

토픽모델링 기반 행복과 불행 이슈 분석 및 행복 증진 방안 연구¹

A Topic Modeling Approach to the Analysis of Happiness and Unhappiness

양승준(Seung-Joon Yang) SLcoms 연구소

이보연(Bo-Yeon Lee) 연세대학교 정보대학원

김희웅(Hee-Woong Kim) 연세대학교 정보대학원²

ABSTRACT

Though Korea has received attention through an exceptional economic growth and the big K-POP fever all over the world, its happiness level is not so high. Therefore, this research aims to find not only the Korean's condition of the happiness and unhappiness, but also the way to enhance their happiness. We collected various web data(89,127 cases from 2013/01 to 2014/12) through searching our own 26 keywords based on Alderfer's ERG Theory. Also, we tried to analyze the subjects related to happiness and unhappiness by using LDA topic modeling. As the result, the condition of happiness and unhappiness were the top topics extracted from each field. We conducted the second detailed analysis based on the data of condition of the happiness and unhappiness which are the top topics of the previous analysis. From the second analysis result, we proposed several ways to enhance happiness from the perspective of government, corporate, family, education, social welfare. This paper is meaningful because it catches the condition of happiness and unhappiness based on a real web data as well as transform the data into the knowledge. Also, this paper provides the practical methods from the view from all walks of life that may enhance happiness and relieve unhappiness.

Keywords: Happiness; Unhappiness; Topic Modeling

¹ 논문접수일: 2015년 12월 20일; 1차 수정: 2016년 3월 11일; 2차 수정: 2016년 4월 11일; 게재확정일: 2016년 4월 11일

이 논문은 2015년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2015S1A3A2046711). 이 논문은 2015학년도 연세대학교 미래선도연구사업 (부분적인) 지원에 의하여 작성된 것임 (2015-22-0053).

² 교신저자

I. 서론

우리나라는 반 세기 만의 눈부신 경제 성장으로 1인당 명목 GDP 세계 28위, 아시아 기준 8위로 상대적으로 높은 경쟁력의 경제 강국이 되었다(2015년 10월 기준). 반 세기 전 6.25 전쟁 발발로 500여 만 명의 엄청난 난민이 발생, 자력 회복 불가 판정을 받아 국제 사회의 원조를 받는 등 비관적 상황에 처했지만 한국인 특유의 뚝심과 놀라운 집중력으로 1960~80년대 '경제개발 5개년 계획', '새마을 운동', '중화학공업 정책', '88 서울 올림픽'등을 성공적으로 수행해 '한강의 기적'을 이루어냈다. 한 때의 원조 수원국이 원조 공여국으로 비약적 지위 격상을 한 것이다. '1960년 이후 대한민국의 경제 성장은 어떠한 경제 이론으로도 설명하기 어려운 기적과 같다(노벨 경제학상 수상자, 로버트 루카스)', '한국은 놀랍게도 빠르고 성공적인 기술 변화를 이뤄냈다. 정말로 놀라운 일이다(미래학자, 앨빈 토플러)' 등 저명한 연구자들이 남긴 말에서 알 수 있듯, 한국인들은 그들만의 독특한 국민성을 발휘해 세계인들을 놀라게 하는 급속한 발전을 했다.

비단 경제 영역뿐만 아니라, 우리만의 독특한 감성으로 창조한 'K-POP', 'K-DRAMA' 등이 큰 인기를 구가하며 세계적 한류 열풍을 주도 중이다. 우리만의 주체적, 기발한 방식으로 새로이 빚어낸 '문화 콘텐츠'가 큰 경쟁력이 되어 세계로부터 인정받고 있는 것이다. 하지만 이러한 발전 양상과는 대비적으로 우리나라는 타국보다 척박한 환경에서 고된 삶을 대물림하고 있다. 아동 학업 스트레스 지수 세계 1위(2013년, OECD 조사), GDP 대비 복지 지출 비율 OECD 최하위, 여성 사회참여 OECD 꼴찌, 자살률 11년 연속 OECD 1위, 의료비 증가율 OECD 최고, 사기사건 하루 발생 600건의 일명 '사기 공화국', 한국 국회의원 경쟁력 OECD 최하위

수준, 국민의료비 공공부담 OECD 최하위, 은퇴 후 생활 자신감 최하위 등 수한 통계자료가 이를 증명한다.

뿐만 아니라, 2015년 4월 UN 세계 행복 보고서³에서 우리나라는 158개국 중 47위로 명시됐고, 2015년 6월 '갤럽·헬스웨이 2014 글로벌 웰빙 보고서⁴(Gallup·Health ways Well-Being)'에서는 145개국 중 117위를 기록했다. 또한 약 10년 째 한국 어린이, 청소년 행복지수 OECD 최하위권, OECD 행복지수(Better life Initiative, BLI)는 36개국 중 27위라는 지표로써 행복 기준에 있어서는 다른 나라보다 명백히 낮아지고 있다는 것을 알 수 있다.

상기 언급된 사례들로부터 우리나라가 전세계인들이 쉽사리 무시할 수 없는 저력과 독특한 국민성으로 큰 발전 동력을 가지고 있지만, 그에 뒤처지지 않는 오명 또한 가지고 있다는 점은 매우 흥미롭고 역설적이다. 이러한 맥락에서, 한국인들이 거처온 사회문화적 특성을 바탕으로 그들이 품고 있는 행복요인 및 불행요인을 발굴하여, 행복 증진 및 불행 경감 방안을 확립하는 것은 건강한 사회 확립을 위해 매우 필요한 일이다. 한국인들을 행복 또는 불행으로 이끄는 근원적 조건을 탐구하고, 사회 각계 측면에서 문제 현안에 대해 이를 바라본다면, 더욱 바람직한 사회를 만들어 가는데 큰 기여를 할 것이기 때문이다.

따라서 본 연구는 한국인의 행복과 불행요인을 도출하고, 이에 대한 행복 증진 및 불행 경감 방안을 제시하는 것에 목표를 둔다. 더욱이 본 연구에서는 현재 다양하게 활용되고 있는 빅 데이터 분석법 중 하나인 토픽 모델링을 활용해 연구를 진행하고자 한다. 정보의 원천인 온라인 상 데이터를 수집 및 분석해 그로부터 지식을 만들어내고, 이를 기반으로 정부 혹은 산업계 등지에서 당면한 현안을 풀어내갈 수 있는 해법을 합리적으로 모색하고자 하는 것이다. '빅 데이터'라는 새로운 기술 분석법을 통해 창출된 과학적 지식인 만큼, 현 시대 한국 사회를 살아가는 우리 국민들의 삶을 윤택하

³ UN 세계 행복 보고서, <http://worldhappiness.report/>

⁴ <http://www.well-beingindex.com/download>

게 하는데 큰 도움을 줄 것이라 기대한다.

