역사	61	법률	238	생산	436	남북 정상회담	1,321	의원	585
28	솔루션	59	운영	230	문서	430	주장	1,287	조사	582
29	계획	57	대표	223	보관	417	이관	1,276	생산	565
30										

할 수 있다. 전 시기에 이어 '자료관', '솔루션', '구축' 등과 같은 단어가 지속적으로 상위 빈도에 놓여 있는 것을 볼 때, 전 시기에 시행되기 시작한 공공기록물법의 적용이 확산되는 시기였다는 것을 알 수 있다. 또한 '공공', '공공기관'와 같은 단어를 통해 기록관리의 공공성이 강조되는 시기였음을 알 수 있다. 그리고 이러한 단어는 2004년 이후에는 상위 빈도에서 사라졌는데, 이는 기록관리의 공공성 개념이 정착되었기 때문인 것으로 해석할 수 있다.

또한 '공개'라는 단어가 이 시기에 처음 등장하였는데, 2004년에 정보공개법이 개정되면서 기록의 보존을 중시하던 이전 시기를 벗어나, 이 시기에는 기록정보의 공개와 활용이 사회적으로 요구되기 시작했다는 것을 알 수 있다. 마찬가지로 '폐기' 역시 이 시기에 처음 등장하였는데, 이를 통해 이 시기에는 기록관리의 기능이 보존 이외로 확장되었음을 알 수 있다. 그리고 이 단어는 이후의 시기에서도 지속적으로 상위 빈도를 차지하게 된다([부록 1] 참조).

3) 이명박 대통령 시기(2008~2012)

이 시기에는 '청와대'와 '노무현'과 같은 정치적 개념을 상징하는 단어가 처음으로 상위 빈도에 등장하였다. 이 시기는 노무현 전 대통령의 퇴임 후 봉하 이지원(e-지원) 시스템 반출에 대해 '대통령 기록물 관리에 관한 법률(2007)' 위반 여부 논란이 있었던 시기로, '대통령 기록물'과 함께 '유출', '고발' 등과 같은 단어도 상위 빈도에 함께 포함되었다. 4.1.2의 분석 결과와 마찬가지로, '대통령기록물'은 이후에도 다양한 정치적 사안과 맞물리며 정치 주체의 뉴스의 보도량이 증가하는 계기가 되며, 지속적

으로 상위 빈도어에 놓이게 된다.

이외에도 '평가', '열람', '생산' 등이 새롭게 상위 빈도어로 추가되었는데, 이는 앞에서 분석한 바와 같이, 기록관리의 기능이 확장되고 있음을 의미한다. 또한 2009년의 '대학', 2010년의 '지자체'는 기록물 관리 기관이 대통령과 중앙정부 외에 다른 공공기관으로 점차 확장되고 있음을 나타낸다([부록 1] 참조).

4) 박근혜 대통령 시기(2013~2016)

기록물 관리 뉴스 기사가 가장 많이 생산된 이 시기는 전 시기에 이어 '국정원', '검찰', '수사', '새누리당', '민주당' 등과 같은 정치적 단어들이 상위 빈도어에 새롭게 추가되었다. 이 시기는 기록관리와 관련된 다양한 정치적 이슈와 논란이 존재했던 시기로, 먼저, 노무현 전 대통령 임기 중 있었던 '남북정상회담', '회의록' 중 'NLL(2013년에 1회 등장, [부록 1] 참조)', '대화록'이 유출되면서 이것이 대통령기록물인지 공공기록물인지의 여부와 함께 회의록 공개 여부를 두고 장기간 동안 정당 간 논쟁이 치열했던 시기였다. 이후, 최순실 국정농단에 대통령 기록물법 위반 사항이 포함되면서 '최순실', '연설문'이 2016년에 상위 빈도어로 등장하게 되었다([부록 1] 참조). 사실 이 시기는 다양한 사건으로 인해 기록물 관리 본연의 내용보다는 기록물 관리와 맞물린 정치적 사안들이 뉴스 주체의 주를 차지하게 되었다. 이러한 이유로 <그림 6>에서 분석했던 바와 같이, 이 시기에는 기록관리와 관련하여 정치적 주체의 기사들이 대량 보도된 것을 알 수 있다.

5) 문재인 대통령 시기(2017~2018, 현재
입기 중)

이 시기는 최순실 국정농단을 시작으로, 우리나라 헌정수립 최초로 대통령 탄핵과 이에 따른 조기 대선이 있었던 시기이다. 그리고 이 과정에서 '후보'간 대통령기록물 이슈가 다시 한번 부각되었다. 이 시기에 처음 상위빈도어로 등장한 '박근혜'는 박근혜 정부의 국정농단 사태가 기록물과 밀접한 영향을 갖고 있음을 증명한다. 또한 탄핵 대통령의 대통령기록물 '이관'이라는 초유의 사태가 벌어지면서 박근혜 전 대통령의 기록물의 무단 '파기' 여부를 놓고 논란이 있었던 시기이기도 하다.

정치적 이슈 이외에도 이 시기에는 '위안부'라는 단어가 처음 등장하였는데, 얼마 전 한국인 위안부의 모습이 담긴 영상 기록이 최초로 공개되면서 큰 사회적 파장을 일으켰다. 이외에도, '이명박', '4대강', '다스' 등 MB 정부와 관련된 단어가 2018년에 처음 등장하면서 현재 진행 중인 이명박 대통령의 비리 수사에도 기록물이 밀접한 관계가 있음을 유추할 수 있다.

시기별 보도 내용의 분석 결과를 종합해보면, 각 대통령 임기별 세부적인 보도 내용의 차이는 있으나, 일반적으로, 김대중, 노무현 전 대통령 임기였던 기록관리 도입 전반기 약 10년간은 기록관리의 개념이 도입, 적용, 확장되는 시기에서 발생하는 이슈들이 주된 뉴스 주제로 보도되었다가, 2007년 대통령기록물관리법이 제정된 이후의 이명박, 박근혜 전 대통령과 문재인 현 대통령의 임기 약 10년 간은 대통령기록물을 중심으로 한 정치적, 사회적 이슈와 관련된 주제로 많은 양의 뉴스가 보도된 것을 확

인할 수 있다.

