

텍스트마이닝을 활용한 송례문 관련 기사의 트렌드 분석

Trend Analysis of News Articles Regarding Sungnyemun Gate using Text Mining

김민정*, 김철주**

숙명여자대학교 소비자경제학과*, (주)동방문화유산**

Min-Jeong Kim(min-jeong.kim@sm.ac.kr)*, Chul Joo Kim(eckehomo@hanmail.net)**

요약

국보 제1호인 송례문은 2008년 2월 10일 화재로 일부가 소실되었으나 화재 이후 복구 작업을 통해 2013년 5월 4일 시민에게 공개되었다. 이로 인해 송례문은 국가적으로 큰 이슈가 되어 언론의 관심을 받으며 동시에 많은 연구의 대상이 되었다. 본 연구는 문화재로서 송례문을 키워드로 하여 2002년부터 2016년까지 신문 기사에 대한 빈도분석을 통해 송례문 관련 어떤 키워드들이 자주 나타나고 있는지에 대해 파악하였다. 또한 추출된 송례문 관련 키워드들간 연관관계 분석을 통해 키워드간 연결의 맥락을 파악하고 분석하였다. 다음으로 송례문 화재 전후, 언론사별 주요 키워드 추출을 통해 공통점과 차이점을 보여줌으로써 관점의 다양성을 제공하였다. 본 연구를 통해 문화재로서 송례문 관련 키워드는 화재 이후에 나타난 키워드가 전체 기사에서 고빈도어로 나타남을 알 수 있었고 몇 가지 키워드간 상관관계가 높게 나타났다. 또한 화재 전후 키워드에는 명확한 차이를 보이고 있었으며 언론사별 키워드에서 상위 키워드들은 명확한 차이는 보여주지 않았지만 차상위 키워드들은 차이가 발생하여 언론사별로 주로 다루어진 기사들의 내용은 차이가 있다는 것을 발견했다. 본 연구는 문화재로서 송례문 관련 기사에 대해 텍스트마이닝 기법을 활용하여 다량의 데이터를 분석하였음에 의미가 있으며 정보생산자 및 정보소비자들에게 송례문 관련 기사의 동향과 정보를 제공할 수 있을 것이라 기대한다.

■ 중심어 : | 송례문 | 문화재 | 신문기사 | 텍스트마이닝 | 군집분석 | 연관관계 분석 |

Abstract

Sungnyemun Gate, Korea's National Treasure No.1, was destroyed by fire on February 10, 2008 and has been re-opened to the public again as of May 4, 2013 after a reconstruction work. Sungnyemun Gate become a national issue and draw public attention to be a major topic on news or research. In this research, text mining and association rule mining techniques were used on keyword of newspaper articles related to Sungnyemun Gate as a cultural heritage from 2002 to 2016 to find major keywords and keyword association rule. Next, we analyzed some typical and specific keywords that appear frequently and partially depending on before and after the fire and newspaper companies. Through this research, the trends and keywords of newspapers articles related to Sungnyemun Gate could be understood, and this research can be used as fundamental data about Sungnyemun Gate to information producer and consumer.

■ keyword : | Sungnyemun Gate | Cultural Heritage | News Articles | Text Mining | Clustering | Association Rule Mining |

I. 서론

송례문은 조선이 개경에서 한양으로 도읍을 옮기면서 한양도성의 축성 시에 남쪽의 정문으로 세워진 문루로서 한양정도 이후 600여 년 동안 우리나라의 역사와 함께 해왔다[1]. 600년 이상의 역사는 현존하는 서울의 목조건물 중 가장 오래된 건물로 남게 되었고 1962년 12월 20일 국보 제1호로 지정되어 우리나라의 문화유산으로서 가치가 매우 크다고 할 수 있다. 그러나 송례문은 2008년 2월 10일 밤 8시 50분 방화로 인해 화재가 발생하여 다음날 새벽 2시경 진화는 되었지만 이때 발생한 화재로 흥예기단과 일층문루를 남기고 이층은 전소되었다[1].

이러한 송례문 화재사고는 문화재 관리에 종사하는 공직자, 소방대원뿐만 아니라 온 국민들이 모두 충격을 받았는데 당시의 방송, 신문은 국가적 최대 비극이 발생한 것으로 보고 기사보도는 물론, 논평, 사설, 칼럼 등에서도 소실상황 및 원인규명 그리고 앞으로의 개선방안을 제시하면서 이 사건을 국가적 재난이라고 보도하였다[2]. 이와 같이 국민들의 문화유산에 대한 의식과 언론의 문화재 전반에 대한 기사들은 정부로 하여금 문화재정책을 새롭게 생각하는 기회도 만들었다[3].

2008년 5월 20일, 문화재청장은 송례문 복구 기본원칙을 발표하는데 문화재청은 이때 공식적으로 ‘송례문 복구’라는 용어를 사용했다[4]. 송례문 복구 기본원칙은 불탄 “성문은 화재 전 모습대로 복구”하고 화재 전 일제 때 이미 없어진 성곽과 변형된 지반은 “원형대로 복원한다”고 ‘복구’와 ‘복원’을 구분했다. 즉 ‘복구’는 화재 전으로, ‘복원’은 원형으로 되돌리는 것임을 알 수 있다. 이러한 송례문 복구 기본원칙에 따라 2008년 6월부터 2009년 12월까지 문헌조사 및 발굴조사 같은 고증조사를 실시하고 이를 바탕으로 복구를 위한 설계를 진행하였으며 2010년 1월부터 2012년 12월까지 복구공사를 실시하였다[4]. 이와 같이 송례문은 약 3년간의 복구 공사를 마치고 2013년 5월 4일 복구기념식과 함께 시민에게 공개되었으며 올해 6월 23일부터는 휴무일을 제외하고 매일 송례문 광장에서 파수의식을 재현하고 있다. 이와 같이 화재 이후 송례문은 문화재 전문가들만의 것이 아

니라 시민의 것이었으며 이는 언론의 힘도 큰 역할을 했다고 볼 수 있다. 2000년부터 2016년 현재까지 ‘송례문’과 ‘문화재’를 키워드로 보도된 일간지 기사는 2448건[5]으로 그중 230건만 송례문 화재 전(2008년 2월 9일까지) 기사이고 나머지는 화재 후 보도된 것만 봐도 화재 이후 많은 기사들이 생산된 것을 알 수 있다.