II. 개념적 배경

2.1. 행복의 개념

행복(幸福, Happiness)은 첫째, 복된 좋은 운수, 둘째, 생활에서 충분한 만족과 기쁨을 느끼는 흐뭇한 상태를 의미한다. 이에는 ‘만족감’, ‘기쁨’, ‘즐거움’, ‘웃음’, ‘보람’, ‘가치감’, ‘평온감’ 등 일정 기준 이상의 좋음의 느낌을 나타낸다. 영미권에서 행복을 뜻하는 ‘Happiness’가 행운을 의미하는 고대 스칸디나비아어 ‘Happ’에서 기원했다는 사실 또한, 행복이란 최상의 결과로부터 일어나는 심리적 상태라는 사실을 뒷받침 한다.

뿐만 아니라 ‘행복’은 철학, 심리학, 경제학 등 다양한 학계에서 오랫동안 연구되어와, 그에 걸 맞는 다양한 정의가 존재한다. 첫째, 철학적 관점에서, 아리스토텔레스는 행복을 만물이 지향하는 최고선(最高善)이자 덕(德)의 실천에 따른 자기실현 상태라 했으며, 이를 쾌락과 동일시하지는 않았다. 그러나 헬레니즘 철학(Hellenistic philosophy)의 한 갈래인 에피쿠로스학과(Epicurean school)는 행복이란 무엇이며, 어떻게 하면 행복한 삶을 살 수 있을까를 고민하며 ‘쾌락’을 중시하였다.

둘째, 심리학계에서는 최근 ‘긍정 심리학(Positive Psychology)’연구가 이슈화 되고 있다(권석만 2010). 인간의 긍정적인 심리적 측면을 과학적으로 연구하고 인간의 행복과 성장을 지원하는 학문분야로서, 과거 심리학이 인간의 불행과 장애에만 편향적으로 관심을 기울여왔다는 반성적 성찰에서 출발했다. Reis et al.(2003)은 긍정 심리학을 인간이 나타낼 수 있는 최선의 기능 상태에 대한 과학적 연구라 정의한 바 있다. 긍정 심리학은 개인, 집단 그리고 사회가 번창하도록 하는 요인들을 발견 및 촉진하는 것을 목표로 하고 있다.

셋째, 경제적 관점에서 ‘행복’은 주로 GDP나 GNP 등의 각종 객관적 경제 지표를 가지고 비교·측정되어 왔다. Bruno S. Frey et al(2001)에 따르면, GDP 15,000 달러 이상 국가의 경우, 국가의 평균 소득과 그 국가의 평균 행복감과는 큰 차이를 보이지 않는다고 했다. 이것은 경제적 선진국에 거주하는 중산층 또는 상류층 시민들에게는 더 많은 경제적 충족이 그들의 행복감을 크게 증진시켜주지 않는다는 것을 의미한다. 따라서 끝없는 물질주의적 태도 탐닉에서 벗어나 다른 방향으로 행복을 탐구해야 함을 시사하고 있다.

2.2. 행복 영향 요인 선행 연구

행복에 영향을 미치는 요인에 관한 연구는 크게 객관적 행복감과 주관적 행복감에 관한 연구로 나뉜다. 객관적 행복감은 GNP, GDP, 기대수명, 영아 사망률, 문맹률 등으로 측정되어왔다. 우리나라는 2013년 UN개발계획 HDI에서 187개국 기준 15위를 기록했고, 2014년 EIU에서는 10점 만점 8.11점으로 167개국 기준 20위이다. 하지만 객관적 지표가 표현하는 행복에는 인간이 느끼는 행복감을 정확히 측정하기에 한계가 있다. 즉, 객관적 행복 지표는 인간이 지각하는 행복의 본질적 요소를 내포하고 있지 않다(Campbell 1976). 또한 객관적인 삶의 조건과 지표들의 목적은 삶의 객관적 조건 개선으로써 ‘전체적 시민’의 행복을 추구하고자 하는 것이지 ‘각 개인’의 행복을 추구하고 위해 생성된 지표가 아니다. 따라서 객관적 삶의 질 향상이 반드시 모든 개인의 행복을 약속하는 것은 아니며, 개인의 행복 정도는 각 개인 수준에서의 경험적 접근을 통해서만 증명될 수 있다는 한계점이 존재하여,(김명소, 한영석 2006) 주관적 행복감에 대한 연구가 등장하게 되었다.

주관적 행복감(Subjective Well-being)은 주관적인 삶의 평가, 인지와 같은 인지적 측면과 정서적 측면의 만족감, 또는 높은 삶의 만족도, 적은 부정 정서 경험, 그리고 많은 정적 정서 경험을 하는 상태, 생활 만

족의 가장 주된 요소로서 생활 전반에 대한 만족 정도로 정의된다. 특히 Diner는 본 분야의 효시가 되는 연구를 행해왔는데, 그는 주관적 삶의 질을 결정짓는 요소를 유쾌감(Pleasant affection), 불쾌감(Unpleasant affection), 만족도 평가차원(Life satisfaction), 평가 영역(Domain Satisfaction)의 4가지로 구분했다(권석만

2010). 이러한 요소들과 관련 있는 변수들-이르테면, 유전, 성격, 목표, 적응력 등-이 주관적 행복감에 어떤 영향을 미치는지 많은 방면에서 연구되고 있다. 또한 국내·외적으로 행복감 측정을 위한 수많은 행복 지수들이 개발되고 있다. <표 1>와 같이 ‘행복 지수’에 관한 선행 연구들을 정리하였다.

