

텍스트 마이닝을 활용한 경제정책기록서비스 연구: 경제정책방향을 중심으로

A Study on the Archival Information Services of Economic Policy Using Text Mining Methods: Focusing on Economic Policy Directions

연지현(Jihyun Yeon)¹, 김성원(Sungwon Kim)²

E-mail: jhofarim@gmail.com, sungwonk@cnu.ac.kr



¹ 제 1저자 충남대학교 일반대학원 기록관리학과 석사과정
² 교신저자 충남대학교 사회과학대학 문헌정보학과 교수

논문접수 2022.04.20
최초심사 2022.04.25
게재확정 2022.05.15

ORCID

Jihyun Yeon
https://orcid.org/0000-0003-0079-8094
Sungwon Kim
https://orcid.org/0000-0003-2959-997X

초 록

자의적으로 구성된 기록 콘텐츠만으로는 이용자가 필요한 기간과 맥락에 대한 이해 없이 이용하게 됨으로써 주요한 경제정책기록에 효율적으로 접근하기에 어려움을 겪는다. 이러한 현재의 기록 서비스를 개선하기 위한 방안을 모색하고자 한다. 본 연구에서 1991년부터 2021년까지 30년간의 경제정책방향을 대상으로 경제정책기록에 텍스트 마이닝 기법을 활용하여 정부별 주요하게 다뤄진 경제 키워드와 변화과정을 도출하였다. 대책 배경, 주요 내용, 본문 텍스트를 수집하여 전처리를 진행한 후 텍스트 빈도분석, TF-IDF, 네트워크 분석, 시계열 분석을 진행하였다. 분석 결과 ‘일자리’, ‘경쟁력’, ‘구조조정’ 순으로 가장 높은 빈도수를 기록하였다. 정부별로 주요 키워드를 한눈에 볼 수 있었으며 ‘일자리’, ‘부동산’, ‘기업’의 연도별 상대비율을 시계열 순으로 분석하였다. 본 연구 결과를 바탕으로 향후 경제정책기록서비스의 발전과 저변확대를 위한 시사점을 제안하였다.

ABSTRACT

The archival content listed arbitrarily makes it difficult for users to efficiently access the records of major economic policies, especially given that they use it without understanding the required period and context. Using the text mining techniques in the 30-year economic policy direction from 1991 to 2021, this paper derives economic-related keywords and changes that the government mainly dealt with. It collects and preprocesses major economic policies' background, main content, and body text and conducts text frequency, term frequency-inverse document frequency (TF-IDF), network, and time series analyses. Based on these analyses, the following words are recorded in order of frequency: "job(일자리)," "competitive(경쟁력)," and "restructuring(구조조정)." In addition, the relative ratio of "job(일자리)," "real estate(부동산)," and "corporation(기업)," by year was analyzed in terms of chronological order while presenting major keywords mentioned by each government. Based on the results, this study presents implications for developing and broadening the area of archival information services related to economic policies.

Keywords: 경제정책, 기록정보서비스, 텍스트마이닝, 네트워크분석, 시계열분석
Economic Policy, Archival Information Service, Text Mining, Network Analysis,
Time Series Analysis

© 한국기록관리학회

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

1. 서론

1.1 연구목적

기존의 경제정책기록서비스는 대통령기록물에서 경제정책을 분석하여 생산기관의 기능과 역할에 기반해 정부 시책의 수립과 집행성과를 서술하고 분류 검색 및 키워드 검색을 기능을 지원하는 서비스와 정책의 배경과 추진 과정을 설명하고 그에 따른 성과와 사업을 제공하는 서비스가 있다. 또한 정책관련 콘텐츠를 수집하여 제공하는 정책정보 포털, 주요 이슈별 정책 보고서, 동영상, 이미지, 정책 기사, 관련 웹사이트 등을 최신 콘텐츠 위주로 선별하여 원문을 제공하는 정책 큐레이션 서비스와 수집한 경제정책을 주제별로 분류하여 시간 순으로 나열하는 서비스 등이 제공되고 있다. 기관에서는 기록에 대해 검색과 열람 서비스를 제공하고 있으며 그 외에도 기록 전시와 기록콘텐츠 개발이 이루어지고 있으나 기록연구지원에 대한 기능은 여전히 부족하다. Duff와 Yakel(2017)은 기록서비스는 기록전문가나 기록시스템에 의해 이루어지는 상호작용이며 이용자의 관심에 맞는 자료, 증거 혹은 정보를 이용자에게 의미 있고 활용할 수 있는 형태로 제공하는 것이라고 하였다. 이용자가 기록에서 쉽게 의미를 찾을 수 있는 도구를 제공하게 된다면 기존의 서비스에서 나아가 기록서비스의 패러다임 변화를 이끄는 주요한 서비스가 될 것이다.

기록정보서비스의 기록 콘텐츠 개발 방식을 살펴보면 주로 주제 중심형으로 개발되고 있음을 확인할 수 있다. 콘텐츠 설계자가 특정주제를 선정하고 그 주제에 따라 기록을 수집하여 제공한다. 국가기록원의 6.25전쟁, OECD 가입 컬렉션 등이 그 예이다. 주제 중심형 콘텐츠의 경우 의도가 실체를 왜곡할 가능성이 높고, 콘텐츠 설계자의 편향된 관점에 따라 기록 해석을 유도하는 경우도 발생할 수 있다(설문원, 2012). 설문원은 이에 의도의 표출을 최소화하여 맥락의 해석을 열어두거나 설계자의 관점과 의도를 명확히 밝혀 이용자가 주제적으로 콘텐츠를 평가하도록 해야 한다고 주장하였다. 경제정책 기록서비스도 이와 다르지 않다. 수집한 기록에 대해 단순 검색만을 제공하고, 설계자가 선정한 주제중심으로 모아둔 콘텐츠에서는 콘텐츠 설계자의 관점으로 재구성한 해석만을 보여주고 있다. 이용자는 주제 순, 시간 순으로 나열해놓은 기록 콘텐츠를 실제 필요한 맥락에 대한 이해를 없이 이용하게 된다. 주요한 경제정책기록에 효율적으로 접근하는 것에 어려움을 겪을 수밖에 없다. 이러한 현재의 기록서비스를 개선하기 위한 방안을 모색하는 것이 본 연구의 목적이다.

본 연구에서는 경제정책기록 중 경제정책방향을 대상으로 시기별 주요 경제정책방향을 맥락을 제공하여 이를 바탕으로 이용자가 기록정보에 대한 정보요구를 수립하고 정보요구에 적합한 기록에 접근할 수 있는 환경을 제공하고자 한다. 기록의 맥락정보에 대한 선제적인 제공을 통해 시기별 주요정책방향을 중심으로 정보요구를 수립할 수 있도록 유도할 수 있다.