4.2.3 언론사 유형별 주제 분석

일반적으로 언론사 유형은 기사의 주제를 구성하는 방향과 무관하지 않다. 그러므로 이 연구에서는 언론사 유형별로 출현 단어의 빈도를 분석하여 상위 빈도 주제어의 언론사 유형별 특성을 바탕으로 주제별 특성까지 함께 유추하고자 하였다.

〈표 4〉는 언론사 유형별 뉴스에서 추출한 상위 30개 빈도의 단어를 나타낸다. 기사 전체에서 추출된 상위 30개 고빈도어와 각 언론사 유형별 상위 빈도어의 일치도를 비교해 보면, 중앙지와 경제지가 26개로 가장 높고, 그 다음으로 지역지, 방송사, 전문지의 순으로 나타났다.

중앙지와 경제지의 경우, 전체 기사에 대한 상위 빈도어와 유사하게 기록관리와 관련된 정치적 이슈를 나타내는 단어들이 상위 빈도어에 포함된 것을 볼 수 있는데, 이는 〈그림 8〉에서 보는 바와 같이 중앙지와 경제지가 가장 높은 뉴스 보도 점유율을 나타내고 있기 때문이며, 이를 통해 중앙지와 경제지가 다루는 기사의 주제가 유사했음을 알 수 있다. 지역지의 경우는 전체 기사의 상위 빈도어 중 15개와 일치하는 단어 외에 '보존', '교육', '평가', '사업', '행정', '생산', '학교', '운영' 등 기록관리 업무 과정과 관계가 있는 용어들이 상위 빈도어에 출현한 것으로 나타났다. 방송사는 '앵커', '인터뷰', '얘기', '말씀', '입장', '사람' 등과 같이 기록물 관리와 관련된 개념보다는 방송사 뉴스 멘트에서 자주 출현하는 단어들이 상위 빈도어에 고유하게 출현한 것으로 나타났다.

〈표 4〉 언론사 유형별 상위 30위 빈도 단어 리스트

순위	전체	빈도	중앙지	빈도	경제지	빈도	지역지	빈도	방송사	빈도	전문지	빈도
1	기록물	17,323	대통령	6,370	대통령	4,158	기록물	6,786	대통령	2,449	관리	1,125
2	대통령	15,622	기록물	5,426	기록물	3,496	관리	5,448	검찰	1,209	기록물	935
3	관리	12,116	관리	3,231	검찰	2,089	기록	2,543	청와대	1,089	기록	548
4	기록	7,482	기록	3,057	관리	2,032	대통령	2,478	앵커	1,003	구축	517
5	검찰	6,913	검찰	2,760	청와대	1,969	보존	1,167	인터뷰	796	사업	455
6	청와대	6,417	청와대	2,532	공개	1,262	국가 기록원	1,083	얘기	791	솔루션	386
7	공개	4,926	공개	2,410	대화록	1,242	업무	1,082	수사	695	자료관	316
8	정부	4,589	국가 기록원	1,938	대통령 기록물	1,176	자료	1,035	기록물	680	문서	293
9	국가 기록원	4,387	정부	1,878	기록	1,175	교육	1,013	최순실	648	전자	275
10	대화록	4,073	대화록	1,795	수사	1,141	평가	951	문건	601	기업	270
11	자료	4,000	회의록	1,690	정부	1,071	정부	932	의원	536	업무	264
12	의원	3,863	자료	1,592	의원	1,039	기관	923	상황	512	정보	260
13	수사	3,782	의원	1,585	문건	1,015	문서	828	생각	502	정부	259
14	대통령 기록물	3,443	열람	1,493	협의	999	검찰	801	박근혜	501	RFID	258
15	문서	3,257	대통령 기록물	1,413	국가 기록원	996	사업	793	조사	482	시장	238
16	열람	3,113	수사	1,409	자료	985	공개	785	협의	451	홈페이지	237
17	문건	3,027	이관	1,243	열람	857	청와대	782	입장	450	보존	231
18	회의록	2,966	문서	1,167	유출	804	운영	761	정부	449	기관	225
19	이관	2,747	국정원	1,164	회의록	796	추진	741	말씀	443	국가 기록원	224
20	유출	2,713	폐기	1,115	문서	774	대상	725	후보	440	통합	212
21	협의	2,574	유출	1,059	의혹	770	대화록	702	공개	379	활용	212
22	보존	2,557	문건	1,041	위반	746	의원	683	사실	364	데이터	206
23	조사	2,485	주장	989	이관	734	기록관	667	의혹	341	대표	205
24	의혹	2,454	국가	980	사건	733	선정	645	정도	339	공공기관	203
25	국가	2,409	새누리당	978	법률	704	행정	643	대화록	331	서비스	202
26	폐기	2,400	사건	951	주장	697	생산	611	문재인	329	문의	201
27	주장	2,398	법률	949	국정원	663	계획	604	가지	322	국가	199
28	법률	2,391	국회	922	확인	639	학교	595	가능성	316	제품	196
29	업무	2,347	의혹	917	조사	624	폐기	584	유출	315	직종	196
30	사건	2,334	노무현	914	새누리당	601	열람	577	사람	311	경력	195
기사 전체 상위 빈도어와의 일치 단어 수			26		26		16		14		10	

령도 제정되기 이전이기 때문에 대통령기록물을 대신하는 표현으로 '통치사료'라는 단어를 사용한 것으로 보인다.