<표 1> 행복 지수 관련 선행 연구

연구	연구 목적	연구 결과
미시건대 행복지수 측정 (2001)	각국의 행복지수 및 행복을 위한 방법 제시	- 전 세계의 ‘행복도’와 ‘생활 만족도’에 대한 설문조사 수행으로 ‘주관적 웰빙 순위(Subjective Well-being Ranking)’ 산출 - 1999~2001년도 연구에서 대한민국은 49위 차지 - 행복으로 가는 길 10가지 제시(유전적 성향, 결혼, 우정, 선행 등)
김명소 외 (2003)	한국인의 행복에 대한 인구통계학적 분석	- 성별, 월평균소득, 학력에 따라 행복 정도에 유의성 있음 - 행복한 삶의 세부 항목에 대해서는 개인의 인구 통계학적 변인의 차이에 따라 각기 다른 행복감 가짐 - 한국인 전체적으로는, 성취 및 자기 수용감, 경제력, 배우자(이성) 과의 사랑 및 신뢰, 여가등이개인 삶의 만족도에 영향 미침
추홍규 (2005)	개인의 행복지수 수준과 결정 요인 분석	- 개인별 행복지수 및 자가 평가 건강수준의 상관관계 파악 - 행복지수는 전반적 건강상태와 양(+)의 상관관계 가짐 - 또한 사회적 고립감, 정서적 긴장 등과 음(-)의 상관관계 가짐
영국 BBC 다큐멘터리 행복 (2006)	종합 행복지수 측정 및 산출	- 심리학, 의학, 사회학 분야의 기존 연구를 토대로 행복의 실천적 방법을 분석해, 행복 실천방법 및 행복 종합 지수 측정
김명소 외. (2006)	한국인의 행복수준 공식 개발	- Alderfer의 ERG Theory에 따른 한국인의 행복지수 산출 - 한국인의 행복지수=2.5*생존+2.5*관계+5*성장(평균=57.71) - 여성 행복지수 > 남성 행복지수 - 성장 요소가 한국인의 행복감 지각에 상대적으로 큰 영향을 미침
신경제 재단 (New Economics Foundation) (2006)	행복 지구 지수 발표 (HPI, Happy Planet Index)	- 생활 만족감, 기대 수명, 생태 발자국 지수 등 반영 - 하지만 정치적 자유, 인권, 노동 등 요소 완전 무시 - 생태 발자국 지수가 정교하지 않은 측면이 있음 - 생활 만족감의 기준이 개인 별로 매우 주관적임
유엔 개발계획 (UNDP)	인간개발지수 (HDI, Human Development Index)	- 기대수명, 교육수준, GNI의 3요소로 측정 - 우리나라는 2013년 187개국 중 15위 차지
동아 행복지수 (2015) ⁵	동행지수	- 심리적 안정, 가정생활 만족도, 경제적 만족도, 업무 만족도 등 심층 설문 수행 - 최근 1년간의 주요 뉴스 중 행복에 영향 미친 이슈를 빅데이터 분석 - SNS 분석으로 행복 관련 트렌드 분석 - 연령대, 지역, 성별 등 다양한 관점에서 행복 지수 분석 비교

⁵ 동아행복지수, <http://news.donga.com/3/all/20151207/75216781/1>

III. 이론적 배경

3.1. ERG THEORY

본 연구에서는 김명소와 한영석(2006)의 연구에서 ‘한국인의 행복의 구성 요소’라고 판단한 16가지 기준(경제력, 외모 자녀의 성장, 여가, 등)에 의거한 하위 키워드를 선정하였다(아래 <표 2> 참조). 위 16가지 기준은 독일의 심리학자 클레이턴 폴 엘더퍼(Clayton Paul Alderfer)의 생존-관계-성장 욕구 위계 이론(ERG Theory)에 근거해 선정되었다. ERG 이론이란, 생존-관계-성장(Existence-Relatedness-Growth)의 약어로서, 아브라함 매슬로우(Abraham Harold Maslow)의 인간 욕구 단계설(Maslow’s hierarchy of needs)을 확장시킨 이론이다. 생리적, 안전과 같은 인간의 기본적인 하위 욕구를 생존의 영역에 통합하고, 사회적, 개인적 대인 관계 형성 시 충족되는 안전 욕구, 사회 욕구, 존경 욕구의 일부를 관계성으로 분류했으며, 마지막 성장의 영역에는 자기실현과 자기 존중의 욕구 등 개인의 노력과 성취에 의해 충족될 수 있는 욕구를 포함시켰다.

인간 욕구 단계설에서는 하위 단계의 욕구 충족 후 상위 단계의 욕구 충족이 가능하다 주장하나, 본 이론에서는 각 단계의 욕구 충족은 동시에 일어날 수 있으

며, 각 단계의 욕구 별 중요도는 개인차 또는 환경, 문화적 차이가 존재할 수 있다 하였다. 또한 ‘좌절’, ‘퇴행’의 개념을 새롭게 도입했다는 점에서 인간 욕구 단계설과 차별성을 지닌다. 상위 욕구가 충족되지 않으면 더 쉬운 하위 욕구로 좌절 또는 퇴행할 수 있다고 간주하는 것이다.

본 이론은 각 욕구 별 개인 차이를 인정하며, 타 욕구 이론보다 탄력적이며 유연한 구조를 지니므로써 조직 구성원의 욕구와 만족에 대한 높은 설명력을 이끌어 낼 수 있으며, 조직 행동론 등 많은 분야에서 실제 적용해 많은 인정을 받고 있다. 그러나 복잡한 인간의 동기 유발 과정을 지나치게 단순화하고 있으며, 구성원의 욕구 충족을 조직 목표 달성 또는 조직 발전의 충분조건으로 자칫 오인하여 잘못 적용할 가능성도 지적되고 있다.

본 연구에서는 ERG 이론을 근거로 도출된 16가지 ‘한국인의 행복의 구성 요소’들을 기초로 그와 상등한 의미를 포함하고 있다고 생각되는 키워드를 선정해 행복, 불행과 함께 데이터를 검색 및 수집했다. 즉, 위 16가지 요소의 의미를 내포하고 있는 총 26가지 키워드(경제, 돈, 교육, 직업, 만족 등)를 도출했다 (<표 2>).