이를 위해 주요 경제동향을 파악할 수 있는 기록을 선정하여 기록에서 배경, 주요내용, 원문을 수집하고 수집한 자료의 텍스트 마이닝을 통해 시기별 정책기록의 맥락정보를 제공한다. 각 분석 결과를 통해 도출한 내용을 이용자에게 제공하는 방식의 기록서비스를 제시하고자 한다.

1.2 연구방법

문자로 된 데이터에서 가치 있는 정보를 얻는 분석방법을 텍스트 마이닝이라고 한다. 텍스트 마이닝은 대량의 비정형 데이터에서 특정한 주제영역을 추출해 연구동향 분석을 수행하는데 유용하게 사용되고 있다. 네트워크 분석은 사회연결망의 개별 개체가 되는 개인이나 집단 간의 관계를 노드와 링크로 모델링해 그 관계를 계량적으로 분석하는 방법이다. 노드와 연결선으로 대상의 구조를 나타내고 연구대상의 형태에 따라 다양한 네트워크 구조가 생성된다. 경제정책기록에 텍스트 마이닝 분석을 사용하는 것은 장기간에 걸친 경제정책의 추세를 기록의

내용분석으로 통해 파악할 수 있는 적절한 방법이다.

경제정책시계열서비스에서 1991년도부터 2021년까지 수집한 경제정책방향 기록의 정책배경과 주요내용에서 키워드를 추출하여 텍스트를 수집한 뒤 텍스트 마이닝을 진행하여 빈도분석과 TF-IDF, 네트워크분석, 시계열 분석을 진행하고자 한다. 수집된 배경과 주요내용의 텍스트는 R 프로그램을 사용하였고 배경, 주요내용, 전문 텍스트는 파이썬을 사용하여 KoNLP 패키지의 형태소 분석 기능을 사용해 데이터를 전처리를 진행하였다. 텍스트 빈도분석을 통해 정권별, 시기별 가장 많이 나온 단어를 구한다. 가장 많이 사용된 단어 빈도표를 작성하여 정권별로 비교하여 흐름과 변화를 찾아내고자 한다. 빈도분석에 따른 시각화를 위해 결과를 정권별로 워드클라우드로 표현한다. 다음으로 네트워크분석을 진행해 연결중심성에 따른 단어 간 관계를 파악하고 주요키워드를 선정하여 시계열화 하여 어느 시점을 기준으로 단어의 언급양상 흐름을 한눈에 파악하는 시각화 자료를 만들고자 한다.

2. 이론적 배경

2.1 정책기록의 정의 및 특징

공공기록물 관리에 관한 법률에서는 공공기관이 업무와 관련하여 생산한 기록물 중에 국가적으로 보존할 가치가 있다고 인정되는 기록정보 자료를 법률에 의해 보호 관리하고 있다. 법령에서 일컫는 기록물이란 공공기관이 업무와 관련하여 생산하고 접수한 문서, 도서, 대장, 카드, 도면, 시청각물, 전자문서 등 모든 형태의 기록정보 자료와 행정박물을 말한다. 이에 따라 정책기록은 정부의 정책을 기록화한 자료로 공공기록물 관리에 관한 법률에 의해 보호되는 기록물이라 할 수 있다.

Schellenberg(1956)은 기록의 가치를 1차적 가치와 2차적 가치로 구분하였다. 1차적 가치는 일상의 행정업무를 지원하기 위한 용도의 행정적 가치, 법적 권리를 보호하기 위한 용도인 법적 가치, 재정상의 책임과 책무를 정립하기 위한 용도인 재정적 가치이며 1차적 가치는 주로 기록생산을 초래한 활동이나 행위가 종결되었을 때 상실된다. 2차적 가치는 법적 의미가 아닌 역사적 의미의 증거와 활동의 증거로서 증거적 가치보다 광범위하고 관료조직의 활동의 대상인 정보적 가치로 구분하였다. 정책기록은 이에 따르면 정부활동의 증거적, 정보적 가치를 가지고 있다. 정책기록은 우리나라 정부 활동의 역사적 증거가 되는 동시에 1차적 가치도 지닌다. 정책이 발표되면 이를 근거로 하여 부처에서 다양한 사업을 추진하기 때문이다.

정책이란 정부 또는 정치단체가 그들의 정치적, 행정적 목적을 앞으로 실현하기 위해 마련한 방책이나 방침을 일컫는다. 정부기관에 의해 공익을 실현하기 위한 정책목적은 달성하고자 시행되는 모든 정책수단의 지침으로 표현할 수 있다(서호준, 2019). 정책입안자가 생산한 정책기록은 정책결정자에게 정책의 입안, 결정, 평가 등을 위해 필요한 정보를 기록한 것으로 일반 국민 및 정책연구자가 정책에 대해 이해하기 위해 필요한 모든 정보를 담은 기록이다. 정부는 매년 다양한 분야의 정책을 발표한다. 정부가 발표하는 정책 중 경제정책은 국민들의 생활과 밀접한 관계를 가지고 있으며 정부 정책 중 가장 핵심이 되는 분야이다. 경제정책의 범위는 매우 광범위하며 그에 따라 복합적인 주제를 가지고 있다. 경제정책은 그 대상과 목표에 따라 미시경제정책과 거시경제정책으로 나누어지는데 경기침체를 완화하기 위한 거시경제정책을 경기부양정책으로 일컫고 장기적으로 경제효율을 높여서 성장잠재력을 높이는 정책을 미시경제정책이라 정의한다(현정택, 2006). 경제와 행정에 관한 연구들은 양적연구에 치중하여 설문자료와 통계분석에 치중하는 경향이 있으며 행정주체와 정책의 이해관계자의 의도와 인지와 같은 질적 연구가 소외되는 경향을 보이고 있다. 텍스트분석은 경제 분야의 방대한 자료의 체계적 분석이 어렵거나 관찰, 조사 등이 여의치 않을 때 유용하게 사용 할 수 있는 방법이다(김준현, 2015). 텍스트 분석에서 질적

내용분석을 수행하는 것은 방대한 텍스트에서 단순한 유사의미의 발견이나 분류를 수행하는 것이 아니며 분석을 통해 숨겨진 의도를 발견할 수 있다(Weber, 1990). 내용분석의 목표는 연구하고자 하는 현상을 이해하고 지식을 만드는 것이기 때문이다(Wamboldt, 1992).