2) 노무현 대통령(2003~2007) 시기

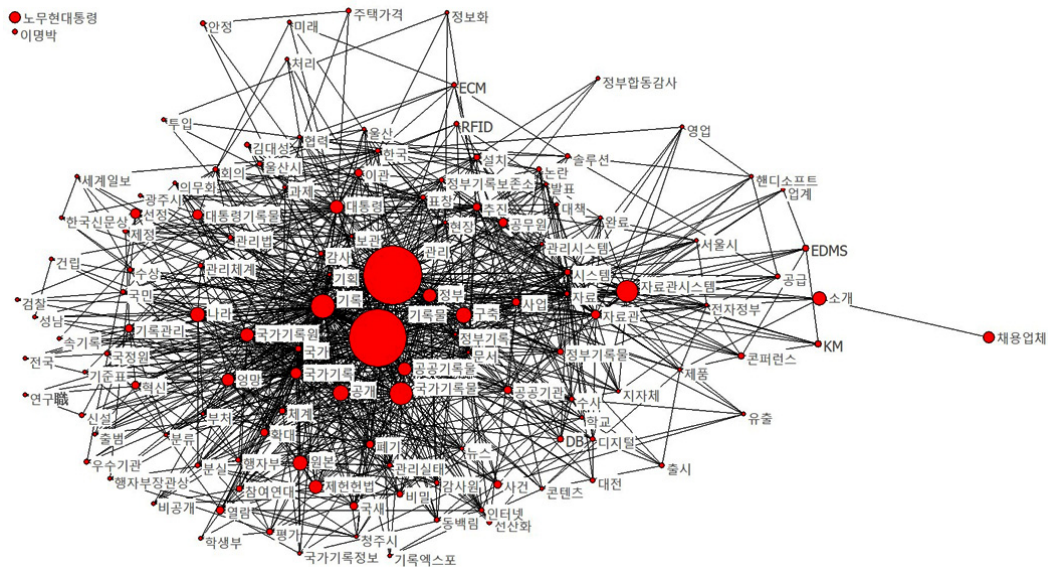
이 시기의 핵심어들은 '기록물'과 '관리'라는 두 개의 단어를 중심으로 하나의 커다란 네트워크를 형성하고 있다. 이 시기는 기록관리 개념이 확장되던 시기로, 정부기록, 공공기록, 국가기록, 대통령기록물 등 다양한 유형의 공공기록 관리에 관한 이슈가 언론을 통해 본격적으로 논의되었던 것으로 볼 수 있다. 또한 4.2.1에서 언급한 바와 같이 '국가기록원'의 전신인 '정부기록보존소'가 함께 네트워크 상에 존재하고 있으며, 기록관리시스템의 원형이라 할 수 있는 자료관시스템 구축 이슈가 기사에서 다루어졌던 것을 알 수 있다(그림 12) 참조).

3) 이명박 대통령(2008~2012) 시기

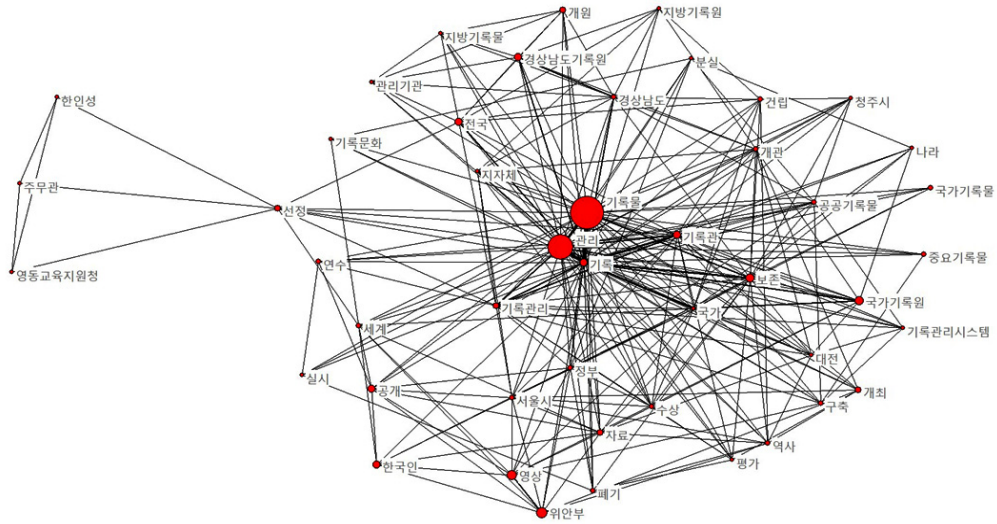
이 시기의 핵심어들은 '기록물'을 중심으로 하나의 커다란 네트워크를 형성하고 있다. 이 시기에는 대통령기록물 이슈가 처음으로 네트워크에 등장하였는데, 이후 대통령기록물은 정쟁의 주요 원인으로 자리잡게 된다. <그림 13>에서 보는 바와 같이, 네트워크 왼쪽으로는 지자체 및 공공기관 기록물의 관리와 관련된 개념들이 위치하고 있으며, 오른쪽으로는 노무현 대통령의 봉하 이지원 반출 및 남북정상회담 대화록 유출 등 대통령기록물 이슈와 관련된 단어들이 위치하는 것을 알 수 있다. 이 시기에는 처음으로 기록물 관리 주체가 정치적 사안과 맞물려 언론에 보도되는 시기로, 이러한 현상은 이후에도 지속적으로 유지되고 있다.