<표 2> 연구 분석 시 선정된 키워드

욕 구	한국인의 행복요소	선정된 키워드
생존 (Existence Needs)	경제력, 건강, 외모, 사회, 정치, 문화, 환경	경제, 돈, 물가, 소득, 부동산, 지출, 자동차, 건강, 국가, 복지, 분배, 수명, 실업
관계 (Relatedness Needs)	자녀의 성장, 부모 및 친지 관계, 이성과의 사랑, 타인과의 관계, 사회적 지위 및 안정	교육, 직업, 취업, 친구, 학교
성장 (Growth Needs)	자기 수용감, 긍정적 인생관, 자기계발 및 목표, 자립성, 종교, 사회봉사, 여가	만족, 안정, 저축, 수준
중립적 단어		요인, 조건, 본질, 삶

IV. 연구방법론

4.1. 토픽모델링

토픽 모델링(Topic Modeling)이란 문서의 주제를 도출하기 위해 텍스트 내 단어를 분석하는 방법론이다 (Steyvers and Griffiths 2007; Blei 2012; 강범일 외. 2013). 최근 기하급수적으로 증가하고 있는 정형 및 비정형 데이터의 양으로 인해 이를 분석하는 빅 데이터가 향후 기업 또는 국가의 성패를 좌우하는 새로운 가치의 원천으로 크게 강조되며(김정선 외. 2014) 다양한 활용법이 모색되고 있는 가운데, 텍스트 빅 데이터를 분석해 인사이트를 얻고자 하는 토픽 모델링 분석법 또한 큰 주목을 받고 있다. 이로 이러한 방법론 적 특성으로 인하여, 토픽 모델링은 각종 문헌들을 연구 대상으로 삼는 다양한 분야에서 적절한 분석 도구로서 활용되어 왔다. 특히 토픽 모델링 분야에서 Blei et al.(2003)가 제안한 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 알고리즘은 각 주제 별 단어 수 분포를 기반으로, 주어진 문서에서의 기 단어 수를 분석하여, 해당 문서가 어떤 주제들을 함께 다루고 있을지를 예측하는 알고리즘으로서, 높은 성능을 보이며 학계에서 표준으로 인식되고 있다.

LDA는 알고리즘 자체의 단순성, 데이터 차원 축소의 유용성, 의미적으로 일관성 있는 주제 생산 등 여러 장점을 가진다(Mimno and McCallum, 2008). 따라서, 소셜 미디어, 저널, 신문 기사 등 다양한 데이터 자원들을 기반으로 많은 학계에서 토픽 모델링 분석법을 이용한 연구를 진행하고 있다. 먼저, 소셜 미디어는 사회적 이슈와 공통된 콘텐츠를 공유하는 무수한 사용자들의 연결고리로 얽혀 있으며, 실시간으로 생성되는 엄청난 데이터를 통해 다양한 함의를 이끌어 낼 수 있다는 특성으로 인해, 본 분야에서 가장 활발한 연구의 장이 되고 있다. Zhao et al.(2011)은 대표적 소셜 네트워크 서비스(SNS) 중 하나인 트위터(Twitter) 상에 출현하는

토픽 분포와 함께, 동 기간 내 New York Times 기사의 주제 분포를 비교 분석하여 트위터의 뉴스 매체적 특성을 확인한 바 있다. 최돈정 외(2011)는 마이크로 블로그를 통한 단어의 동시 발생빈도를 활용해 토픽 연관 단어를 추출, 단어 간 연관 정도를 파악하였다. 배정환 외(2013)는 2012년 대선과 관련한 트위터에서 추출된 토픽과 그 변화 양상이 실제 사건 또는 신문 기사들과 어떠한 연관성을 가지고 있는지 분석하였다. 차윤정 외(2015)는 트위터 비정형 데이터를 중심으로 한 소셜 미디어 상 애플(Apple), 삼성(Samsung), 샤오미(Xiaomi) 스마트폰 기종에 대한 고객 리뷰를 수집하여 토픽모델링을 수행하고, 추출된 브랜드 별 토픽을 비교 및 분석해 각 사에 알맞은 최적의 마케팅 전략을 제안한 바 있다.

둘째, 생물 정보학, 문헌 정보학, 정보 시스템학 등 다양한 학계에서 국내 및 국외의 학술자료 트렌드를 파악해, 이로부터 연구자들에게 가이드를 제공하고자 하는 연구 또한 진행되고 있다. Gerrish and Blei.(2010)는 시간의 흐름에 따른 논문 코퍼스(Corpus)로부터의 주제 변화를 파악하여, 이로부터 개별 문헌의 영향력 및 파급력을 분석하고자 했다. Song et al(2013)은 생물 정보학 분야의 논문으로부터 화두가 되고 있는 주제들을 추출하여 해당 학문 분야의 지적 구조를 분석하고 하위 분야에 대한 연구 경향을 살펴보았다. 박자현과 송민(2013)은 국내 문헌 정보학 분야에서의 연구 동향을 분석해 그로부터 의의를 도출하고자 하였다. 안정국 외(2015)는 정보 시스템 학계에서의 연구 트렌드 변화를 토픽모델링과 네트워크 분석을 통해 연구하여, 현재 및 향후의 정보시스템 학계의 이슈를 도출 및 예측했다.

셋째, '신문' 또한 보도매체로서 시사적이며 시의적 주제를 가지고 대중에게 소식을 타전하는 특징을 가지므로 토픽 모델링 분야에서 여러 차례 분석이 진행되어 왔다. Newman and Block(2006)은 초기 미국 사회와