정부정책이 기록된 정책기록을 이해하기 위해서는 정부정책이 어떻게 수립되는지 파악할 필요가 있다. 우선 국회의 입법 과정을 통하여 법률 규범이 만들어지게 된다. 법률 규범에 따라 법정계획이 존재하고 정부는 이에 따른 다양한 중장기 전략들을 수립하게 된다. 법정계획으로서 기본계획은 현재 329개이며 이와 유사한 종합계획이 122개가 존재한다. 이외에도 정부의 정책을 담은 수많은 계획과 로드맵들이 생산되고 있다. 450개의 법정계획들과 자체 계획까지 포함하면 1000여개의 정부부처의 분야별 계획이 존재하고 있다고 볼 수 있다. 이렇게 생산된 정책기록들은 각 부처의 보도 자료를 통해 배포된다. 그러나 모든 정책을 보도 자료를 통해 배포되지 않고 국민들에게 알려지지 않고 비공개로 진행되는 대책도 존재한다. 현재 국내 정책기록의 수집하는 기관으로는 국가기록원, 정책브리핑, 정책정보 포인트, KDI 경제정보센터가 있다.

2.2 경제정책방향

경제정책방향은 현재 기획재정부 소관으로서 연간 2회 주기로 발표되는 자료로 매년 12월에는 내년도 경제정책방향 발표되며 6-7월경에는 하반기 경제정책방향이 발표된다. 재정, 금융, 산업, 복지 등 경제전반에 영향을 끼치는 정책의 큰 방향을 제시하는 정책기록으로 거시경제와 미시경제정책 모두를 포함하고 있으며 정부가 경제의 청사진을 제시하기 위해 발표하는 정책이니 만큼 다른 정책에도 영향을 미치는 정책이다. 경제정책은 역사는 다음과 같이 시작된다. 경제정책을 이끌 추진력을 가진 부처가 필요하게 됨에 따라 1961년 경제기획원이 설립되었다. 경제를 기획하고 이를 실행하기 위해 예산 수립과 통계 작성을 합친 부서가 탄생되었다. 경제기획원 장관은 내각 서열 2위인 부총리가 각 경제부처를 소관 하였으며 경제성장률 목표를 세우고 목표달성여부를 점검하고, 외국에서 빌린 차관을 각 산업에 배분 하는 일, 산업의 우선순위를 골라 집중 투자하는 일 등 경제정책과 관련된 일을 책임지고 수행하였다. 경제기획원의 설립과 함께 1960년대부터 1989년까지 경제운용계획이 발표되었고 1982년과 1983년에는 경제운용방향이라는 이름으로 발표되었다. 이는 정부 주도에서 민간자율경제로 변화하여야 한다는 여론이 발생함에 따라 경제를 운용하되 방향을 제시한다는 의지를 반영한 것으로 해석할 수 있다. 1991년 경제운용방향이 등장하였고 1994, 1995년도에는 운용대신 운영이라는 단어를 사용하여 경제운영방향으로 이름을 변경하였다. ‘운영’과 ‘운용’의 차이점은 ‘운영’은 ‘학교, 당, 기업, 상점, 학회, 대회’ 등과 어울려 사용되며 ‘운용’은 ‘기금, 예산, 물품’ 등과 어울려 사용된다. 이러한 문맥상의 차이에 근거하면 두 단어는 무엇인가를 움직여 나간다는 점은 공통적인 의미를 가지고 있다. 그러나 ‘운영’은 조직이나 기구 등의 대상을 관리하면서 움직여 감을 의미하고 ‘운용’은 대상을 움직여 가면서 사용함을 의미한다. 이는 민주화가 뿌리내리기 시작되는 시대상을 비추어 경제운영이라는 더 넓고 추상적인 단어를 사용함으로써 정부가 경제에 개입하지 않고 큰 운영방향만을 제시하겠다는 의지로 해석할 수 있다. 1997년 김대중 정부부터는 경제정책방향을 사용하였으며 노무현 정부에서는 다시 경제운용방향을 사용하였다. 이명박 정부 2010년부터 경제정책방향을 사용하였다. 이후로 현재까지 계속 경제정책방향을 사용해오고 있다.

현재 경제정책방향은 과거에 비해 영향력이 떨어졌다는 의견도 있다. 정부가 경제정책방향을 제시하였으나 시간이 지난 후 그에 대한 평가가 제대로 이루어지지 않으며 추진하겠다는 언급한 정책이 무산되는 경우가 많아 국민에게 신뢰도를 저하시켰기 때문이다. 그러나 우리나라 경제는 여전히 정부의 역할이 큰 만큼 총체적인 정책방향의 제시는 필요하다는 입장도 존재한다.

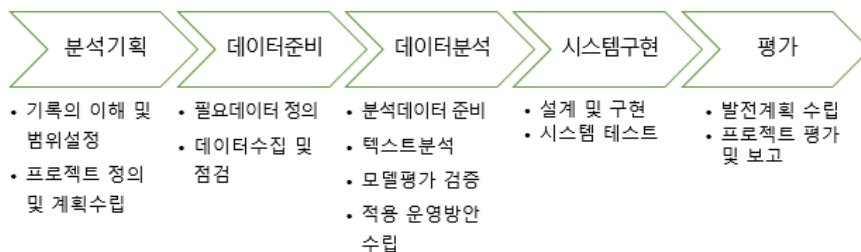
〈표 1〉 경제정책방향 대책명의 변화(1991~2003년도)

대책명	부처명	발표일
1991년 하반기 경제운용방향	경제기획원	1991.06.25.
1992년도 경제운용방향	경제기획원	1992.01.24.
1992년 하반기 경제운용방향	경제기획원 경제기획국 종합기획과	1992.07.09.
1993년도 경제운용방향	경제기획원 경제기획국 종합기획과	1992.12.29.
1994년도 경제운영방향: 「신경제 5개년계획」 제2차년도 추진계획	경제기획원 경제기획국 종합기획과	1994.01.11.
1995년도 경제운영방향	재정경제원 경제정책국 종합정책과	1995.01.09.
1995년 하반기 경제운영방향	재정경제원 경제정책국 종합정책과	1995.07.01.
1996년 경제운영방향	재정경제원 경제정책국 종합정책과	1996.01.05.
1996년 하반기 경제운영방향	재정경제원 경제정책국 종합정책과	1996.07.02.
1997년 경제정책방향	재정경제원 경제정책국 종합정책과	1997.01.15.
1998년 하반기 경제정책방향	재정경제부 경제정책국 종합정책과	1998.08.27.
1999년 경제정책방향	재정경제부 경제정책국 종합정책과	1998.12.29.
1999년 하반기 경제정책방향	재정경제부 경제정책국	1999.07.01.
2000년 경제정책방향	재정경제부 경제정책국 종합정책과	2000.01.17.
2000년 하반기 경제정책방향	재정경제부 경제정책국 종합정책과	2000.06.23.
2001년 경제운용방향	재정경제부 경제정책국 종합정책과	2000.12.29.
2001년 하반기 경제운용방향	재정경제부 경제정책국 종합정책과	2001.07.02.
2002년 경제운용방향	재정경제부 경제정책국 종합정책과	2001.12.24.
2002년 하반기 경제운용방향	재정경제부 경제정책국 종합정책과	2002.06.26.
2003년 경제운용방향	재정경제부 기획관리실	2003.01.08.