4) 박근혜 대통령(2013~2016) 시기

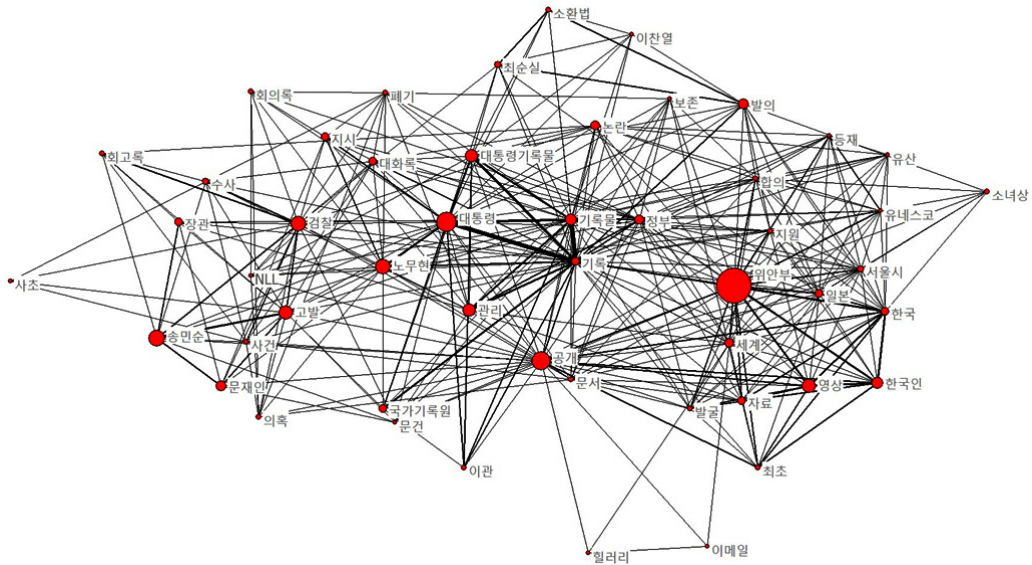
이 시기는 '대화록', '검찰', '기록물', 'NLL' 등



<그림 12> 노무현 대통령(2003~2007) 시기의 뉴스 기사 핵심 단어 네트워크



〈그림 19〉 문화 주제의 뉴스 기사 핵심 단어 네트워크



〈그림 20〉 국제 주제의 뉴스 기사 핵심 단어 네트워크

의 네트워크에서는 중앙 및 지방의 기록관의 기록관리 이슈를 확인할 수 있으며, 약하게 표현되어 있으나 기록 보존과 관련된 기록관리시스템 이슈와 서울시의 한국인 위안부 영상 기록 발견

주제의 보도 내용도 포함되어 있는 것을 알 수 있다. 이 위안부 영상 기록 발견 이슈는 국제 주제에서 더욱 두드러지게 네트워크 상에 표현되어 있는데, 그 내용은 〈그림 20〉과 같다.

이 주제에서는 기록관리와 관련된 국제적 이슈를 중심으로 네트워크가 형성되어 있는데, 문화 주제 네트워크에서 약하게 표현되었던 위안부 영상 기록 이슈가 이 네트워크에서는 오른쪽 제의 핵심 단어로 등장하며 위안부 기록에 대한 유네스코 기록 유산 등재 등과 관련된 주제의 네트워크 형성에 크게 기여하고 있다. 네트워크 오른쪽에는 송민순 전 외교부 장관의 회고록과 관련된 NLL 대화록 이슈가 자리하고 있다.

4) 지역

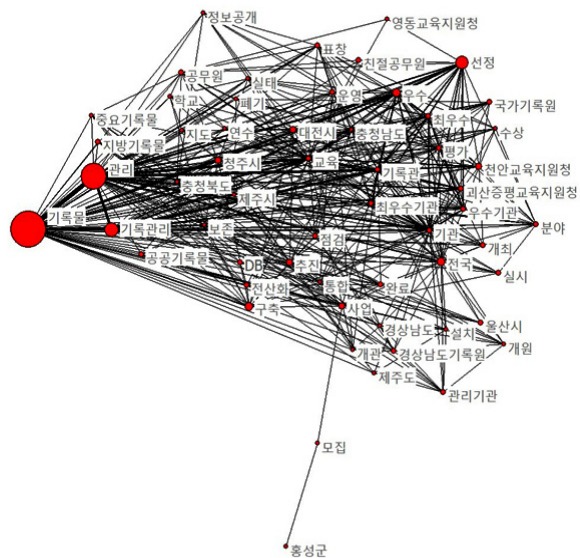
지역을 주제로 한 뉴스 기사의 단어 네트워크 분석 결과는 <그림 21>과 같다. 이 주제에서는 '기록물'을 중심으로 지자체 및 지역 교육청 기록관리 이슈를 나타내고 있다. 지역 기록관리의 이슈는 주로 전산화, 수상 및 표창, 기록관 개관 등을 포함하는데, 이를 통해 지역 주제의

기록관리 뉴스는 주로 지역 기록관의 소식 알리고 이들의 우수사례를 홍보하는 것을 목적으로 보도되고 있음을 확인할 수 있다.

5) IT_과학

이 주제에서는 '기록물'과 '관리' 두 용어를 중심으로 하여 IT_과학 주제의 연관어로 구성된 네트워크를 형성하였으며, 그 내용은 <그림 22>와 같다. 이 네트워크의 왼쪽에는 기록관리 시스템 구축과 관련된 단어들이 위치하고 있으며, 오른쪽에는 전자기록관리와 관련된 기술을 주제로 하는 단어들이 위치하고 있다. 이 주제에서는 기록관리가 지식관리(KM), 기업 콘텐츠 관리(ECM), 기록물의 디지털화, DB 기술, RFID, 전자정부 등과도 관련된 개념임을 확인할 수 있다.

주제별 언어 네트워크의 분석 내용을 종합해 보면, 정치와 사회 주제가 밀접한 관련을 맺고



<그림 21> 지역 주제의 뉴스 기사 핵심 단어 네트워크

제정된 1999년부터 약 10년 동안은 뉴스 보도의 양이 적었으나, 대통령기록물관리법이 처음 제정된 2007년 이후로 보도량의 증가 추세를 보여주었다. 특히 2008년, 2013년, 2017년에 보도량이 급격히 증가한 것으로 나타났는데, 이 시기에는 대통령기록물과 관련된 다양한 정치적 이슈가 존재했기 때문인 것을 확인할 수 있었다.

언론 보도량의 시기별 차이에 대해서는 대통령의 임기를 기준으로 20년을 5개의 시기로 나누어 분석하였는데, 김대중 전 대통령 시기에 보도량이 가장 적었고, 박근혜 전 대통령 시기에 보도량이 가장 많은 것으로 나타났다.