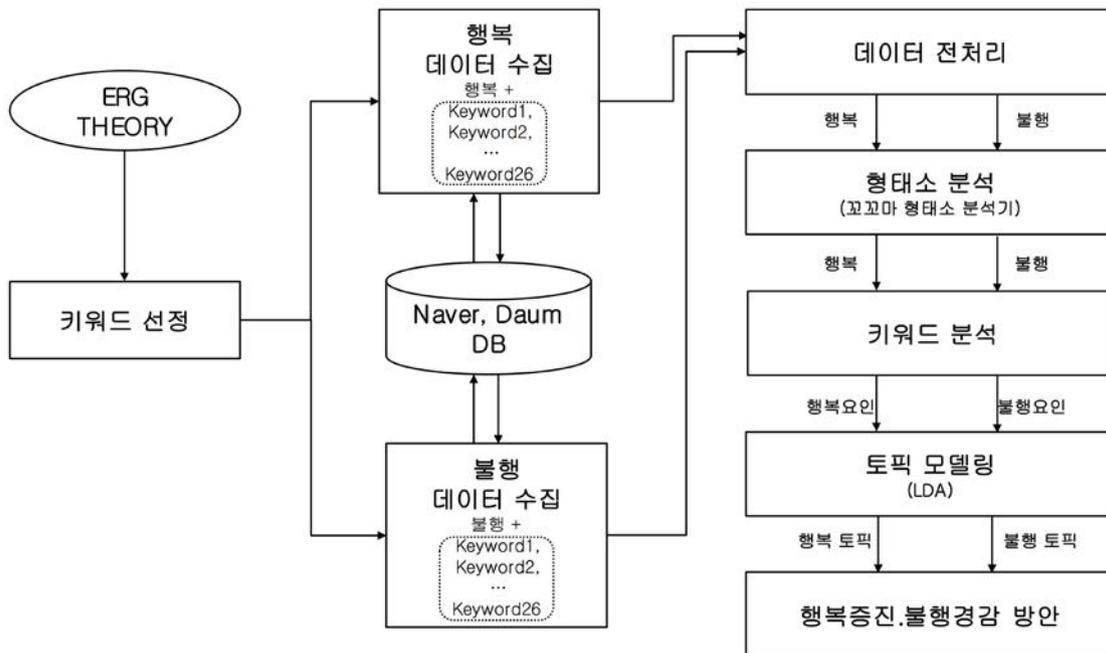
출판 문화의 이해를 위해, 토픽 모델링을 이용하여 18세기 신문 텍스트에서 주제를 추출하고, 각각의 주제들이 시간의 흐름에 따라 어떻게 변하는지를 분석하였다. Grimmer(2010)는 미 상원의원들의 보도 자료에서 의원들이 강조하는 어젠다(Agenda)를 추출하는데 토픽 모델링을 활용해, 입법자들의 홍보방식을 분석하였다. Gam and Song(2012)은 신문기사의 키워드 빈도 분석과 클러스터링, 분류 결과를 통해 신문사 간 논조 차이점을 분석하고자 하였다.

위와 같은 기존 연구 검토에 따르면, 토픽 모델링은 주로 특정 분야의 데이터를 가지고 이에 대한 동향 분석을 하거나 이를 탐색해 사실 관계를 파악하는 정도

에 머물러 왔다. 그러나 신문 기사와 블로그, 웹 콘텐츠 등 다양한 성격의 데이터를 토픽모델링으로 복합적으로 분석해, 그로부터 도출된 결과에 대한 실질적 해결책을 제안한 연구는 아직까지 이루어지지 않았다. 이에 본 연구는 행복·불행에 대해, 각종 웹 콘텐츠를 통합적으로 분석해 이에 영향을 미치는 행복 요인과 불행 요인들을 유추하고 실질적 해결 방안을 제안하고자 한다.

4.2. 데이터 수집 및 분석 방법

본 장에서는 기 연구의 데이터 수집, 정제, 분석 절차에 대해 설명하고자 한다. 개괄적인 데이터 분석 절차는 <그림 1>과 같이 진행하였다.



<그림 1> 연구 절차

4.2.1. 데이터 수집

본 연구에서는 데이터 수집을 위해 국내 약 90% 이상 시장 점유율을 차지 중인 웹 포털 사이트 N사 (<http://www.naver.com/>)와 D사(<http://www.daum.net/>)로부터 2013년 1월부터 2014년 12월까지 포스팅 (Posting)된 뉴스, 블로그, 통합 웹 문서를 대상으로 웹

크롤링(Web Crawling)을 수행했다.

데이터 수집 시, 김명소 외(2006)가 선정한 한국인이 생각하는 행복 구성 요소 16가지 기준(경제력, 건강, 외모, 여가 등)에 의거한 하위 키워드 26개(위 <표 2> 참조)를 선정하여 데이터 수집 기준을 명확히 하고자 했다. 위 26개의 키워드와 동시에 행복·불행을 각각 검색

해 출현하는 데이터를 얻었다. 예컨대, ‘행복, 경제’, ‘불행, 경제’ ‘행복, 돈’, ‘불행, 돈’ 등 행복과 불행을 각 키워드와 함께 검색함으로써 도출되는 데이터를 크롤링한 것이다. 그 결과 총 89,127건(N사 뉴스 기사: 41,077건, D사 블로그 및 웹 문서: 48,050건)의 데이터를 추출했으며, 이를 기초로 연구를 진행하고자 했다.

4.2.2. 데이터 전처리

웹 크롤링(Web Crawling)과정을 통해 수집된 데이터를 아래 과정을 따라 전처리 했다. 먼저 최초로 수집된 데이터는 웹 페이지 형태이므로, Html 또는 Xml 마크업 기호, 각종 콘텐츠의 불필요 정보(블로거 명, 주소 정보, 전화번호 등)를 가져 이를 정제 처리하였다. 다음으로 본 연구 분석 대상이 되는 데이터는 대부분 한글이라는 특성을 가지므로, 정제 후 문장을 유의미한 최소의 단위로 쪼개는 형태소 분석(Morphological analysis) 과정을 시행하였다. ‘형태소 분석’이란 문장을 문법적, 관계적 의미를 내포하는 각각의 단어로 나누어 이에 대한 형태소 속성을 추출하고 단어 원형을 복원하는 과정을 의미한다. 본 연구에서는 현재 시중에 나와있는 여러 한글 자연어 처리기 중 가장 고성능이며, 오류율이 낮아 자연어 처리 분석 시 가장 빈번히 사용되고 있는 ‘꼬꼬마 형태소 분석기⁶⁾’를 활용해 수집한 말뭉치 저장하고, 이를 형태소 단위로 구분해 각종 통계 데이터 생성하는 등 연구를 위한 기반을 마련하였다. 연구 목적 상 명사 형태소가 가장 핵심적인 의미를 지니고 있을 것으로 판단해 명사 중심으로 추출했으며, 그 외 속성을 가지는 형태소들은 제거 후 아래 단계의 분석 과정을 수행하고자 했다. 더불어 명사이지만 토픽 분석에 큰 의미를 가지지 않을 것으로 간주되는 단어(예: 년, 월, 일 등)를 불용어(Stopwords)로 처리해 향후 진행할 토픽 모델링 분석이 더욱 원활히 진행될 수 있

도록 했다.