3. 경제정책기록서비스 연구 설계

3.1 방법론

본고에서는 빅데이터 분석 방법론의 5단계를 적용한 경제정책기록서비스 방법론을 제시하여 보았다. 분석기획 단계에서는 기록의 이해를 바탕으로 분석할 데이터의 범위를 설정한다. 또한 프로젝트의 정의와 계획을 설정해야 한다. 본 연구에서는 경제정책기록 중 경제정책방향을 선정하였으며 경제정책 방향에 대한 선행연구를 진행하였다. 데이터준비 단계에서는 필요한 데이터를 정의하고 데이터의 수집 및 점검이 이루어져야 한다. 경제정책방향을 정부별로 분류하였으며 배경설명과 요약자료, 원문 자료의 수집과 점검을 진행하였다.



〈그림 1〉 경제정책기록서비스 절차

데이터분석 단계에서는 분석용 데이터를 준비한다. 전처리를 통하여 데이터를 정제하고 텍스트 분석을 진행한다. 모델에 대한 평가와 검증을 진행하고 분석 결과를 운영할 방안을 수립한다. 시스템 구현 단계에서는 분석

결과를 어떻게 설계하고 구현하여 시스템 상에서 운영할 방안을 마련한다. 시스템에서 구현이 완료되었으면 시스템의 테스트 및 운영을 수행하여야 한다. 본 연구에서는 시스템의 구현 예시까지 제안하고자 한다. 평가단계에서는 분석 결과를 바탕으로 발전계획을 수립하고 프로젝트에 대한 평가와 보고가 이루어져야 한다.

3.2 분석기법별 기대효과

본 연구에서 사용된 텍스트 마이닝 분석기법에 대하여 예상되는 기대효과를 살펴보도록 한다. 빈도분석 기법은 분석 결과를 활용함으로써 방대한 양의 기록을 직접 찾아 읽지 않아도 가장 많이 높이 등장하는 단어 활용해 주요장점을 빠르고 쉽게 파악할 수 있으며 정책의 변화과정을 관찰하여 흐름을 한눈에 보여줄 수 있다. TF-IDF 기법은 특정 단어가 문서 내에 얼마나 자주 등장하는지 나타나는지를 확인하여 문서에서 단어의 중요도 순으로 기록의 내용을 파악할 수 있다. 네트워크분석 기법은 정책 간 연결 정도를 확인하여 정책 간의 연계와 상관관계를 파악하는 서비스를 제공할 수 있다. 시계열분석 기법은 시간의 흐름에 따라 필요한 키워드를 선정하여 흐름과 변화 양상을 시각적으로 한눈에 파악할 수 있다.

4. 텍스트 마이닝 분석

4.1 분석설계

4.1.1 자료수집

앞서 언급하였듯이 1991년부터 2021년까지 수집한 총 58개의 경제정책방향 기록을 정부별로 구분하면 다음과 같다. 노태우 정부는 4회의 경제운용방향을 발표하였으며 김영삼 정부는 6회, 김대중 정부 10회, 노무현 정부 11회, 이명박 정부 10회, 박근혜 정부 9회, 문재인 정부는 21년까지 8회의 경제정책방향을 수집하였다. 경제정책시계열서비스에서 제공하고 있는 정책 배경설명과 요약자료와 기록 원문 텍스트를 수집하여 분석 대상으로 하였다.

4.1.2 전처리

전처리작업을 위해 먼저 리스트 형태로 단어의 토큰화를 진행하였다. 고유명사, 일반명사를 제외한 용어를 불용어 처리하였다. 또한 합성명사의 식별을 위해 바이그램(Bigram)을 진행하여 명사명사 형태단어에 대하여 상호정보량(MI: Mutual Information)을 구하여 적정수준이상의 MI 값을 가지는 단어를 합성명사로 보고 다음과 같은 단어를 추가하였다. ‘건강’, ‘보험’(MI:130) ‘공정’, ‘거래’(MI:112) ‘창업’, ‘지원’(MI:102) ‘대외’, ‘경제’(MI:89)와 같이 계산된 값을 합친 499개의 단어를 합성어 단어사전으로 구성하여 추가하였다. 또한 부처명칭은 삭제하였다. 전처리한 단어들은 정부별 분석을 위해 정부별로 데이터를 분리하였다.

4.1.3 분석도구

R은 수리나 통계분석도구, 프로그래밍 언어 및 시각화 도구의 기능을 지닌 오픈 소스의 소프트웨어 프로그램이다. 다양한 운영체제에서 사용가능하며 빠른 분석속도와 시각화 기능을 가지고 있다. 패키지 중 형태소 분석 라이브러리인 KoNLP 패키지는 R에서 한국어 데이터의 자연어 처리를 지원한다.

파이썬 또한 R과 유사한 특징을 가지고 있는 프로그램이며 파이썬은 통계분석 외에 다른 부분에서 상대적으로 장점을 가지고 있으며 최신 머신러닝과 딥러닝 알고리즘을 제공한다. 초보자가 배우기 더 쉽다는 장점을 가지고 있으며 데이터전처리 부분과 속도에서 R보다 장점을 지니고 있다. 본 연구에서는 R로는 배경, 주요내용의 키워드

빈도분석을 진행하였고 파이썬으로는 배경, 주요내용, 본문을 포함한 모든 텍스트의 전처리를 수행하고 이어 TF-IDF, 네트워크분석, 시계열분석을 진행하였다.

4.2 분석결과

4.2.1 키워드 빈도분석

배경설명과 내용요약 텍스트의 전처리 진행 후 불용어를 삭제하고 총 687개의 단어리스트가 생성되었다. 빈도분석에서는 본문을 제외하고 배경과 주요내용 텍스트만으로 진행하였다. 배경, 요약정보는 본문에서 중요한 주제가 모여 있게 된다. 즉 본문에서 이야기하는 주제를 중심으로 인위적으로 요약한 결과이기 때문에 빈도분석에서 주요한 단어를 더 정확하게 볼 수 있는 결과를 얻을 수 있다. 배경과 요약정보 대한 빈도분석결과 ‘일자리’가 빈도 131회로 가장 높았으며 그다음으로 ‘경쟁력’이 110회, ‘구조조정’이 91회로 나타났다. <표 2>는 상위 30개의 단어를 정리하였다. 노동, 고용에 관한 단어와 기업, 산업, 대외경제 분야의 단어들 상위권에 나타난 것을 확인할 수 있다.