그러므로 문화재로서 송례문에 대한 신문기사 분석을 통해 의미있는 지식을 발견한다면 문화재 전문가들뿐만 아니라 국민들의 역사문화의식이 더욱 확대될 수 있을 것이라 사료된다. 현재까지 신문기사로부터 문화재 관련 ‘용어’에 대하여 텍스트마이닝 기법을 활용한 분석은 거의 이루어지지 않았다. 그러나 일정 기간 동안 신문기사를 수집하여 전통목조건축 화재실태를 화재의 발생빈도, 발화요인, 진압과정 측면에서 구체적으로 내용을 분석한 연구[6]는 있다.

따라서 본 연구에서는 신문기사로부터 문화재로서 송례문 관련 용어의 분석을 통해 어떤 용어가 핵심적이고 쟁점이 되는지를 파악해본다. 이를 위해 텍스트마이닝과 연관관계 분석을 이용한다. 텍스트마이닝을 통해서 신문기사에서 가장 많이 등장한 키워드를 선별하고 연관관계 분석으로 추출된 키워드와 함께 연결의 맥락을 파악하고 분석한다. 다음으로 화재전후, 언론사별 기사에 대해 텍스트마이닝을 수행하여 기사에서 언급되는 키워드의 공통점과 차이점을 도출한다.

용어분석을 통한 키워드 용어의 파악은 문화재로서 송례문이 직면해 온 상황을 그대로 반영한다고 볼 수 있다. 이를 통해 정보생산자들이 좀더 다양한 양질의 정보를 생산하는데 도움이 되고 정보소비자들은 송례문에 대한 정보와 문화재 관련 지식을 함양하는데 도움이 이 되기를 기대한다.

II. 관련 연구

빅데이터는 데이터 형식이 다양하고 생성 속도가 매우 빨라 새로운 관리 및 분석 방법이 필요한 대용량의 데이터로[7] 모바일 기기의 확산과 소셜미디어의 활성화와 맞물리며 개인과 조직으로부터 축적되는 정보의

양이 폭발적으로 증가하게 되면서 나타나기 시작했다 [8]. 빅데이터는 데이터의 형태에 따라 정형 또는 비정형 데이터로 나눌 수 있으며 이러한 데이터의 분석 기술에는 텍스트마이닝, 평판분석, 소셜네트워크분석, 클러스터 분석 등의 4가지로 나누어 볼 수 있다[9].

본고에서 활용한 주요 기술은 텍스트마이닝 기법으로 방대한 텍스트 문치에서 의미있는 정보를 추출해내고, 다른 정보와의 연계성을 파악하는 등 단순한 정보 검색 그 이상의 결과를 얻어낼 수 있다[9]. 그러므로 텍스트마이닝 기법을 활용하려는 시도들이 활발하게 이루어지고 있는데 기업들은 자사 제품에 대해 텍스트마이닝을 활용하여 소비자의 니즈를 발견하여 신속한 대응을 하고 있고, 정치권에서는 선거 기간 동안 소셜미디어의 데이터를 분석하여 선거의 흐름과 판도를 예측하기도 한다[10].

이러한 텍스트마이닝 기법은 대량의 텍스트에 대한 분석 관련 연구로 이어져 특허정보, 학술지 논문정보, 기사, 소셜 미디어 등 다양한 분야에서 연구가 활발히 진행되고 있다. Lee et al.[11]의 연구에서는 특허문서로부터 기술동향 패턴을 파악하고 특정 제품과 기술에 관한 특허의 진화패턴을 발견하였다. 조수곤과 김성범 [12]은 산업공학분야 국제학술지 논문의 제목과 주제어에 대하여 텍스트마이닝을 수행하였다. 이 외에도 국제학술지 논문에 대하여 텍스트마이닝 기법을 이용한 연구는 계속 이루어지고 있다[13-15].

또한 기사에 대해 텍스트마이닝을 활용한 연구들도 다수 진행되었는데 감미아와 송민[16]은 텍스트마이닝 기법을 활용하여 경향신문, 한겨레, 동아일보 기사의 내용 및 논조에 차이가 있는지를 분석하였다. 이외에도 사설에 나타난 의료민영화 관련 이슈용어 분석[17], 신문기사와 사설, 정부간행물에 나타난 재난대응 용어 분석[18], 에볼라 주제에 대한 연구논문과 뉴스 텍스트에 대한 비교분석[19] 등이 있다.

최근 들어 소셜 미디어에 대해 텍스트마이닝을 활용한 연구들이 활발하게 이루어지고 있는데 이오준 외 [10]의 연구에서 트위터를 대상으로 영화 흥행 요인을 분석하였고 한진성과 윤지환[8]은 블로그와 트위터를 대상으로 부산영화제 개최 이전 한달과 이후 한달의 테

이터를 수집하여 분석하였는데 영화제 개최 이후 대중의 관심이 급격히 높아지는 것으로 나타났다. K항공의 램프리턴을 사건 전후의 소셜미디어 텍스트를 분석한 연구에서는 텍스트마이닝과 의미연결망분석 방법을 이용하였는데 이 사건을 계기로 항공사의 이미지가 부정적으로 변화하였음을 분석하였다[20].

이와 같이 다양한 문서와 주제에 대하여 텍스트마이닝을 통해 문서에 잠재되어 있는 핵심용어들을 파악한 연구들은 존재하지만 문화재로서 승례문에 대해서 신문기사가 전달하고자 하는 메시지를 파악한 연구는 찾기 힘들다. 그러므로 본 연구는 신문기사에 대한 내용 분석 방법이 질적인 고찰부문에 강점은 있지만 연구자의 해석에 따라 신뢰도에 영향을 줄 수 있는 약점이 있다[21]는 사실에 착안하여 문화재로서 승례문의 가치를 파악하는데 텍스트마이닝 기법을 연구방법으로 도입하고자 한다.

그러나 문화재로서 승례문에 대한 신문 기사를 텍스트마이닝하는 과정에서 통계 및 기술적 문제점들이 발견될 수 있다. 한국어 텍스트마이닝의 경우 조사, 부사 등 불용어의 존재, 중의어 및 다의어의 존재, 그리고 띄어쓰기와 붙여쓰기 등 한국어의 특수성으로 인해 빈도 분석에 혼란을 야기할 수 있다[22].