4.2.3. 데이터 분석

본 연구 진행을 위해 수집된 데이터로부터 1차적으로 키워드 분석을 시행해 가장 많이 언급되고 있는 주제를 추출했다. 이후 가장 많이 언급되고 있는 주제를 가지고 토픽 모델링 기법의 방법론 중 가장 성능 좋은 방법론으로 인식되고 있는 MALLET의 잠재 디리클레 할당(Latent Dirichlet Allocation, LDA) 알고리즘을 활용해 토픽 모델링을 시행했다. 샘플링을 3,000회 반복하고, 각각 10개의 주제를 추출했으며, 그 결과 행복과 불행 각각의 영역에서 나타난 핵심 키워드로부터 유추된 토픽에 대해 특징을 확인하고 공통점과 차이점을 비교·분석하고자 했다. 토픽 모델링 결과 분석에 있어, 유추한 토픽은 데이터 수집 및 분석 기간 동안 이슈가 되고 있는 사항들을 참조하였다. 또한 토픽 선정 시 도출된 핵심 키워드 중 서로 연관성 있는 의미들을 포함하고 있다 생각되는 단어들을 중심으로 연구자들 간의 토의 과정을 거쳐 공통적으로 합의되는 토픽을 추출함으로써 주관성의 여지를 최소화하고자 했다는 점을 밝히고 넘어가고자 한다.

V. 분석 결과 및 행복 증진 방안

5.1. 행복 주제 구성 분석

본 절에서는 키워드 분석, 토픽 모델링 수행 과정을 통해 추출된 ‘행복’ 영역에서의 결과에 대해 이야기하고자 한다. 또한 ‘행복’ 영역에 가장 큰 비율로 도출된 키워드인 ‘행복 요인’ 데이터를 토대로 2차 세부 토픽 모델링을 수행하여, 이로부터의 결과를 더욱 자세히 논의한다.

⁶⁾ 꼬꼬마 프로젝트는 서울대학교 IDS (Intelligent Data Systems) 연구실에서 자연어 처리를 하기 위한 다양한 모듈 및 자료를 구축하기 위한 과제로 크게 ‘형태소 분석기 및 자연어 처리 모듈 개발’ 부분과 ‘세종 말뭉치 활용 시스템’으로 구분된다. (<http://kkma.snu.ac.kr/>).

5.1.1. 행복 주제별 키워드 분석

다음은 ‘행복’과 그에 관련된 키워드(Keyword)들을 웹 포털에서 검색 후 토픽 모델링을 통해 적합한 토픽을 유추한 결과이다. 토픽모델링을 통해 <표 3>과 같은

핵심 키워드가 추출되었다. 이와 같이, 조건, 안정된 삶, 복지, 진로, 주변인(이웃)과의 소통, 경제 관련 토픽들이 주를 이루고 있다.

<표 3> 행복 주제 별 키워드 분석 결과

	토픽	핵심 키워드	빈도	비율(%)
1	행복요인	행복, 삶, 돈, 요인, 조건, 건강, 수명, 인간, 인생, 무엇	8,347	19.7
2	안정된 삶	행복, 조건, 요인, 안정, 복지, 삶, 소득, 돈, 직업, 수명	6,570	15.5
3	지역사회복지	행복, 복지, 지역, 도시, 사업, 사회, 지원, 나눔, 주민, 추진	4,666	11.0
4	진로	행복, 교육, 취업, 직업, 꿈, 진로, 대학, 여성, 기업, 행복교육	4,203	9.9
5	행복소통	행복, 가족, 우리, 사람, 친구, 생각, 사랑, 시간, 마음, 엄마	3,874	9.1
6	행복정책	행복, 박근혜, 정치, 국민행복, 사회, 분배, 복지, 정책, 국민, 경제	3,442	8.1
7	행복한 노후	행복, 저축, 소득, 연금, 노후, 국민, 설계, 수명, 은퇴, 생활	3,422	8.1
8	경제, 물가	행복, 나라, 물가, 가격, 만원, 부동산, 이벤트, 은행, 서비스, 제품	3,177	7.5
9	주택공급	행복, 주택, 부동산, 임대, 주거, 금융, 대출, 초년생, 신혼부부, 사회	2,641	6.2
10	운세	행복, 오늘, 건강, 운세, 뉴스, 사랑, 결혼, 지출, 가족, 음력	2,098	4.9
총 계			42,440	100.0

전체 출현 빈도에서는, 행복요인이 총 42,440건 중 8,347건으로 19.7% 등을 차지해 ‘행복을 위한 근원적 조건’에 대해 고민하고 있는 주제가 가장 큰 비율을 차지하고 있는 것을 알 수 있다. 추후 절에서 ‘행복요인’ 대한 2차 세부 토픽 모델링을 수행해 자세한 주제 분석을 할 예정이다. 또한 안정된 삶, 지역사회복지, 진로, 행복소통, 행복정책, 행복한 노후, 주택공급 등 다양한 영역의 주제가 약 5~10% 이상으로 고르게 분포되었다. 즉, 행복한 삶을 위해서는 주변인(이웃)과의 소통, 경제, 복지 요소 등이 다양한 측면에서 중시되고 있다.

먼저 ‘행복소통’에서는 행복, 가족, 우리, 친구, 마음 등의 키워드가 도출되어 행복을 위해서는 가족, 친구 등 주변과의 진솔한 대화가 중요함을 알 수 있다. ‘행복한 노후’ 토픽에서는 행복, 저축, 소득, 연금, 노후, 설계, 은퇴와 같은 행복한 노후 대비를 위한 키워드가 출현했다. 우리나라가 급속히 고령화 사회로 진입함에 따

라, 노후 준비에 대한 관심이 집중되고 있음을 알 수 있음을 알 수 있는 대목이다. 그 외에도 ‘행복요인’과 같이 행복의 요인을 근원적으로 탐구하기 위한 토픽이 발견되었다. 실제로, 데이터 수집 기간 동안 휴식, 행복을 모토로 한 도서가 베스트 셀러였다는 사실이, 사람들의 주요 관심사가 행복과 힐링(Healing)임을 여실히 반영한다. 또한 ‘경제, 물가, 주택 공급’과 같은 경제적 요소와 관련된 토픽이 2차례 도출되어, 행복한 삶 영위를 위해서는 기본 이상의 경제력 확보가 매우 중요하게 생각된다는 사실을 알 수 있다. 비단 경제분야뿐만 아니라, 지역사회복지와 같은 ‘복지 요소’의 토픽이 추출되었다. 따라서 운택한 ‘복지’가 국민의 높은 삶의 질과 이어지고 있음을 볼 수 있다. 마지막으로 ‘진로’ 토픽이 추출되어, ‘미래’, ‘꿈’, ‘교육’, ‘대학’ 등의 키워드가 출현하고 있다. 그러므로 진로는 개인의 미래, 꿈과 행복을 좌우하는 중요한 매우 요소이고, 교육이 매우 큰 부분을 차

지하는 것으로 판단된다.