<표 2> 경제정책방향 키워드빈도

순위	단어	빈도	순위	단어	빈도
1	일자리	131	16	글로벌	34
2	경쟁력	110	17	서비스	34
3	구조조정	91	18	노동시장	32
4	중소기업	89	19	거시경제	31
5	가계소득	76	20	경제정책방향	31
6	선진화	69	21	경제협력	31
7	구조개혁	58	22	금융시장	31
8	리스크	45	23	외국인	30
9	중장기	45	24	노사관계	29
10	시스템	4	25	벤처기업	29
11	내실화	42	26	성장동력	29
12	서비스산업	42	27	효율화	28
13	경기회복	41	28	근로자	25
14	인프라	40	29	성장잠재력	25
15	부동산	35	30	외환시장	24



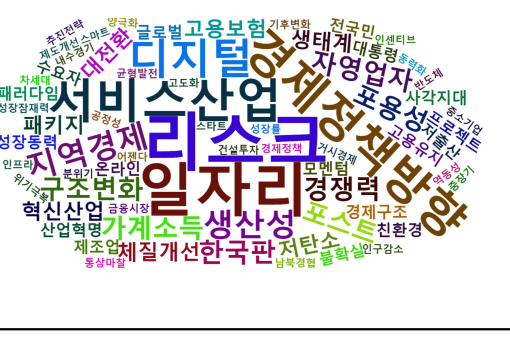


<그림 2> 경제정책방향 출현 키워드 워드클라우드

빈도분석 결과를 각 정부 별 가장 높은 빈도의 단어를 10위까지 표시하였으며 전체 결과를 워드클라우드 형태로 시각화하였다.

〈표 3〉 정부별 단어출현빈도 및 워드클라우드

정권	순위	단어	빈도	워드클라우드
노태우	1	국제수지	7	
	2	산업경쟁력	7	
	3	경쟁력	5	
	4	고도화	4	
	5	산업구조	4	
	6	중소기업	4	
	7	건설투자	3	
	8	구조조정	3	
	9	부동산투기	3	
	10	자율화	3	
김영삼	1	경쟁력	16	
	2	중소기업	10	
	3	지방자치단체	10	
	4	선진화	9	
	5	생산성	8	
	6	노사관계	6	
	7	사회간접자본	6	
	8	구조조정	5	
	9	금융개혁	5	
	10	임금안정	5	
김대중	1	구조조정	38	
	2	경쟁력	23	
	3	구조개혁	23	
	4	내실화	18	
	5	서민층	14	
	6	금융기관	13	
	7	벤처기업	13	
	8	외국인	13	
	9	시스템	11	
	10	거시경제정책	10	
노무현	1	경쟁력	37	
	2	선진화	37	
	3	중소기업	33	
	4	일자리	24	
	5	구조조정	22	
	6	가계소득	21	
	7	내실화	18	
	8	시스템	17	
	9	서비스산업	16	
	10	중장기	16	

정권	순위	단어	빈도	워드클라우드
이명박	1	일자리	43	
	2	가계소득	24	
	3	글로벌	22	
	4	중소기업	20	
	5	선진화	17	
	6	경쟁력	16	
	7	경기회복	13	
	8	거시경제	12	
	9	불확실	12	
	10	서비스산업	12	
박근혜	1	일자리	45	
	2	리스크	26	
	3	구조개혁	20	
	4	가계소득	17	
	5	구조조정	11	
	6	노동시장	11	
	7	중소기업	10	
	8	서비스	9	
	9	경쟁력	8	
	10	민주화	8	
문재인	1	리스크	12	
	2	일자리	11	
	3	서비스산업	9	
	4	디지털	8	
	5	생산성	6	
	6	지역경제	6	
	7	가계소득	5	
	8	경쟁력	5	
	9	구조변화	5	
	10	자영업자	5	

노태우 정부에서는 ‘국제수지’가 가장 많이 나타났으며 ‘산업경쟁력’, ‘산업구조’, ‘건설’, ‘부동산투기’가 나타났다. 노태우 정부에서는 1986년 발생한 국제수지 흑자가 인플레이션 압력을 야기하였고 부동산가격이 폭등하고 민간소비가 증가하여 물가상승이 나타난 바 있다. 김영삼 정부에서도 ‘경쟁력’이 가장 많이 나타났으며 ‘지방자치단체’와 김영삼 정부의 금융실명제 정책과 관련 있는 ‘금융개혁’이 상위권에 위치한다. 김대중 정부에서는 IMF의 배경으로 ‘구조조정’이 가장 높은 빈도를 보였고 벤처 붐과 함께 ‘벤처기업’도 나타났다. 노무현 정부의 경제정책기조는 반시장적, 정부 개입적, 복지 지향적이었던 평가를 받는다. 노무현 정부에서는 ‘경쟁력’, ‘선진화’, ‘중소기업’, ‘가계소득’, ‘서비스산업’이 높은 출현 빈도를 보였다. 이명박 정부는 ‘일자리’, ‘가계소득’, ‘글로벌’, ‘경기회복’이 나타났다. 이 시기는 2008년 경기가 둔화세를 보이며 미국의 서브프라임 모기지 사태로 글로벌 금융위기가 촉발되어 급격히 경제가 위축되었다. 실물경기가 침체되고 고용사정이 악화되며 ‘일자리’가 가장 높은 빈도를 보인 단어로 나타났다. 박근혜 정부도 글로벌 금융위기 여파가 계속되는 가운데 경제의 창조와 혁신을 강조하였으며 이를 위한 기업환경개선 정책이 추진되었으며 ‘일자리’, ‘리스크’, ‘중소기업’, ‘노동시장’ 등이 높은 빈도를 보였다. 문재인 정부에서는 ‘리스크’, ‘일자리’, ‘디지털’, ‘지역경제’, ‘자영업자’ 등이 높은

빈도로 나타났다. 공통적으로 ‘일자리’가 역대정부에서 높은 빈도를 보였으며 ‘경쟁력’, ‘구조조정’, ‘중소기업’ 등이 다수 출현하였다.

4.2.2 TF-IDF

TF-IDF는 문서 내에 특정단어의 중요도를 나타내는 가중치로 TF(단어빈도, term frequency)는 특정단어가 문서 내에 얼마나 자주 등장하는지 나타내는 값이다. 이 값이 높을수록 문서에서 중요하다고 생각할 수 있다. 그러나 특정단어 하나의 문서에만 나오지 않고 자주 등장하면 단어의 중요도는 낮아진다. DF(문헌빈도, document frequency)의 값의 역수를 IDF(역문헌 빈도, inverse document frequency)라 한다. TF-IDF는 TF와 IDF를 곱하여 점수가 높은 단어가 다른 문서에는 많지 않고 해당문서에 자주 등장하는 단어임을 의미한다.