III. 자료의 수집

본 논문에서 신문기사의 수집은 신문기사 데이터베이스 시스템인 BIG KINDS[23]에서 2002년부터 2016년 7월까지 문화재로서의 승례문에 대해 보도된 일간지 기사를 대상으로 하였다. 이 중 승례문 기사에 “전통건축”, “전통문화”, “문화재 복원”, “문화재 복구”, “문화재 수리”, “문화재 보수”, “문화재 보존” 키워드가 포함되어 있는 기사를 수집하였다. 이 중 다른 키워드에 중복되는 기사를 제거하고 최종적으로 506건의 기사를 대상으로 하였으며 분석대상 기사들은 [표 1]에 요약하였다. [표 1]에서 나타난 것처럼 문화재로서 승례문에 관한 기사는 2008년 2월 10일 화재사고 이후 126건으로 가장 많이 보도되었으며 복구공사에 따른 여러 가지 이

슈로 2014년까지 기사수가 증가하다가 2015년 이후 화재사고 전과 같은 수준으로 기사가 보도됨을 볼 수 있다.

표 1. 분석대상 : 승례문 관련 기사

연도	언론사별 기사수	전체
2002	경향(1), 동아(2)	3
2003	경향(2), 동아(2), 문화(1), 한겨레(1)	6
2004	경향(1), 동아(1), 한겨레(2)	4
2005	경향(2), 국민(2), 동아(5), 문화(1), 서울(1), 세계(1), 한겨레(2)	14
2006	경향(4), 국민(1), 동아(1), 문화(2), 서울(3), 세계(6), 한겨레(3)	20
2007	경향(1), 동아(1), 문화(3), 서울(1), 세계(5)	11
2008	경향(22), 동아(23), 문화(19), 서울(21), 세계(24), 한겨레(17)	126
2009	경향(6), 국민(1), 동아(6), 문화(2), 서울(6), 세계(6), 한겨레(2)	29
2010	경향(3), 국민(5), 동아(7), 문화(9), 서울(3), 세계(6), 한겨레(4)	37
2011	경향(2), 국민(2), 동아(3), 문화(5), 서울(5), 세계(4), 한겨레(2)	23
2012	경향(2), 국민(5), 동아(5), 문화(8), 서울(9), 세계(2), 한겨레(5)	36
2013	경향(14), 국민(6), 동아(9), 문화(13), 서울(13), 세계(6), 한겨레(6)	67
2014	경향(11), 국민(15), 동아(12), 문화(11), 서울(21), 세계(14), 한겨레(12)	96
2015	경향(2), 동아(2), 문화(1), 서울(4), 세계(5), 한겨레(1), 한국(1)	16
2016	경향(2), 국민(2), 동아(2), 문화(2), 서울(3), 세계(2), 조선(1), 중앙(1), 한겨레(1), 한국(2)	18
총계		506

IV. 분석 방법

1. 분석 도구

분석에 사용한 프로그램은 오픈소스 프로그램인 R 프로그램 3.3.1 버전을 사용하였다. 우선 수집된 기사에 대해서 텍스트마이닝을 수행하기 위해 tm(text mining) 패키지를 사용하였으며 추출된 고빈도어들이 어떻게 군집을 이루는지를 확인하기 위해서 fpc(flexible procedures for clustering) 패키지를 사용하였다. 또한 추출된 고빈도어간 연관관계 분석을 위해 연관관계용 패키지인 arules(association rules) 패키지를 사용하였으며 이를 시각화하기 위해서 arulesViz(Visualizing association rules) 패키지를 사용하였다.

2. 분석 과정

첫째, 2002년부터 2016년 7월까지 문화재로서 승례문에 관한 일간지 기사 506건을 대상으로 기사 텍스트를 수집하였다.

둘째, 수집한 기사에 대하여 텍스트 처리를 수행하여 고빈도 키워드들을 1차로 추출한다. 1차 추출된 고빈도어로부터 추가적인 말뭉치(corpus) 수정작업을 수행한다. 예를 들어 한글의 경우 ‘문화재청이’, ‘문화재청을’ 등 조사를 포함하는 단어에서 조사를 제거하거나 일반어를 제거하는 등 말뭉치 수정작업을 수행하였다.

셋째, 최종적으로 수정된 말뭉치를 기반으로 2차 키워드 빈도분석을 수행하여 10회 이상 출현한 분석대상 핵심 용어를 추출하고 추출한 키워드의 전반적인 개요와 구조 등을 관찰하여 승례문 관련 기사의 내용을 파악하였다.

넷째, 한 기사에서 존재하는 키워드간의 연관규칙을 발견하는 연관관계 분석을 통해 연관관계가 높은 키워드들을 찾아내고 그 관계를 시각적으로 표시하였다.

다섯째, 승례문 관련 기사들은 시간적 경과 속에 기사 내용도 달라지는지 여부를 승례문 화재 전후 키워드들의 빈도분석을 통해 키워드들의 변화를 조망하였다.

여섯째, 언론사에 따라 승례문 관련 기사에서 사용하는 키워드간 공통점과 차이점이 있는지를 분석하기 위해 언론사별 키워드들의 빈도분석을 수행하여 키워드 특성을 관찰하였다.

본 연구의 자료 수집과 분석 과정을 아우르는 연구의 내용 및 방법을 요약하면 [그림 1]과 같다.



그림 1. 연구의 내용 및 방법

V. 분석 결과

1. 주요 키워드

본 연구의 분석대상인 506건의 기사에서 추출된 용어의 수는 32081개인데 최종적으로 얻어진 단어는 총 10회 이상 출현한 단어들 중 연구 분석에서 불필요한 무의미한 단어나 중의적 단어를 제외하고 101개의 단어를 선정하였다. [표 2]에 나타난 것처럼 송례문 기사에서 가장 많은 빈도로 랭크된 키워드는 ‘문화재청’으로 나타났다. 이는 송례문이 국보1호인 이유도 있지만 화재 이후 문화재청에서 복구공사를 진행하는 과정에서 관련된 이슈 및 그동안 서울시 중구가 맡아온 송례문 관리를 문화재청이 직접 맡으면서 관련 기사가 많아진 이유로 판단된다. 최상위에 랭크된 키워드는 ‘문화재청’, ‘송례문 복구’, ‘송례문 복원’, ‘전문가’, ‘문화재 복원’, ‘문화재 수리’, ‘문화재 보존’, ‘관계자’, ‘송례문 화재’, ‘문화재연구소’ 등 송례문 화재로 인한 복원, 복구 등의 키워드로 나타났다. 이는 15년 동안 문화재로서 송례문 관련 기사는 화재 이후에 집중되어 있으며 복원, 복구

등의 내용을 주로 보도하였음을 볼 수 있다.