또한 언론 보도량의 뉴스 주제별 차이에 대해 분석한 결과, 기록관리 분야 뉴스에서 가장 많은 비중을 차지하는 주제는 '정치'와 '사회'인 것으로 나타났는데, 이는 기록관리가 우리나라의 정치 및 사회적 이슈와 밀접한 관계가 있다는 것을 증명한다.

언론 보도량의 언론사 유형별 차이에 대해 분석한 결과, 절대적인 수치로는 지역지가 가장 많은 뉴스를 보도하였으나, 각 언론사 유형별 매체 수를 고려하면 중앙지와 경제지가 가장 많은 뉴스를 보도한 것으로 나타났다.

다음으로, 기록관리 주제에 대한 언론 보도 내용의 차이 특성을 분석하기 위해 뉴스 기사 본문에서 추출한 키워드를 대상으로 빈도 분석과 뉴스 제목에서 추출한 키워드를 대상으로 언어 네트워크 분석을 실시하였고, 이를 통해 언론 보도 내용의 시기별, 주제별, 언론사 유형별 차이를 분석하였다.

전체 뉴스 기사를 대상으로 출현 단어의 빈도를 분석한 결과, 빈도가 가장 높은 단어로는

'기록물', '대통령', '관리', '기록', '검찰', '청와대' 등으로 나타났으며, 상위 30위에 해당하는 빈도를 차지하는 단어를 통해 유추해보면, 기록관리 관련 뉴스 기사의 주요 주제가 대통령기록물과 관련이 깊다는 것을 확인할 수 있었다.

또한 언론 보도 내용의 시기별 차이에 대해 분석한 결과, 각 대통령 임기별 세부적인 차이는 있으나, 기록관리 도입 전반기 약 10년 간은 기록관리의 개념이 도입, 적용, 확장되는 시기에서 발생하는 이슈들이 주된 뉴스 주제로 보도되었다가, 2007년 대통령기록물관리법이 제정된 이후 약 10년 간은 대통령기록물을 중심으로 한 정치적, 사회적 이슈와 관련된 주제로 많은 양의 뉴스가 보도된 것을 확인하였다.

언론사 유형별 분석을 통해 주제에 따른 보도 내용의 차이를 분석한 결과, 중앙지와 경제지가 전체 뉴스 기사의 주제와 유사하게 기록과 관련된 정치적, 사회적 이슈를 다룬 뉴스를 많이 보도한 것으로 나타났으며, 지역지는 기록관리 업무에서 발생하는 이슈들을 중심으로, 전문지는 기록물의 전자적 관리와 시스템 구축 및 관련 시장과 관련된 이슈를 중심으로 뉴스를 보도한 것으로 나타났다.

언어 네트워크 분석 결과, 김대중 전 대통령 임기에 보도된 뉴스를 대상으로 한 언어 네트워크는 단어 간 연결 강도가 낮은 형태로 네트워크를 형성하였고, 박근혜 전 대통령 임기의 언어 네트워크가 가장 연결 강도가 높은 형태로 네트워크를 형성한 것으로 나타났다. 임기별로 각각 생성된 언어 네트워크의 특징은 단어 빈도에 의한 내용 분석에서 살펴본 바와 마찬가지로, 임기별로 세부적인 차이가 있으며, 김대중, 노무현 전 대통령 시기에는 주로 기록

관리 개념의 도입과 확장 과정에서 일어난 이슈들을 중심으로 뉴스가 보도되었던 반면, 이명박, 박근혜 전 대통령 및 문재인 대통령 시기에는 대통령기록물을 중심으로 한 정치적, 사회적 이슈들을 중심으로 뉴스가 보도된 것을 확인하였다. 그리고 이러한 결과를 통해 단어 빈도에 의한 내용 분석 결과를 입증할 수 있게 되었다.

마지막으로, 주제별 뉴스에 대한 언어 네트워크의 분석 결과, 정치와 사회 주제가 밀접한 관련을 맺고 유사한 단어를 포함하는 네트워크를 형성하였으며, 경제, 지역, 국제, IT_과학 등 나머지 주제에 대해서는 기록관리와 관련된 각 주제의 특성을 반영한 뉴스가 보도되었음을 알 수 있다. 이러한 분석 결과는 언론사 유형별 단어빈도 정보에 기반한 내용 분석 결과와 유사

한 것으로 나타나, 언어 네트워크 분석을 통해 다시 한 번 단어 빈도에 기반한 내용 분석 결과를 입증할 수 있게 되었다.

이 연구는 약 20년이라는 장기간의 뉴스 빅데이터를 대상으로 대표적인 빅데이터 분석 기법인 언어 네트워크 분석법과 시계열 내용 분석 방법론을 적용하여 최초로 기록관리 분야의 뉴스 보도 특성을 분석하였다는 데 의의가 있다. 그간 문헌정보학/기록관리학 분야에서의 동향 분석 및 시계열 분석 연구는 학술지 연구 데이터를 기반으로 진행해 왔다. 이 연구를 계기로 후속 연구에서는 이 연구가 제시하고 있는 연구 모형을 바탕으로 하여 문헌정보학/기록관리학을 구성하는 특정 주제를 대상으로 뉴스 빅데이터를 기반 시계열적 분석을 시도할 수 있다.

참 고 문 헌

- 강범일, 송민, 조화순 (2013). 토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구. 한국문헌정보학회지, 47(4), 315-334. <http://dx.doi.org/10.4275/KSLIS.2013.47.4315>
- 국립중앙과학관 (2018. 6. 1.) 빅데이터란 무엇인가? Retrieved from https://smart.science.go.kr/scienceSubject/bigdata/view.action?menuCd=DOM_000000101001013000&subject_sid=1201
- 김봉제 (2018). Big Data 분석을 통한 한국사회의 도덕·윤리 용어 사용 특성 연구. 도덕윤리과교육, (58), 27-58.
- 김수연, 김대욱, 최명일 (2013). 한국 광고홍보학 연구경향 언어 네트워크 분석: 『광고연구』, 『광고학연구』, 『홍보학연구』, 『한국광고홍보학회』에 게재된 논문의 핵심어를 중심으로. 한국광고홍보학회, 15(1), 59-85.
- 김희정 (2006). 국내외 기록관리학 연구영역의 시계열적 분석 및 확장성 연구. 한국기록관리학회지, 6(2), 5-25.