이는 단어의 출현빈도만으로 단어의 가중치를 결정하는 것에 대한 한계를 개선하기 위한 방법으로 본 연구에도 적용하여 사용하였다. 정부별로 발표한 문서를 합쳐 계산하였다. 여러 문서를 합쳐 값을 계산하는 경우 단어가 출현한 문서의 수가 적을수록 높은 IDF값을 받는 경우가 생긴다. 이에 문서수도 함께 표시하였다.

〈표 4〉 정부별 TF-IDF

정권	순위	단어	TF	IDF	문서수	스코어
노태우	1	토지	3	1.386294	1	4.158883
	2	이용	2.5	1.386294	1	3.465736
	3	개혁	4	0.693147	2	2.772589
	4	소비재	1.75	1.386294	1	2.426015
	5	부진	1.75	1.386294	1	2.426015
	6	남북	1.5	1.386294	1	2.079442
	7	기간	1.5	1.386294	1	2.079442
	8	활용	3	0.693147	2	2.079442
	9	연구	1.5	1.386294	1	2.079442
	10	농어촌	6.75	0.287682	3	1.941854
김영삼	1	직업훈련	4.333333	1.098612	2	4.760653
	2	훈련	5.5	0.693147	3	3.812309
	3	세계	9.333333	0.405465	4	3.784341
	4	시대	2	1.791759	1	3.583519
	5	wto	3.166667	1.098612	2	3.478939
	6	지역	4.833333	0.693147	3	3.350211
	7	통신	3	1.098612	2	3.295837
	8	기술	7.5	0.405465	4	3.040988
	9	선진	2.5	1.098612	2	2.746531
	10	부두	1.5	1.791759	1	2.687639
김대중	1	IT	13.3	0.356675	7	4.743777
	2	월드컵	4.4	0.693147	5	3.049848
	3	국가균형발전	1.3	2.302585	1	2.993361
	4	해양	4.3	0.693147	5	2.980533
	5	DDA	2	1.203973	3	2.407946
	6	동북아	4.6	0.510826	6	2.349798
	7	지역	22.3	0.105361	9	2.349539
	8	종합	9.9	0.223144	8	2.209121
	9	신용회복	1.7	1.203973	3	2.046754
	10	복지	8.3	0.223144	8	1.852091

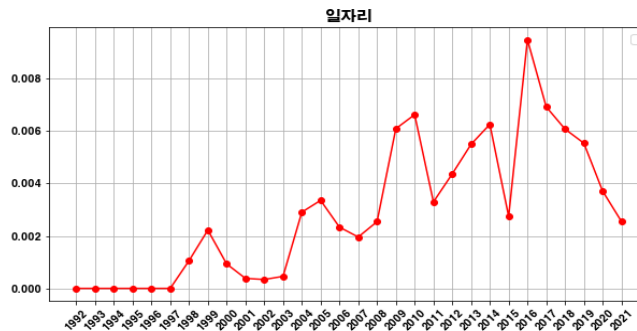
정권	순위	단어	TF	IDF	문서수	스코어
노무현	1	상생협력	6.363636	0.606136	6	3.857228
	2	통계	11.18182	0.318454	8	3.560892
	3	한미	2.727273	1.299283	3	3.543499
	4	사업전환	4.454545	0.788457	5	3.512219
	5	부업	2.636364	1.299283	3	3.425382
	6	서브프라임	1.272727	2.397895	1	3.051867
	7	채권	8.818182	0.318454	8	2.808183
	8	물류	28.63636	0.09531	10	2.729337
	9	brics	4.363636	0.606136	6	2.644956
	10	장애인	5.818182	0.451985	7	2.629732
이명박	1	FTA	9.2	0.693147	5	6.376954
	2	동반성	6.7	0.693147	5	4.644086
	3	창조경제	1.7	2.302585	1	3.914395
	4	경제민주화	1.5	2.302585	1	3.453878
	5	다문화	6.7	0.510826	6	3.422532
	6	가족	6.6	0.510826	6	3.371449
	7	봉사	2.3	1.203973	3	2.769137
	8	국정과제	1.7	1.609438	2	2.736044
	9	농어	7.4	0.356675	7	2.639395
	10	부실채권	1.6	1.609438	2	2.575101
박근혜	1	메르스	3	2.197225	1	6.591674
	2	사업	12.33333	0.405465	6	5.000736
	3	확대	39.44444	0.117783	8	4.645886
	4	협업	5.666667	0.81093	4	4.595271
	5	해외	7.666667	0.587787	5	4.506364
	6	부처	7.666667	0.587787	5	4.506364
	7	점검	11.11111	0.405465	6	4.505168
	8	분야	10.77778	0.405465	6	4.370013
	9	현장	6.555556	0.587787	5	3.853268
	10	제도	9.333333	0.405465	6	3.784341
문재인	1	코로나	29	0.980829	3	28.44405
	2	방역	16.75	0.980829	3	16.42889
	3	그린	19.125	0.693147	4	13.25644
	4	녹색	9	0.980829	3	8.827463
	5	디지털	29.25	0.287682	6	8.414701
	6	비대면	15.125	0.470004	5	7.108805
	7	감염병	6.875	0.980829	3	6.743201
	8	탄소	13.625	0.470004	5	6.403799
	9	중립	6.5	0.980829	3	6.37539
	10	한국판뉴딜	5.75	0.980829	3	5.639768

TF-IDF를 통하여 문헌에 가중치를 부여한 후 나타난 결과는 다음과 같다. 본 결과는 각 정부에만 특징적으로 보이는 단어들이 구체적으로 나타나게 된다. 노태우 정부에서는 ‘토지’가 가장 높은 스코어를 나타냈다. 그 뒤를 이어 ‘개혁’, ‘소비재’, ‘남북’, ‘농어촌’이 나타났다. 노태우 정부에서 시행한 토지공개념 제도와 시장개방이 주요 주제였음을 알 수 있다. 김영삼 정부에서는 ‘직업훈련’, ‘세계’, ‘시대’, ‘WTO’가 나타났는데 1995년 WTO의 출범과 함께 세계화가 주요 주제였음을 알 수 있다. 김대중 정부에서는 ‘IT’, ‘월드컵’, ‘국가균형발전’, ‘동북아’, ‘신용

1991년부터 2021년까지 전체정부의 단어에 대하여 네트워크 분석결과 핵심 주제어 67개를 선정하였다. 연결정도(degree)에 따라 높은 순으로 나열하면 다음과 같다. ‘사업’(1512), ‘지역’(1067), ‘일자리’(711), ‘물류’(556), ‘청년’(493), ‘관광’(488), ‘에너지’(477), ‘취업’(423), ‘뉴딜’(397), ‘문화’(390), ‘해외’(384), ‘스마트’(358), ‘통계’(349), ‘펀드’(327), ‘선정’(308), ‘부처’(285), ‘서비스산업’(278), ‘공공기관’(268), ‘디지털’(257), ‘인프라’(244) 등의 순서로 나타났다.