또한 ‘신 대목장’, ‘대목장’ 등의 키워드가 고빈도어로 추출되었는데 ‘신 대목장’, ‘대목장’은 송례문 복구공사 도편수인 ‘신용수 대목장’을 지칭하는 것으로 2013년 말까지는 송례문 복구공사의 대목장으로서의 관련 기사가 많이 보도되었으며 2014년 초부터는 목재 관련 부실공사 책임을 묻는 기사가 많이 보도되었기 때문에 고빈도어로 추출된 것으로 판단된다. 또한 고빈도어에 ‘자격증 불법’, ‘송례문 단청’, ‘감사’, ‘대장간’, ‘입건’, ‘홍 단청장’, ‘자격증 대여’, ‘법률’, ‘기술자 자격시험’, ‘기술자 자격증불법’, ‘용산경찰서’, ‘지능범죄 수사대’ 등의 키워드가 추출되었다. 이는 송례문 복구공사 동안 목재를 포함한 기와, 단청, 못 등 송례문에 사용된 자재에서 원래 계획과 다른 흔적들이 발견되었고 이에 따른 부실공사의 의혹과 문화재 기술자 자격증 대여 의혹도 제기되었기 때문이다.

본 연구에서는 총 101개 키워드들이 506건의 기사에 분포되어 있는 키워드-문헌행렬(101×506)을 구성하였고 이 행렬을 대상으로 계층적 군집방법 중 Ward법을

표 2. 송례문 기사에서 나타난 최빈 키워드의 빈도 및 범위

키워드	빈도	키워드	빈도	키워드	빈도	키워드	빈도
문화재청	615	박 대통령	35	자격증 대여	19	발굴조사	13
송례문 복구	185	목재	33	건축문화	18	송례문 좌우	13
송례문 복원	164	전통건축	32	문화재 행정	18	종합점검	13
전문가	133	복구공사	31	문화재 훼손	17	건축학과	12
문화재 복원	128	세계문화유산	31	신한은행	17	나 청장	12
문화재 수리	125	자격증 불법	30	건축가	16	삼국시대	12
문화재 보존	115	지자체	30	정비사업	16	예술가	12
관계자	103	고궁박물관	28	광화문 헌판	15	화재	12
송례문 화재	101	미륵사지	27	대들보	15	기술자 자격시험	11
문화재연구소	75	소나무	27	대한 상공회의소	15	문화재 복구	11
박물관	72	송례문 단청	27	법률	15	송례문 공사	11
신 대목장	68	문화재 보호	26	사무관	15	송례문 기둥	11
조선시대	64	감사	24	아버지	15	송례문 방화	11
우리나라	63	대한 민국	24	전시관	15	아름다움	11
복원	62	대장간	23	콘크리트	15	자문위원	11
대통령	55	경복궁 복원	22	모니터링	14	기술자 자격증불법	10
문화재	53	입건	22	부석사	14	누각	10
광화문 복원	46	관람객	21	송례문 국보	14	문화재 화재	10
문화재위원회	46	문화재 전문가	21	송례문 주변	14	송례문광화문	10
프로그램	46	청와대	21	유홍준	14	안전진단	10
대목장	44	건설업체	20	한국전통문화	14	용산경찰서	10
가능성	42	홍 단청장	20	관리 실태	13	이탈리아	10
지정문화재	41	관련자	19	광화문 광장	13	지능범죄 수사대	10
송례문	38	박근혜 대통령	19	국립중앙박물관	13		
관광객	37	암각화 보존	19	대책 마련	13		
문화재 보수	35	유네스코	19	박 정부	13		

표 3. 계층적 군집분석을 사용한 키워드간 군집결과

키워드	군집명
문화재청	송례문 소관부서
송례문 복구, 문화재 수리, 문화재 보존, 전문가, 문화재 복원, 송례문 복원	문화재 수리
프로그램, 신 대목장, 대목장, 전통건축, 대통령, 박대통령, 관광객, 지정문화재, 광화문 복원, 분석사, 신한은행, 가능성, 문화재 보호, 대한민국, 복구공사, 지자체, 미르사지, 세계문화유산, 문화재위원회	송례문 복구공사
송례문 화재, 흥 단청, 건설업체, 입건, 법률, 용산경찰서, 송례문 단청, 문화재 보수, 자격증 불법, 유흥준, 발굴조사, 사무관, 누각, 문화재행정, 이탈리아, 나 청장, 송례문 공사, 자격증 대여, 기술자 자격시험, 기술자 자격증 불법, 지능범죄수사대, 문화재, 대들보, 송례문, 화재, 지문위원	부실공사 이슈
소나무, 목재, 건축가, 유네스코, 대장간, 고궁박물관, 관람객, 아버지, 경복궁 복원, 암각화 보존, 삼국시대, 청와대, 대책 마련, 관련자, 박근혜 대통령, 건축학과, 전시관, 문화재 화재, 문화재 복구, 모니터링, 안전진단, 관리 실태, 종합 점검, 대한상공회의소, 문화재 전문가, 한국전통문화, 송례문 국보, 감사, 박정부, 복원, 콘크리트, 광화문 현판, 송례문 좌우, 국립중앙박물관, 문화재 훼손, 송례문 방화, 정비사업, 예술가, 송례문 기둥	문화재 보존·관리·활용 이슈
송례문 주변, 건축문화, 광화문 광장, 송례문광화문, 아름다움	송례문 주변환경
문화재연구소, 관계자, 우리나라, 박물관, 조선시대	문화재

사용하여 키워드들 간의 거리를 측정함으로써 군집을 형성하였다. 각 군집에 포함되는 키워드들은 다른 군집에 있는 키워드들에 비해서 한 기사에 함께 나타날 가능성이 크다는 것을 의미하며 일반적으로 특정 주제에서 자주 언급되는 키워드들은 한 기사에서 같이 나타날 가능성이 크므로 각 군집에 포함되어 있는 키워드들을 바탕으로 그 군집이 나타내는 주제를 추정할 수 있다. [표 3]은 키워드간 군집분석 결과 및 각 군집의 키워드들의 특징을 가지고 주제어를 부여한 것이다.