- 박대민 (2013). 뉴스 기사의 빅데이터 분석 방법으로서 뉴스정보원연결망분석. 한국언론학보, 57(6), 234-262.
- 박대민 (2016). 장기 시계열 내용 분석을 위한 뉴스 빅데이터 분석의 활용 가능성: 100만 건 기사의 정보원과 주제로 본 신문 26년. 한국언론학보, 60(5), 353-407.
- 박승정, 전진오, 김선우, 김성태 (2017). 국내 주요일간지의 대통령 이슈소유권에 대한 빅데이터 분석: LDA(Latent Dirichlet Allocation) 토픽 모델링(Topic Modeling)을 중심으로. 정치정보연구, 20(3), 25-55. <http://dx.doi.org/10.15617/psc.2017.10.31.3.25>
- 박준형, 류법오, 오효정 (2018). 시계열 기반 국내 기록관리학 토픽 트렌드 분석. 한국기록관리학회지, 18(1), 29-47. <http://dx.doi.org/10.14404/JKSARM.2018.18.1.029>
- 박한우, Leydesdorff, L. (2004). 한국어의 내용분석을 위한 KrKwic 프로그램의 이해와 적용: Daum.net 에서 제공된 지역혁신에 관한 뉴스를 대상으로. Journal of the Korean Data Analysis Society, 6(5), 1377-1388.
- 서은경, 유소영 (2013). 국내 정보학분야 연구동향 분석: 2000-2011. 정보관리학회지, 30(4), 215-239. <http://dx.doi.org/10.3743/KOSIM.2013.30.4.215>
- 성지영, 김성균 (2017). 뉴스 빅데이터를 통한 덕수궁 돌담길의 장소 담론 해석. 한국디지털콘텐츠학회 논문지, 18(5), 923-932. <http://dx.doi.org/10.9728/dcs.2017.18.5.923>
- 신성현 (2016, 8, 10). [미디어 혁신가] “기사 3000만 건 분석 빅카인즈.. IBM보다 2년 앞서.” 조선비즈. Retrieved from http://biz.chosun.com/site/data/html_dir/2016/08/10/2016081001399.html
- 유사라 (2010). 메타데이터 주제 국내 연구동향 분석. 한국문헌정보학회지, 44(2), 405-426. <http://dx.doi.org/10.4275/KSLIS.2010.44.2.405>
- 이수상 (2012). 네트워크 분석 방법론. 서울: 논형.
- 이수상 (2017). 신문기사에 나타난 경주지진 사건의 사회적 이슈분석. 한국도서관·정보학회지, 48(2), 53-72. <http://dx.doi.org/10.16981/kliss.48.201706.53>
- 이은별, 전진오, 백지선 (2017). 서울의 다문화 공간 연구: 뉴스 빅데이터 분석 시스템 ‘빅카인즈’를 이용한 국내 언론의 외국인 마을 보도(1990~2016) 분석. 미디어 경제와 문화, 15(2), 7-43. <http://dx.doi.org/10.21328/JMEC.2017.5.15.2.7>
- 이재윤, 문주영, 김희정 (2007). 텍스트 마이닝을 이용한 국내 기록관리학 분야 지적구조 분석. 한국문헌정보학회지, 41(1), 345-372.
- 장정우, 최경호 (2012). 언어네트워크분석을 이용한 통계법 내용분석. 통계연구, 17(2), 53-66.
- 최윤정, 권상희 (2014). ‘빅데이터’ 관련 신문기사의 의미연결망 분석. 사이버커뮤니케이션학보, 31(1), 241-286.
- 최형욱, 최예진, 남소연 (2018). 문헌정보학 분야의 지적구조 및 연구 동향 변화에 대한 시계열 분석:

2003년부터 2017년까지. 정보관리학회지, 35(2), 89-114.

<http://dx.doi.org/10.3743/KOSIM.2018.35.2.089>

카카오 (2015). 다음카카오, 공식 블로그에 다음뉴스 배치와 주요 데이터 제공. 카카오 보도자료.

Retrieved from

<https://www.kakaocorp.com/kakao/prChannel/pressReleasesView?boardIdx=8356¤tPage=1¤tGroup=1>

한국언론진흥재단 (2018). 빅카인즈 사용자 매뉴얼. Retrived from

https://www.kinds.or.kr/manual/%EB%B9%85%EC%B9%B4%EC%9D%B8%EC%A6%88_%EC%82%AC%EC%9A%A9%EC%9E%90%EB%A7%A4%EB%89%B4%EC%96%BC.pdf

Blei, D. M. (2012). Probabilistic topic modeling, *Communications of the ACM*, 55(4), 77-84.

<http://dx.doi.org/10.1145/2133806.2133826>

Borgatti, S. P. (2002). *NetDraw: graph visualization software*. Harvard: Analytic Technologies.

Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *Ucinet 6 for Windows: software for social network analysis*. Harvard, MA: Analytic Technologies.

• 국문 참고문헌에 대한 영문 표기

(English translation of references written in Korean)

Choi, Hyung Wook, Choi, Ye-Jin, & Nam, So-Yeon (2018). Time series analysis of intellectual structure and research trend changes in the field of library and information science: 2003 to 2017. *Journal of the Korean Society for Information Management*, 35(2), 89-114.

<http://dx.doi.org/10.3743/KOSIM.2018.35.2.089>

Choi, Yoon-jung, & Kweon, Sang-Hee (2014). A semantic network analysis of the newspaper articles on big data. *Journal of Cybercommunication Academic Society*, 31(1), 241-286.