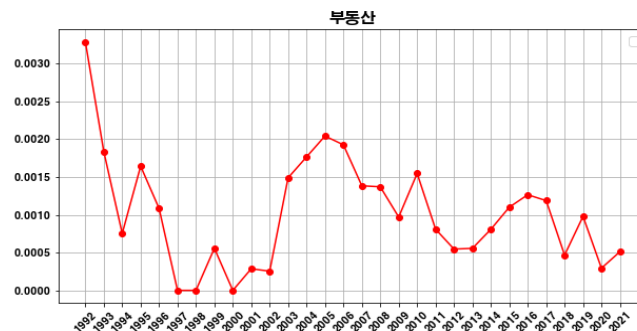
4.2.4 시계열분석

TF-IDF 스코어가 높은 단어 중 관심도가 높은 ‘일자리’, ‘부동산’, ‘기업’에 대하여 시계열 분석을 진행하였다. 시계열 분석의 추세는 해당연도 자료에 언급된 단어 중 비교단어의 상대비율로 제시하였다. 먼저 ‘일자리’가 언급된 것은 1998년부터이다. 1998년은 IMF를 겪은 해로 이 시기부터 ‘일자리’가 경제정책방향에 언급 되었다. 1998년에 0.001057, 1999년에 0.002224의 수치를 보이며 상승하였으며 이후 2004년에 0.002909로 상승하였다. 이후 2000년대 중반부터는 평균적으로 높은 비율을 보여주고 있으며 2009년, 2014년, 2016년에 상승곡선을 보이고 있다. 일자리는 평균 0.0029의 수치를 보였다.



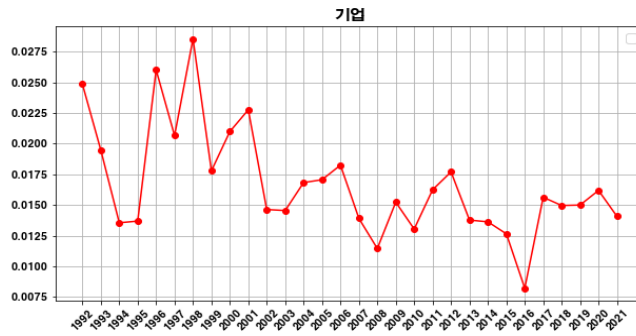
〈그림 4〉 일자리 시계열 그래프

‘부동산’에 관한 언급은 평균 0.001 일자리와 달리 1992년 노태우 정부부터 가장 높은 비율로 나타났다. 노태우 정부는 1989년 긴급부동산투기억제대책을 발표한 바 있다. 1995년, 1999년에 상승하였으며 2003년부터 2008년 동안 높은 비율을 보여주고 있다. 부동산 폭등이 있던 2005년 2006년에 높게 나타난다. 이후로는 하락곡선을 보이다 2010년에 한 번 더 높아진다. 이후 문재인 정부 2019년 부동산 폭등 현상과 함께 경제정책방향에서 ‘부동산’의 출현비율은 0.0055를 기록하였다.



〈그림 5〉 부동산 시계열 그래프

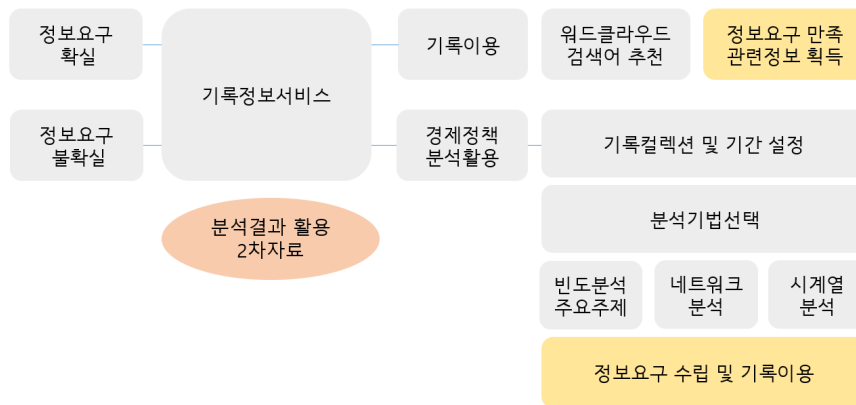
‘기업’에 대한 언급은 평균 0.016712의 비율로 나타나고 있다. 1992년 0.024854로 상당히 높은 비율을 보여주고 있다. 이후 1996년 0.026012 1998년 0.028541의 비율을 보였다. 2000년대 들어서는 평균 0.016567 2010년대는 0.014261로 90년대에 비해 점점 상대비율이 낮아지는 경향을 보인다.



〈그림 6〉 기업 시계열 그래프

5. 경제정책기록정보서비스 모형제안

본 장에서는 경제정책기록서비스의 2장에서 언급한 방법론의 4단계로 시스템을 구현하는 방안을 제시하였다. 먼저 기록의 이용자의 구체적 정보요구가 있는 경우와 구체적인 정보요구가 없는 경우에 대하여 제안해보았다.



〈그림 7〉 정책기록정보서비스 모형

구체적인 정보요구가 있는 경우 이용자는 검색기능을 통해 구체적인 기록컬렉션에 접근할 수 있다. 네트워크분석결과를 통해 주제가 검색되면 검색어와 연결 중심성이 높은 단어를 검색어로 추천하는 기능을 제공한다. 또한 특정 기록이 검색되었다면 검색결과에 따른 고빈도의 단어의 워드클라우드를 함께 제시하여 제공할 수 있다. 이용자는 정보요구의 만족을 얻게 되고 추가 관련 정보의 획득을 통해 정보요구를 확장할 수 있다.

기록에 대한 요구는 있으나 구체적인 정보요구가 없는 이용자에게는 경제정책분석기능을 통해 기록의 맥락을

서비스를 제공하여야 한다. 정책기록이 정책연구의 핵심 연구 자료로 인식되며 정책기록서비스가 가치를 지니고 있다는 것을 입증하기 위해서는 이에 대한 이해와 연구가 필요하다. 본 연구에서는 정책기록만을 대상으로 분석하였으나 정책기록과 사회현상, 법령, 제도 등과 같은 데이터와 연계한다면 더 나은 정책기록서비스를 제공할 수 있을 것이다.