키워드간 군집분석 결과에 따라 정확한 주제를 추론하는데 한계가 있으나 앞서도 설명했듯이 각 군집에 포함되는 키워드들은 다른 군집에 있는 키워드들에 비해서 한 기사에 함께 나타날 가능성이 크기 때문에 단어집합을 바탕으로 군집명을 추론하였다. 예를 들어 첫 번째 군집은 ‘문화재청’으로 한 개의 키워드만으로 하나의 독립된 군집이 형성되었는데 다른 키워드에 비해서 빈도수가 월등히 높기 때문인 것으로 판단되며 ‘송례문 소관부서’로 군집명을 명명하였다. 두 번째는 ‘송례문 복구’, ‘문화재 수리’, ‘문화재 보존’, ‘전문가’, ‘문화재 복원’, ‘송례문 복원’ 등의 키워드가 포함되어 있는데 이는 빈도분석 최상위권에 추출된 키워드로 많은 기사에 함께 출현하였기 때문인 것으로 판단되며 송례문을 포함한 문화재 전반적인 수리와 관련된 키워드들로 판단되어 ‘문화재 수리’라 하였다. 세 번째는 송례문 화재 후 대한민국 차원의 복구공사가 진행되면서 ‘신 대목장’, ‘대목장’, ‘대통령’, ‘박대통령’, ‘복구공사’ 등과 같은 키워드가 포함되어 ‘송례문 복구공사’로 군집명을 명명하였다. 네 번째는 복구공사 관련 여러 가지 부실 이슈들

에 대한 키워드들이 포함되어 있어 군집명을 ‘부실공사 이슈’로 하였다. 다섯 번째는 ‘유네스코’, ‘경복궁 복원’, ‘암각화 보존’, ‘안전진단’ 등 문화재에 대한 보존, 관리, 활용 관련 키워드가 모인 관계로 ‘문화재 보존·관리·활용 이슈’라고 하였다. 나머지 2개 군집은 ‘송례문 주변 환경’과 ‘문화재’로 명명하였다.

2. 키워드 연관관계

본 절에서는 앞 절에서 추출한 101개 키워드에 대한 연관관계를 분석하고자 한다. 이는 데이터 안에 존재하는 항목 간의 연관규칙을 발견하는 것으로 손님의 장바구니에 들어있는 아이템 간의 관계를 알아본다는 의미에서 장바구니 분석이라고도 한다[15]. 이때 손님의 장바구니는 하나의 트랜잭션이 되며 트랜잭션에 존재하는 아이템 리스트들의 발생하는 규칙을 수치화하는 것이다. 기사 자료의 경우 각 기사가 트랜잭션이 되며 기사에 포함된 특정 키워드가 아이템이 되어 특정 키워드 간 연관성을 찾아내면 되는데 연관성을 측정하는 지표로는 지지도(support), 신뢰도(confidence), 그리고 향상도(lift)가 주로 사용된다.

지지도는 전체 기사 중 두 개의 키워드가 동시에 존재하는 기사의 비이며, 신뢰도는 조건부확률로써 키워드 A를 포함하고 있는 기사 중에 키워드 B가 포함되어 있을 확률을 나타낸다. 지지도와 신뢰도는 1에 가까울수록 연관성이 높다고 할 수 있다. 향상도는 키워드 B를 연관규칙과 관계없이 관찰하는 것에 비하여 연관규칙을 알고 키워드 A를 발견한 경우에 대하여 키워드 B가 관찰되는 경우 얼마나 관찰이 증가하는가를 나타낸

다. 즉 키워드 A와 B가 관련이 없다면 향상도는 1이 되며 1보다 크다면 두 아이템이 서로 양의 상관관계를 1보다 작다면 두 아이템은 서로 음의 상관관계를 의미한다. 각각은 다음과 같이 표시할 수 있다.

$$\text{지지도}(A, B) = P(A \cap B) = \frac{A \text{와 } B \text{가 동시에 포함된 기사}}{\text{전체 기사}} \quad (1)$$

$$\text{신뢰도}(A \rightarrow B) = \frac{P(A \cap B)}{P(A)} = \frac{A \text{와 } B \text{가 동시에 포함된 기사}}{A \text{가 포함된 기사}} \quad (2)$$

$$\text{향상도}(A \rightarrow B) = \frac{P(A \cap B)}{P(A)P(B)} = \frac{P(B|A)}{P(B)} = \frac{\text{신뢰도}}{P(B)} \quad (3)$$

본 연구에서는 지지도를 0.01, 신뢰도를 0.1 이상인 규칙만을 추출하였는데 그 결과 총 1326개의 연관규칙이 도출되었으며 이중 원소가 1개씩인 연관규칙은 580개로 나타났다. 이 중 향상도 기준 상위 10개 규칙을 [표 4]에 요약하였다.

표 4. 키워드간 지지도, 신뢰도, 향상도 분석 결과

선행키워드	후행키워드	지지도	신뢰도	향상도
건설업체	용산경찰서	0.0124	0.6	29.1
용산경찰서	법률	0.0165	0.8	25.87
건설업체	법률	0.0144	0.7	22.63
용산경찰서	입건	0.0144	0.7	21.22
홍단청	법률	0.0103	0.625	20.21
입건	자격증대여	0.0144	0.4375	13.26
법률	입건	0.0124	0.4	12.13
박근혜대통령	박대통령	0.0227	0.5789	11.70
청와대	관련자	0.0103	0.3846	11.66
종합점검	지정문화재	0.0103	0.625	10.45

대부분 연관되어서 사용되는 단어들로 구성된 것을 알 수 있는데 예를 들어 ‘건설업체’와 ‘용산경찰서’가 동시에 발견될 확률인 지지도는 0.0124로 작게 나왔지만 선행키워드 ‘건설업체’가 포함된 기사에서 후행키워드 ‘용산경찰서’가 발견될 확률인 신뢰도는 0.60으로 매우 높았다. 또한 연관성을 의미하는 향상도의 경우 ‘건설업체’와 ‘용산경찰서’의 향상도는 29.1로 동시에 발견될 확률의 비율이 상당히 큼을 알 수 있다.

지지도 0.01 이상, 신뢰도 0.1 이상의 키워드간 연관

관계를 네트워크 그래프를 이용하여 나타낸 결과는 [그림 2]과 같다. [그림 2]에서 키워드와 키워드간의 원의 크기는 전체 기사 중에서 두 개의 키워드가 동시에 존재하는 기사의 비율인 지지도를 의미하며 비율이 높아질수록 원의 크기도 커진다. 또한 키워드와 키워드간의 원의 색깔은 두 키워드간 상관관계인 향상도를 나타내며 상관관계가 커질수록 색깔이 짙어진다. 이와 같은 규칙 하에 살펴보면 ‘문화재청’과 ‘승례문 복구’는 지지도가 가장 큰 것으로 나타났고 ‘문화재청’과 ‘승례문 복원’, ‘문화재청’과 ‘관계자’, ‘문화재청’과 ‘문화재 보존’, ‘문화재청’과 ‘전문가’ 등이 지지도가 큰 것으로 보인다. 상관관계 분석결과 [표 3]에서와 같이 ‘건설업체’와 ‘용산경찰서’, ‘용산경찰서’와 ‘법률’, ‘건설업체’와 ‘법률’, ‘용산경찰서’와 ‘입건’, ‘홍 단청’과 ‘법률’간에는 향상도 값이 큰 것으로 나타났다. 이는 승례문 복구공사 시 부실공사에 따른 기사들로부터 추출된 결과로 추정된다.