5.1.2. 행복 요인 세부 주제 토픽 모델링 분석

위 절에서 ‘행복’에 대한 키워드 분석 수행 결과, ‘행복 요인’에 대한 키워드가 웹 상에서 가장 많이 회자되고 있는 것이 확인되었다(<표 3> 참조). 고로, 본 절에서는 ‘행복 요인’에 대해 어떠한 키워드가 중점적으로 언급되고 있는지 탐색하기 위해 2차 세부 토픽 모델링을 수행하여 더욱 깊이 ‘한국인의 행복 증진을 위한 요인’을 탐색 코자 했다.

그 결과 <표 4>와 같은 토픽과 키워드가 도출되었다. 모델링 시 도출 시도한 10가지 토픽 중 해석내용으

로 다소 불충분한 2가지 토픽을 제외 후, 행복 마음가짐, 불평등 해소, 안정된 직장, 자녀 성장, 문화 생활, 수입, 종교 생활, 건강과 같은 토픽을 도출했다. 즉 행복한 삶 영위를 위해서는 만족스럽고 기쁜 마음(행복 마음가짐), 충분한 경제력(안정된 직장, 수입), 복지(불평등 해소), 가족(자녀 성장), 튼튼한 건강(건강), 문화 및 종교 생활(문화 생활, 종교 생활) 등이 필요함을 알 수 있다. 특징적으로 ‘불평등 해소’, ‘안정된 직장’, ‘수입’ 등 ‘경제’와 관련된 토픽이 8개 중 2개를 차지하는 것으로 보아, ‘충분한 경제력’ 이 행복 요인에 굉장히 중요시되고 있음을 알 수 있다.

<표 4> 행복요인 세부 토픽 모델링 결과

	토픽	키워드
1	행복 마음가짐	삶, 생각, 만족, 마음
2	불평등 해소	행복, 분배, 소득, 복지, 불평등, 소비, 자본주의
3	안정된 직장	요인, 소득, 경제, 국가, 결정, 중요
4	자녀 성장	실업, 취업, 직업, 미래, 학생, 대학, 결혼
5	문화 생활	문화, 책, 교수, 저자, 예술
6	수입	돈, 투자, 중요, 희망, 관리, 부자
7	종교 생활	하나님, 교회, 말씀, 목사, 예수
8	건강	수명, 건강, 치료, 부부, 스트레스, 운동
9	-	책, 저자, 리뷰, 지음, 출간
10	-	행복수준, 테스트, 심리, 발표, 욕구

5.2. 불행 주제 구성 분석

본 절에서는 ‘불행’ 영역에서의 키워드 분석 수행 후, 추출된 주제에 대한 의미를 분석하고자 한다. 더불어, 분석 결과 가장 큰 비율을 차지한 ‘불행 요인’에 대해 2차 세부 토픽모델링을 진행해 한층 더 깊이 분석해 그에 대한 함의를 얻고자 한다.

5.2.1. 불행 주제별 키워드 분석

<표 5>는 ‘불행’ 관련 데이터 수집 후 키워드 분석을

진행한 결과이다. 그 결과 ‘불행 요인’이 전체 분포 중 1/4 이상으로 가장 비중이 크게 나타났다. 행복해지기 위한 요인이 무엇인지 끊임없이 고민하듯, 불행에 빠지는 조건 또한 고민하고 있는 것으로 생각된다. 사람, 일, 건강 등이 ‘불행 요인’의 주요 키워드인 것으로 보아, 주변인과의 사회 생활, 일, 건강 등이 주요한 불행 요소인 것으로 유추할 수 있다. 둘째, 경제적 불행감, 집 문제 등 ‘경제’ 연관 토픽이 약 22%로 뒤를 이어, 경제 또한 불행의 큰 원인임을 알 수 있다. 분배, 실업, 부동산, 아파트, 물가 등의 키워드가 경제 영역에서의 불행에 대

한 주요 키워드로 나타나고 있다. 셋째, 불안정한 미래, 과도한 사교육 등 교육과 진로에 관련된 토픽이 약 16%의 분포를 보인다. 학교, 대학, 과도, 사교육비 등이 주요 키워드로 도출됐다. 실제 우리나라는 치열한 입시 열풍으로 사교육 규모가 팽창 중이며, 사교육비 과다 지출은 가계에 큰 영향을 끼쳐 민생 경제의 큰 난관으로 작용하고 있다. 또한 지나친 경쟁교육은 교육 기회에 있어 사회 계층 간 간극을 넓히고, 위화감을 조성해 비생산적이며 소모적 결과를 낳는다. 따라서 학교 교육 내실화로 공교육 질을 높이는 기존 입시 체제에서 벗어나 다양하고 창의적 학생 선발 전형으로 전환해 사교육비 경감에 힘써야 할 것이다. 무엇보다 입시 구조 변화, 학벌 위주 풍토 개선 등의 근본적 노력이 필요하다 생각

된다.

넷째, 노후 대비 관련 토픽이 7.5%를 차지한다. 우리나라는 이미 2000년 노인 인구(65세 이상)비중이 7%를 넘기며 고령화 사회에 진입했고, 2020년에는 고령사회, 2030년에는 초 고령화 사회에 도달할 것으로 예상된다. 본 토픽의 주요 키워드로 저축, 돈, 부동산, 노후, 준비, 소득, 연금 등이 추출되는 것을 보아, 곧 닥칠 고령화 사회에 대비한 은퇴 대비가 큰 이슈화 되고 있음을 알 수 있다. 그 외 복지지원, 이웃들(주변)의 불행 토픽이 출현했고, 2014년 4월 16일 전남 진도군 인근 해상에서 침몰해 295명의 희생자를 발생시킨 세월호 사태 또한 불행 토픽 중 하나로 분석되었다.