Jang, Jeong-Woo, & Choi, Kyoung-Ho (2012). Statistics act content analysis using semantic network analysis. *Journal of Korean Official Statistics*, 17(2), 53-66.

Kakao Corp. (2015). *Daum Kakao, providing Daum news layout and key data on the official blog*. Kakao Press Release. Retrieved from

<https://www.kakaocorp.com/kakao/prChannel/pressReleasesView?boardIdx=8356¤tPage=1¤tGroup=1>

Kang, Beomil, Song, Min, & Jho, Whasun (2013). A study on opinion mining of newspaper texts based on topic modeling. *Journal of the Korean Society for Library and Information*

- Science, 47(4), 315-334. <http://dx.doi.org/10.4275/KSLIS.2013.47.4315>
- Kim, Bong Je (2018). A study on the characteristics of the use of moral and ethical terms in korean society through big data analysis. *Journal of Moral & Ethics Education*, (58), 27-58.
- Kim, Hee-Jung (2006). A study on the serial analysis and expansion of research areas of records management and archives. *Journal of Korean Society of Archives and Records Management*, 6(2), 5-25.
- Kim, Soo-Yeon, Kim, Daewook, & Choi, Myung-II (2013). Semantic network analysis of advertising and public relations studies in korea - Using key words in Advertising Research, *The Korean Journal of Advertising*, *The Journal of Public Relations Research*, and *The Korean Journal of Advertising and Public Relation*, *The Korean Journal of Advertising and Public Relations*, 15(1), 59-85.
- Korea Press Foundation (2018). Bigkinds user manual. Retrieved from https://www.kinds.or.kr/manual/%EB%B9%85%EC%B9%B4%EC%9D%B8%EC%A6%88_%EC%82%AC%EC%9A%A9%EC%9E%90%EB%A7%A4%EB%89%B4%EC%96%BC.pdf
- Lee, Eunbyul, John, Jino, & Baek, Jeesun (2017). A study of multicultural space in Seoul: Analysing the coverage of foreign communities with news big data analytics 'BigKinds' for 27 years. *Journal of Media Economics & Culture*, 15(2), 7-43. <http://dx.doi.org/10.21328/JMEC.2017.5.15.2.7>
- Lee, Jae-Yun, Moon, Ju-Young, & Kim, Hee-Jung (2007). Examining the intellectual structure of records management & archival science in Korea with text mining. *Journal of the Korean Society for Library and Information Science*, 41(1), 345-372.
- Lee, Soo-Sang (2012). *Network analysis methods*. Seoul: Nonhyung.
- Lee, Soo-Sang (2017). Analysis of social issues of the newspaper articles on Gyeongju earthquakes. *Journal of the Korean Library and Information Science Society*, 48(2), 53-72. <http://dx.doi.org/10.16981/kliss.48.201706.53>
- National Science Museum of Korea (2018, June 1). What is the big data? Retrieved from https://smart.science.go.kr/scienceSubject/bigdata/view.action?menuCd=DOM_000000101001013000&subject_sid=1201
- Park, Daemin (2013). News information source network analysis as a big data analysis method of news articles. *Korean Journal of Journalism & Communication Studies*, 57(6), 234-262.
- Park, Daemin (2016). Automated time series content analysis with news big data analytics: Analyzing sources and quotes in one million news articles for 26 years. *Korean Journal*

- of Journalism & Communication Studies, 60(5), 353-407.
- Park, Han Woo, & Leydesdorff, L. (2004). Understanding the KrKwic: A computer program for the analysis of Korean text. *Journal of the Korean Data Analysis Society*, 6(5), 1377-1388.
- Park, JunHyeong, Ryu, Pum-Mo, & Oh, Hyo-Jung (2018). Timeline-based topic trend analysis of archives management in Korea. *Journal of Korean Society of Archives and Records Management*, 18(1), 29-47. <http://dx.doi.org/10.14404/JKSARM.2018.18.1.029>
- Park, Seung Jeong, John, Jino, Kim, Seon Woo, & Kim, Sung Tae (2017). A big-data analysis of presidential issue ownership of two prestigious korean newspapers: Focusing on LDA (latent dirichlet allocation) topic modeling. *The Journal of Political Science & Communication*, 20(3), 25-55. <http://dx.doi.org/10.15617/psc.2017.10.31.3.25>
- Seo, Eun-Gyoung, & Yu, So-Young (2013). Detecting research trends in Korean information science research, 2000-2011. *Journal of the Korean Society for Information Management*, 30(4), 215-239. <http://dx.doi.org/10.3743/KOSIM.2013.30.4.215>
- Shin, Seongheon (2016, Aug, 10). [Media innovator] "Bigkinds, analyzing 30 million news articles.. 2 years ahead of IBM." *Chosun Biz*. Retrieved from http://biz.chosun.com/site/data/html_dir/2016/08/10/2016081001399.html
- Sung, Ji-Young, & Kim, Sung-Kyun (2017). Interpretation of the place discourse of Deoksugung Doldam-gil through news big data. *Journal of Digital Contents Society*, 18(5), 923-932. <http://dx.doi.org/10.9728/dcs.2017.18.5.923>
- Yoo, Sarah (2010). A diagnostic analysis of metadata R&D status in Korea. *Journal of the Korean Society for Library and Information Science*, 44(2), 405-426. <http://dx.doi.org/10.4275/KSLIS.2010.44.2.405>