문화재로서 승례문 관련 기사에서 추출된 키워드들은 [그림 2]에서 보는 바와 같이 최상위 최빈 키워드들인 ‘문화재청’과 ‘승례문 복구’, ‘승례문 복원’ 등을 중심으로 거의 모든 키워드들이 고르게 연관되어 있는 것으로 나타났다. 그러나 ‘문화재청’과만 연결된 키워드들이 다수 존재했는데 ‘관리 실태’, ‘국립중앙박물관’, ‘나 청장’, ‘누각’, ‘대책 마련’, ‘모니터링’, ‘문화재 화재’, ‘부석사’, ‘사무관’, ‘승례문 공사’, ‘승례문 기둥’, ‘승례문 방화’, ‘승례문 좌우’, ‘안전진단’, ‘자문위원’ 이 그것이다. 또한 ‘미륵사지’와 ‘삼국시대’는 최상위 최빈 키워드와 직접 연결되지 않고 ‘미륵사지’는 ‘문화재연구소’와 ‘전문가’와 연결되어 있으며 ‘삼국시대’는 ‘조선시대’와만 연결되어 있음을 알 수 있다.

3. 키워드 트렌드 변화

본 절에서는 시간적 경과에 따라 기사 내용이 달라지는지 여부를 살펴보고자 크게 승례문 화재 전후 기사에서 최빈 키워드를 추출하고자 한다.

우선 승례문 관련 506건 전체 기사 중 승례문 화재 전 기사는 2002년 7월 16일부터 2008년 2월 9일까지 59건이고, 화재 후 기사는 2008년 2월 10일부터 2016년 7월 15일까지 447건이다. 각각 두 개의 파일을 txt파일로 변

표 5. 송례문 화재 전후 최빈키워드 비교결과

순위	송례문 화재 전	송례문 화재 후
1	문화재청*	문화재청
2	관광객	송례문 복구
3	조선시대*	송례문 복원
4	전문가*	문화재 수리
5	대한제국	전문가
6	문화재 복원*	문화재 복원
7	광화문 광장	문화재 보존
8	무지개 모양	관계자
9	문화재연구소*	송례문 화재
10	프로그램*	문화재연구소
11	박물관*	박물관
12	송례문 중앙	신 대목장
13	가능성*	우리나라
14	개방	대통령
15	미륵사지	조선시대
16	서울시 관계자	광화문 복원
17	송례문 연결	문화재위원회
18	우리나라*	대목장
19	일제	프로그램
20	지정문화재	가능성
21	청계천 복원	박 대통령
22	청계천 주변	지정문화재
23	경천사지	문화재 보수
24	관계자	전통건축
25	궁중문화	복구공사
26	문화재 보존관리	자격증 불법
27	수문장 교대	고궁박물관
28	송례문 역사	세계문화유산
29	유 청장	지자체
30	종로구청	송례문 단청

* 송례문 화재전후 모두 추출된 키워드

조립함으로써 우리가 알 수 있는 현재보다 이른 시기의 모습으로 돌려놓는 것을 의미하기 때문에[24] 송례문 화재와 관계없이 화재 전에도 기사화될 수 있는데 반해, ‘복구’는 천재지변 또는 고의적인 파괴 행위 등으로 손실되고 파괴된 경우 손실 이전의 상태로 회복함을 나타내므로 화재 이후 최빈키워드로 추출된 것으로 보인다. 그러나 화재 후에도 ‘복원’이 상위에 랭크된 이유는 송례문 복구 기본원칙[4]인 불탄 성문은 ‘복구’하고 성곽과 변형된 지반은 ‘복원’함에 대한 기사가 많이 언급되었기 때문인 것으로 판단된다. 또한 ‘수리’, ‘보수’ 등은 문화재뿐만 아니라 건물이나 시설 등에서도 일반적으로 사용하는 용어이므로 최빈키워드로 도출되었으며 ‘문화재 보존’은 문화유산이 지닌 문화적 중요성과 가치를 유지할 수 있도록 하는 모든 과정을 의미하기 때문에[24] ‘복구’, ‘복원’, ‘수리’, ‘보수’ 등의 키워드를 포함하는 상위 개념으로 화재 후 최빈키워드로 추출된 것으로 사료된다.

4. 언론사별 기사의 특징분석

이 절에서는 언론사별로 기사의 키워드 특징이 어떻게 달라지는가를 각 언론사마다 출현빈도가 높은 키워드를 그 언론사 기사들의 특징어로 간주하여 추출, 분

표 6. 언론사별 상위 20위 최빈 키워드 리스트

경향(75)	한겨레(58)	국민(39)	동아(81)	문화(77)	서울(90)	세계(81)
문화재청	문화재청	문화재청	문화재청	문화재청	문화재청	문화재청
문화재 수리	문화재 복원	문화재 수리	전문가	송례문 복구	송례문 복원	송례문 복원
전문가	전문가	송례문 복구	송례문 복원	문화재 보존	문화재 수리	전문가
송례문 복구	송례문 복원	문화재 복원	송례문 화재	송례문 화재	송례문 복구	문화재 복원
송례문 복원	지자체	송례문 복원	관계자	전문가	문화재 복원	문화재 보존
박물관	문화재 보존	대통령	문화재 보존	송례문 복원	관계자	송례문 복구
관계자	문화재 수리	신 대목장	문화재 복원	문화재 복원	박물관	송례문 화재
문화재 보존	관계자	대목장	문화재연구소	문화재연구소	우리나라	관계자
문화재 복원	신 대목장	문화재위원회	송례문 복구	관광객	문화재 보존	관광객
송례문 화재	대통령	감사	우리나라	조선시대	전문가	문화재 수리
신 대목장	박물관	자문위원	대한제국	건축문화	문화재위원회	대통령
조선시대	송례문 복구	전문가	조선시대	문화재 수리	광화문 복원	문화재연구소
문화재연구소	홍 단청	고궁박물관	네티나무	관계자	송례문 화재	프로그램
미륵사지	대목장	광화문 복원	대한상공회의소	복구공사	신 대목장	광화문 광장
지정문화재	문화재위원회	광화문 현판	신 대목장	지정문화재	가능성	대한민국
프로그램	우리나라	문화재 보존	박물관	가능성	건설업체	문화재 보호
대목장	가능성	입건	가능성	경복궁 복원	문화재 보수	세계문화유산
가능성	광화문 복원	송례문복구공사	고궁박물관	대통령	사무관	조선시대
복구공사	대장간	송례문광화문	문화재 수리	프로그램	프로그램	우리나라
예술가	문화재연구소	유 청장	유네스코	광화문 복원	박 대통령	박 교수