본 연구에서 제안한 정책정보서비스를 통하여 정책입안자는 보다 현실적이고 일관성뿐만 아니라 연속성을 지닌 정책을 제시하는 데 도움이 될 것이다. 정책연구자들은 정책기록을 분석함으로써 정책 생태계의 혁신을 위한 발전방안을 도출할 수 있을 것이다. 본 연구가 시발점이 되어 정책기록에서 정책 이슈를 발견하고 그 연계 과정에 대한 분석이 이루어진다면 정부 정책 간의 상호연계가 강화되어 정권이 교체되더라도 지속적 연계가 가능한 정책이 수립될 수 있을 것이다.

참고문헌

- 강범일, 송민, 조화순 (2013). 토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구. 한국문헌정보학회지, 47(4), 315-334.
<https://doi.org/10.4275/KSLIS.2013.47.4.315>
- 국립국어원 (2021.11.3.). 온라인가나다. 국립국어원. 출처: <https://www.korean.go.kr/front/main.do>
- 김수현, 이영준, 신진영, 박기영 (2020). 거시경제 분석을 위한 텍스트 마이닝. 한국경제의 분석, 26(1), 1-70.
<http://dx.doi.org/10.22823/jkea.26.1.202004.1>
- 김연지 (2021). 문헌정보학의 메타데이터 분야 연구동향 분석. 석사학위논문, 상명대학교 일반대학원 문헌정보학과.
- 김준현 (2015). 네트워크 텍스트 분석결과 해석에 관한 소고: 행정학 분야 연구를 중심으로. 인문사회과학연구, 16(4), 247-280.
<https://doi.org/10.15818/ihss.2015.16.4.247>
- 대통령기록관 (2013). 정책기록. 기록컬렉션. 출처: <https://www.pa.go.kr/research/contents/policy/policymain.jsp>
- 박은태 (2014). (현대) 경제학사전. 경기도: 경연사.
- 박준형, 류범모, 오효정 (2018). 시계열 기반 국내 기록관리학 토픽 트렌드 분석. 한국기록관리학회지, 18(1), 29-47.
<https://doi.org/10.14404/JKSARM.2018.18.1.029>
- 송해지, 박경수, 정혜은, 송민 (2013). 텍스트 마이닝 기법을 활용한 한국의 경제연구 동향 분석. 제20회 한국정보관리학회 학술대회논문집, 47-54.
- 신동현 (2017). TF-IDF 기반 이직 대상 기업 추천 시스템 설계 및 구현. 석사학위논문, 숭실대학교.
- 이재운, 문주영, 김희정 (2007). 텍스트 마이닝을 이용한 국내 기록관리학 분야 지적구조 분석. 한국문헌정보학회지, 41(1), 345-372.
- 조재인 (2011). 네트워크 텍스트 분석을 통한 문헌정보학 최근 연구 경향분석. 정보관리학회지, 28(4), 65-83.
<https://doi.org/10.3743/KOSIM.2011.28.4.065>
- 최광 (2018). 기적의 한국경제 70년사. 파주: 북앤피플.
- 한국개발연구원 (2020). 알기 쉬운 한국경제 발전사. 세종: 한국개발연구원.
- 한국개발연구원 (2020.9.21.). 경제정책방향. 경제정책시계열서비스. 출처: <http://epts.kdi.re.kr/polcTmsesSrvc/them#this>
- 한국개발연구원 (2021). 50주년 기념 KDI 비전보고서. 세종: 한국개발연구원.
- 현정택 (2006). 경제정책에 대한 이해 및 최근 경제상황에 대한 정책적 함의. KDI 정책포럼, 172.
- Duff, W. & Yakei, E. (2017). Archival interaction, Currents of Archival Thinking(2nd),
- Schellenberg, T. S. (1956). Mordern archives: Principles and Techniques
- Wamboldt, B. D. (1992). Content analysis: Method, applications, and issues, 14, 313-321.
<https://doi.org/10.1080/07399339209516006>
- Weber, R. P. (1990). Basic Content Analysis

• 국문 참고자료의 영어 표기

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

- Cho, Jane (2011). A study for research area of library and information science by network text analysis. *Journal of the Korean Society for information Management*, 28(4), 65-83 <https://doi.org/10.3743/KOSIM.2011.28.4.065>
- Choi, kwang (2018). *The Miracle of the Korean Economy 70 Years*, Paju: Book&People.
- Hyun, Jung-taek (2006). Understanding of economic policy and policy implications on recent economic conditions. *KDI Policy Forum*, 172.
- 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. <https://doi.org/10.4275/KSLIS.2013.47.4.315>
- Kim, Jun hyeon (2015). An essay for understanding the meaning of the network text analysis results in study of the public administration. *Institute for Humanities and Social Sciences*, 16(4), 247-280. <https://doi.org/10.15818/ihss.2015.16.4.247>
- Kim, Soo hyon, Lee, Young Joon, Shin, Jhinyoung, & Park, Ki Young (2020). Text mining for economic analysis. *Journal of Korean Economic Analysis*, 26(1), 1-70. <http://dx.doi.org/10.22823/jkea.26.1.202004.1>
- Kim, Yeon ji (2021). A Study on the Trend Analysis of Metadata in Journals of Library and Information Science: Using Network Analysis Techniques. Master's thesis, Sangmyung University.
- Korea Development Institute (2020). *A Clear History of Korean Economic Development*. Sejong: Korea Development Institute.
- Korea Development Institute (2020.9.21.) Economic policy directions. Economic policy timeline. Available: <http://epts.kdi.re.kr/polcTmsesSvc/them#this>
- Korea Development Institute (2021). *50th Anniversary Vision Report*. Sejong: Korea Development Institute.
- 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.
- National Institute of the Korean Language (2021.11.3.). Onlineganada. Available: <https://www.korean.go.kr/front/main.do>
- Park, Euntae (2004). *A Dictionary of Economics*. Seoul: Kyungyeonsa.
- Park, Jun Hyeong, 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*, 12(1), 29-47. <https://doi.org/10.14404/JKSARM.2018.18.1.029>
- Presidential Archives (2013). Policy record. Archive collection. Available: <https://www.pa.go.kr/research/contents/policy/policymain.jsp>
- Sin, Dongheon (2018). Implementation of Moving Job-Recommender System Based on TF-IDF: Focusing on reviews posted on the Jobplanet, Master's thesis, Soongsil University.
- Song, Hye-Ji, Park, Kyoung-Soo, Jung, Hye-Eun, & Song, Min (2013). Trend analysis of korean economy in the economic literature by text mining techniques. *Proceedings of the Korean Society for Information Management Conference*, 20, 47-50.