[부록 1] 연도별 상위 20위 빈도 단어 리스트

연도 순위	2018년		2017년		2016년		2015년		2014년		2013년		2012년		2011년		2010년		2009년	
	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도
1	기록물	1,434	대통령	3,020	대통령	2,495	기록물	1,095	기록물	655	기록물	3,428	기록물	980	기록물	466	기록물	773	기록물	850
2	대통령	1,134	기록물	2,301	관리	1,133	914	청와대	547	청와대	547	3,428	대통령	736	관리	381	관리	674	관리	584
3	관리	868	청와대	1,763	청와대	1,104	862	문건	512	문건	512	3,395	관리	660	기록	203	기록	454	기록	304
4	검찰	664	정부	1,182	최순실	1,099	594	검찰	500	검찰	500	2,581	기록	414	평가	104	대통령	260	대통령	186
5	과기	524	관리	1,102	기록물	1,001	525	관리	466	관리	466	2,506	대화록	358	대통령	90	정부	210	사업	162
6	기록	414	공개	1,013	관리	877	483	대통령	330	대통령	330	2,019	의원	340	국가 기록원	82	폐기	204	공개	153
7	수사	410	문건	1,002	협의	714	469	검찰	318	유출	318	1,839	공개	329	교육	81	사업	179	국가 기록원	140
8	4대강	386	후보	968	의혹	694	435	협의	307	의원	307	1,825	새누리당	284	기관	80	분야	168	기록관	113
9	국가 기록원	385	검찰	964	최씨	683	372	비서관	282	경정	282	1,814	후보	234	추진	78	보존	158	보존	112
10	이명박	376	장관	883	수사	677	361	회의	242	협의	242	1,713	NLL	232	자료	71	공개	146	생산	108
11	사업	343	이관	805	기록	599	320	무죄	218	수사	218	1,616	열람	206	정부	69	국가 기록원	143	업무	107
12	협의	340	기록	792	유출	580	318	기소	190	기록	190	1,526	정부	194	지원	69	추진	140	기관	97
13	문서	324	대통령 기록물	721	박근혜	563	294	대통령 기록물	177	사건	177	1,503	국회	190	행정	65	평가	138	예산	96
14	자료	318	박근혜	675	조사	473	288	경정	176	조사	176	1,385	청와대	186	사업	64	사분위	132	평가	94
15	보존	277	수사	664	위반	447	262	선고	152	작성	152	1,272	발인	185	요원	61	학교	131	구축	90
16	문건	264	자료	617	대통령 기록물	446	258	유출	148	정보	148	1,238	노무현	175	의원	58	국가	122	정보공개	89
17	청와대	262	지경	566	자료	433	253	업무	146	비서관	146	1,137	주장	175	진문	58	법률	119	정부	87
18	조사	252	문서	536	연설문	427	241	폐기	144	정부	144	1,117	업무	173	선정	57	지자체	118	대학	85
19	확인	237	문재인	531	적용	426	236	사건	140	공개	140	1,069	폐기	172	보존	54	자료	115	자료	85
20	디스	224	주장	531	문건	408	211	보존	135	의혹	135	1,060	청와대	161	대통령 기록관	52	교수	111	문서	80

연도 순위	2008년		2007년		2006년		2005년		2004년		2003년		2002년		2001년		2000년		1999년	
	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도
1	대통령 기록물	2,034	기록물	624	기록물	308	기록물	595	기록물	568	기록물	381	기록물	304	기록물	110	기록물	105	기록물	110
2	기록물	1,512	관리	493	관리	253	관리	502	관리	451	관리	304	기록물	100	기록물	110	기록물	105	기록물	110
3	관리	768	기록	328	기록	125	기록	306	기록	399	기록	190	기록물	71	기록물	107	기록물	57	기록물	89
4	청와대	684	대통령 기록물	293	홈페이지	124	원본	183	자료관	166	기록	148	기록물	57	기록물	102	기록물	55	기록물	67
5	국가 기록물	653	국가 기록물	195	근무지역	120	문서	176	대통령 기록물	165	정부	135	기록물	50	기록물	87	기록물	35	기록물	51
6	유출	637	사업	171	마감일	120	대통령 기록물	163	정부	153	구축	121	기록물	47	기록물	85	기록물	26	기록물	39
7	자료	619	공개	162	문의	120	국가 기록물	159	보존	152	기록물	115	기록물	44	기록물	83	기록물	24	기록물	33
8	기록	581	보존	142	지원자격	120	보존	152	문서	151	기록물	108	기록물	42	기록물	82	기록물	21	기록물	30
9	정부	372	추진	137	직종	120	국가 기록물	151	혜기	134	공개	106	기록물	37	기록물	67	기록물	21	기록물	30
10	검찰	364	국가 기록물	133	회사정보	120	정부 기록물	149	기관	131	문서	99	기록물	35	기록물	64	기록물	21	기록물	29
11	국가 기록물	311	기관	128	경력	116	구축	147	국가 기록물	125	기록물	97	기록물	33	기록물	63	기록물	20	기록물	29
12	고발	299	정부	127	대통령 기록물	86	업무	120	업무	118	기록물	95	기록물	31	기록물	56	기록물	20	기록물	25
13	방화마을	298	자료	111	서울	79	보관	116	공개	116	기록물	90	기록물	30	기록물	51	기록물	19	기록물	20
14	열람	283	감사	108	구축	74	헌법	116	국가 기록물	99	기록물	90	기록물	30	기록물	45	기록물	17	기록물	20
15	노무현	280	구축	105	나이부관	66	기관	111	구축	97	기록물	86	기록물	29	기록물	45	기록물	16	기록물	20
16	대통령 기록물	243	정보	95	대출	56	국세	109	자료	92	기록물	81	기록물	28	기록물	45	기록물	16	기록물	18
17	관계자	226	계획	94	국가 기록물	53	자료관	94	공공기관	82	기록물	78	기록물	27	기록물	42	기록물	16	기록물	17
18	조사	208	운영	94	국가 기록물	53	홈페이지	89	부처	75	기록물	71	기록물	25	기록물	42	기록물	14	기록물	17
19	법률	205	실치	87	문서	53	비밀	85	추진	69	기록물	65	기록물	25	기록물	40	기록물	14	기록물	17
20	수사	193	RFID	82	사업	53	지중	79	사업	67	기록물	65	기록물	23	기록물	40	기록물	13	기록물	15