() : 보도건수

석하기로 한다. [표 1]에서 나타난 것처럼 전체 506건의 기사 중에서 조선일보와 중앙일보, 한국일보는 각각 1, 1, 3건 밖에 보도되지 않았기 때문에 7개 언론사에 대해서만 보도기사의 특징을 분석하고자 한다.

[표 6]는 7개 언론사에서 보도된 기사들의 상위 20개 키워드를 보여준다. 7개 언론사 모두 ‘문화재청’에 대해 가장 높은 빈도로 언급함을 알 수 있었고 상위 5개 키워드까지는 [표 2]의 최상위 키워드들이 순서만 바뀌면서 언급되어 언론사별 특징은 발견되지 않았다. 그러나 차상위 키워드에 대해 타 언론사 대비 특징적인 언론사들을 살펴보면 한겨레신문은 ‘지자체’, ‘신 대목장’, ‘대통령’, ‘홍 단청’ 등을 높은 빈도로 언급하였고 국민일보는 ‘대통령’, ‘감사’, ‘자문위원’, ‘입건’ 등 부실 복구공사 관련 기사를 자주 다루고 있다고 해석할 수 있다. 즉 한겨레신문과 국민일보는 송례문 복구공사 시 부실공사 논란에 관한 기사들이 많이 다루어졌음을 알 수 있다. 반면 동아일보는 ‘대한제국’, ‘조선시대’, ‘노티나무’, ‘유네스코’ 등 문화재에 대한 기사를 많이 다루고 있으며 문화일보는 ‘관광객’, ‘조선시대’, ‘건축문화’, ‘경복궁 복원’, ‘광화문 복원’ 등 조선시대 건축물 복원 및 그에 따른 관광객에 초점을 맞춘 기사를 주로 쓰고 있고, 서울신문은 ‘신 대목장’, ‘건설업체’, ‘사무관’ 등이 많이 언급된 것으로 보아 인용관련 기사가 많은 것으로 유추된다.

VI. 결론 및 시사점

본 연구에서는 문화재로서 송례문에 관련한 핵심 용어의 분석을 위해 2002년부터 2016년까지 15년간 보도된 신문기사 506건을 수집하고 해당 기사들에서 핵심 키워드를 추출하는 분석을 수행하였다. 우선 506건의 기사에서 가장 많은 출현빈도를 가지며 문화재로서 송례문과 관련한 101개의 핵심키워드를 도출하였으며 도출된 키워드를 분석하여 현재까지 송례문 관련 신문기사의 특징을 살펴보았다. 또한 도출된 101개의 핵심키워드간 관계를 연관관계분석을 통해 수치화 및 시각화하였는데, 연관관계 분석 결과 ‘문화재청’ 키워드를 가장 큰 허브로 ‘송례문 복구’와 ‘송례문 복원’을 다음 순

의 허브로 하는 키워드 그룹이 형성되는 구조를 보였다. 이는 지금까지 송례문 관련 기사는 다양한 주제를 다루지만 ‘문화재청’, ‘송례문 복구’ 및 ‘송례문 복원’을 중심으로 하여 일반적인 키워드들과 함께 출현하고 있음을 보여준다.

다음으로 2008년 2월 10일 송례문 화재를 계기로 송례문 화재 전과 화재 후의 키워드 동향을 분석하였는데, 화재 전의 주요 키워드는 송례문 관광에 관련된 키워드들이 추출되었으나 송례문 화재 후의 주요 키워드는 송례문 화재에 따른 복구공사에 대한 키워드가 상위에 랭크되었으며 복구공사와 관련되어 이슈가 되었던 사건들과 사건들에 대한 정부의 정책 관련 키워드들이 추출되었음을 유추할 수 있다. 이와 같이 송례문 화재 전후의 키워드 동향은 공통점보다는 차이점이 발견되었으며 화재 이후 송례문 복구를 포함한 문화재 보존, 관리방안에 대한 기사가 많이 다루어졌음을 보여준다.

마지막으로 언론사별 키워드를 분석하였다. 7개 언론사별 상위 키워드의 출현경향에는 차이가 발견되지 않았지만 언론사별로 많이 다루어진 이슈에서 차이가 발견되었다. 한겨레신문과 국민일보에서 자주 언급하는 특정 단어는 부실 복구공사에 관련한 용어로 해석되며 동아일보는 문화재 전반에 관한 기사를 문화일보는 전통건축의 복원에 관한 기사에서 언급될만한 특정 단어가 많이 발견되었다.

분석을 통해 나타난 결과를 요약하면 송례문과 관련된 기사는 송례문 고유의 문화재적 가치를 공유하려는 것 보다는 상대적으로 각종 사건 및 사고에 주목하고 있음을 알 수 있다. 물론 신문기사의 특성상 사회적으로 민감한 이슈에 집중하는 경향이 있지만, 송례문의 탁월한 문화재적 가치와 역사성 등을 지속적으로 대중과 공유하려는 노력이 필요하다고 판단된다.

본 연구는 지금까지 연구되어왔던 한글 신문기사의 직접 분석 방법에서 벗어나 텍스트마이닝 기법을 활용하여 다량의 데이터를 분석하였다는 데에 의의가 있다. 기존의 일반적인 내용분석 방법과는 달리 텍스트마이닝과 연관관계 분석기법을 사용하여 15년간 보도된 송례문 관련 기사들에서 언어의 통계량을 계산하고 이를 연관관계 분석기법과 연계함으로써 기사의 내용을 키

워드 형태로 파악하고 특징적인 메시지를 도출한 실험적인 연구이다. 본 연구 결과물은 승례문 관련해서 주요 이슈가 된 것들을 파악할 수 있고 언론사별로 승례문에 대해서 어떤 기사를 주로 다루는지 확인할 수 있다.

향후 지속적인 연구를 통해 승례문에 대한 키워드 용어를 파악한다면, 문화재에 대한 일반 대중의 인식과 국가 혹은 지자체 주도의 문화재 보호 시책간의 간극을 줄이는데 보다 도움이 될 것이라 사료된다. 한정된 예산과 인력 하에서 보존, 관리, 활용이라는 문화재 보호 정책은 때론 구체적인 시행에서 충돌을 일으키기도 한다. 화재 전후 키워드 분석에서 보이는 바와 같이 화재 전에는 활용에 초점을 맞춘 정책이 우세하다가, 화재 후에는 보존에 초점을 맞추는 정책으로 급변하는 양상을 보였다. 키워드 분석을 통한 정량적 접근법이 문화재 보호 정책을 수립, 시행하는데 효율적이고 균형잡힌 시각을 제공할 수 있을 것으로 생각해본다.

그러나 본 연구의 한계점으로는 승례문 관련 신문기사만을 대상으로 텍스트마이닝을 수행한 점이며 향후 승례문 관련 연구논문을 포함하여 신문기사와 연구논문간의 공통점과 차이점을 보여줌으로써 매체별 정보생산자들이 연구 및 현상 분석을 진행하는 것이 필요하다고 하겠다.

참 고 문 헌

- [1] 윤홍로, “승례문 화재의 전말과 복구대책,” 건축역사연구, 제17권, 제3호, pp.120-130, 2008.
- [2] 한범덕, “문화재 재난관리의 필요성과 발전방향 - 승례문 화재사고사례를 중심으로-,” 한국콘텐츠학회논문지, 제8권, 제10호, pp.189-197, 2008.
- [3] 김란기, “승례문 화재가 우리에게 남긴 것,” 건축역사연구, 제17권, 제1호, pp.156-157, 2008.
- [4] 최종덕, *승례문 세우기-승례문복구단장 5년의 현장 기록*, 돌베개, 2014.
- [5] <https://search.naver.com/>
- [6] 이동희, “신문보도를 통해 살펴본 전통목조건축 화재의 실태분석,” 대한건축학회연합논문집, 제16권, 제1호, pp.83-90, 2014.
- [7] 국가정보화전략위원회, *스마트 국가 구현을 위한 빅데이터 마스터플랜*, 서울 국가정보화전략위원회, 2012.
- [8] 한진성, 윤지환, “소셜 빅데이터 텍스트 마이닝을 활용한 부산국제영화제(BIFF) 활성화 방안 연구 -영화제 개최 전·후 비교 분석을 중심으로-,” 관광학연구, 제40권, 제1호, pp.133-145, 2016.
- [9] 김정숙, “빅 데이터 활용과 관련기술 고찰,” 한국콘텐츠학회지, 제10권, 제1호, pp.34-40, 2012.
- [10] 이오준, 박승보, 정다울, 유은순, “소셜 빅데이터를 이용한 영화 흥행 요인 분석,” 한국콘텐츠학회논문지, 제14호, 제10호, pp.527-538, 2014.
- [11] S. Lee, S. Lee, H. Seol, and Y. Park, “Using patent information for designing new product and technology: keyword based technology roadmapping,” R&D Management, Vol.38, No.2, pp.169-188, 2008.
- [12] 조수곤, 김성범, “텍스트마이닝을 활용한 산업공학 학술지의 논문 주제어간 연관관계 연구,” 대한산업공학회지, 제38권, 제1호, pp.67-73, 2012.
- [13] 배규용, 박주현, 김정선, 이영섭, “텍스트 마이닝 기법을 활용한 기후변화관련 식품분야 논문초록 분석,” 한국데이터정보과학회지, 제24권, 제6호, pp.1429-1437, 2013.
- [14] 임시영, 임용민, 이재용, “텍스트마이닝 기법을 이용한 U-City와 Smart City의 연구 동향에 대한 분석,” 한국지형공간정보학회지, 제22권, 제3호, pp.87-97, 2014.
- [15] 정용복, 박의섭, “텍스트 마이닝을 이용한 암반공학분야 SCI논문의 주제어 분석,” 터널과 지하공간, 제25권, 제4호, pp.303-319, 2015.
- [16] 감미아, 송민, “텍스트 마이닝을 활용한 신문사에 다른 내용 및 논조 차이점 분석,” 지능정보연구, 제18권, 제3호, pp.53-77, 2012.
- [17] 김유호, “의료민영화 논의에 따른 이슈용어의 연결 중심성 분석,” 한국콘텐츠학회논문지, 제12권, 제8호, pp.207-214, 2012.

[18] 강성경, 유환, 이영재, “텍스트 마이닝과 소셜 네트워크 분석을 이용한 재난대응 용어분석,” *Information Systems Review*, 제18권, 제1호, pp.141-155, 2016.

[19] 안주영, 안규빈, 송민, “텍스트 마이닝을 이용한 매체별 에블라 주제 분석 - 바이오 분야 연구논문과 뉴스 텍스트 데이터를 이용하여-,” *한국문헌정보학회지*, 제50권, 제2호, pp.289-307, 2016.

[20] 홍지숙, 오익근, “소셜미디어 빅데이터를 활용한 항공사 이미지 변화 분석-K항공의 램프리턴을 중심으로,” *관광연구저널*, 제30권, 제6호, pp.119-133, 2016.

[21] J. E. Ha-Brookshire and J. Hawley, “Trends of Research published by *Clothing and Textiles Research Journal* (1993 -2012) and *Outlook for Future Research*,” *Clothing and Textiles Research Journal*, Vol.32, No.4, pp.251-265, 2014.

[22] 구주나, 김경아, “한국어 텍스트 마이닝의 특성과 2011 한국경제총조사 자료에의 응용,” *응용통계연구*, 제27권, 제7호, pp.1207-1217, 2014.

[23] <http://www.bigkinds.or.kr/>

[24] <http://www.nrich.go.kr/>

김 철 주(Chul Joo Kim)

준회원



- 1994년 2월 : 서울대학교 경영학과(학사)
- 2015년 1월 ~ 현재 : (주)동방문 화유산 대표

<관심분야> : 고건축 시공, 건축역사

저 자 소 개

김 민 정(Min-Jeong Kim)

정회원



- 2012년 2월 : 서울대학교 산업공학과(박사)
- 2015년 3월 ~ 현재 : 숙명여자대학교 소비자경제학과 조교수

<관심분야> : User & Data Analysis