<표 5> 불행 주제 별 키워드 분석 결과

	토픽	핵심 키워드	빈도	비율(%)
1	불행요인	불행, 행복, 사람, 일, 우리, 무엇, 직업, 건강, 때문, 요인	10,131	25.1
2	경제적 불행감	불행, 우리, 경제, 분배, 빈곤, 실업, 성장, 소득, 청년, 정부	5,728	14.2
3	불안정한 미래 (교육, 진로 등)	불행, 학교, 교육, 학업, 아이, 학생, 취업, 사회, 대학, 직업	5,017	12.4
4	세월호 사태	불행, 국민, 대통령, 정부, 세월호, 경제, 물가, 정책, 박근혜, 사태	4,859	12.1
5	이웃들(주변)의 불행	불행, 가족, 친구, 자신, 집, 아들, 사랑, 그녀, 아버지, 엄마	3,764	9.3
6	집 문제	불행, 일, 부동산, 돈, 취업, 서울, 아파트, 한국, 집, 물가	3,102	7.7
7	노후 대비	불행, 저축, 돈, 수명, 노후, 부동산, 준비, 소득, 투자, 연금	3,035	7.5
8	대한민국 불행요인	불행, 교육, 한국, 대한민국, 우리나라, 행복, 노르웨이, 국민, 지수, 사회	2,173	5.4
9	과도한 사교육	불행, 교육, 아이, 학부모, 과도, 사교육비, 경쟁교육, 학교, 문희상, 교육감	1,574	3.9
10	복지지원	불행, 복지, 지원, 긴급 복지 지원, 박근혜, 국회, 예산안, 연금, 기초, 급여, 보장	967	2.4
총 계			40,342	100.0

5.2.2. 불행 요인 세부 주제 토픽 모델링 분석

위에서 수행된 ‘불행’에 대한 키워드 분석 수행 결과, ‘불행 요인’에 대한 키워드가 가장 많이 언급되었다(<표 5> 참조). 이는 ‘행복’영역에서 ‘행복 요인’이 가장 많

이 언급되는 것과 같은 맥락이다. 따라서, 본 영역에서도 ‘불행 요인’에 대해 주로 어떠한 토픽이 도출되는지 분석하기 위한 2차 세부 토픽모델링을 수행했다. 그 결과, <표 6>와 같이 가정 불화, 가난, 자녀 문제, 부정적

인생관, 인간관계 문제, 직업 불만족, 건강 문제 등 전반적으로 경제력(가난, 미 취업), 가족 및 주변이웃(가정 불화, 인간관계 문제), 건강(건강 문제), 자아성취(직업 불만족) 등이 불행 요인의 토픽으로 추출되었다. 따라서, 불행 경감을 위해서는 충분한 경제력을 가지기 위해 노력하고, 매사에 부정적 마음가짐보다 긍정적 태

도를 가지도록 힘쓰며, 가족 및 주변이웃들과 원만하게 지내고자 하는 태도와 의지가 필요하다고 할 수 있다. 뿐만 아니라 항상 건강을 최우선 가치로 삼으며, 진정한 자아를 성취하는 삶을 누리는 것이 불행을 줄이는 길이라 할 수 있다.

<표 6> 불행요인 세부 토픽 모델링 결과

	토픽	키워드
1	가정 불화	불행, 사랑, 가족, 집, 아버지, 가정, 부모
2	가난	분배, 돈, 소득, 빈곤, 경제, 가난
3	자녀 문제	학교, 위험, 아이, 행동, 상황
4	부정적 인생관	불행, 사람, 인생, 마음, 성공, 공통점
5	인간관계 문제	불행, 자신, 관계, 마음, 생각, 환경, 상황
6	직업 불만족	불행, 사람, 생각, 인생, 직업, 친구
7	건강 문제	불행, 건강, 수명, 질병, 생명, 병, 사고
8	미 취업	오늘, 운세, 불행, 건강, 취업, 뱀띠, 금전
9	부정적 마음가짐	불행, 사람, 마음, 생각, 이기심, 자만심, 피해의식
10	-	예수, 교회, 신앙, 설교, 말씀, 축복

5.3. 행복 증진 및 불행 경감 방안

본 연구는 행복과 불행 요인에 대해 도출된 2차 세부 토픽 모델링 결과를 기반으로 정부, 기업, 가정, 교육기관, 사회복지기관의 5가지 측면에서 행복 증진 및 불행 경감을 위한 실무적 실행 방안을 제안한다(<그림 2> 참조). 연구자들의 토의를 통해 선정한 각 토픽들에 대해, 당시 현안이 되고 있던 이슈들이 무엇인지 파악했으며, 그에 대한 다수 문헌 검토를 통해 근본적인 해결책이라 주장되는 방안들을 제시했다.

첫째, 정부 측면에서는 스타트업(Start up) 환경을 위한 창업 자금 지원 및 강력한 인프라 제공이 필요하다. 현 정부의 핵심 정책인 ‘창조 경제’는 다소 모호한 개념으로 ‘뜯구름 잡기’ 아니냐는 비판과 함께, 정부 주도의 경제 정책은 자율성 보다 많은 규제로 혁신에 해

가 될 수 있다는 우려가 존재했다. 일각에서는 창조 경제의 현실적 실행을 위해서는 공공 데이터 개방 확대를 통한 공공정보 기반 신 비즈니스 창출 및 촉진, 목표 지향형 대상자 별 맞춤형 교육 훈련 시스템 구축 등 개혁의 필요성을 촉구하는 등 많은 의견이 분분하다. 그 중 최근 연구 결과, 스타트업에 대해 정부의 창업 자금 지원 및 인프라 지원 정책이 가장 부족한 것으로 나타났다(김영환, 2015). 스타트업 성장은 우리나라의 고질적 문제인 과도한 대기업 의존을 어느 정도 해결하는 방편이 될 수 있을 것이고, 취업하여 안정을 원하는 청년층에게도 해답이 될 수 있으므로 적극적 지원이 필요하다. 또한 투자 컨설팅, 창업 노하우 전수, 판로, 법제도 등 창업 시 필요한 과정에 대해 전체적으로 신경을 쓰고 세심히 지원해 창업자의 부담을 줄이고, 활성화하는

방안의 마련이 필요하다(한국 정보화 진흥원, 2013). 또한 윗 절에서 논의되었듯, 안정된 직장, 직업 불만족, 미취업이라는 토픽들이 추출될 정도로 온라인 상에서 ‘직업’에 관한 많은 이야기가 오고 가고 있으며, 이것이 행복과 직결되는 큰 문제임을 알 수 있다. 이를 타개하기 위해 정부는 적극적, 합리적인 고용정책을 마련해 경제적, 사회적 불평등을 해소하는 노력을 기울이는 것이 바람직하다. 2015년 7월 정부는 3년간 20만명 일자리 마련을 목표로 하는 ‘청년 고용절벽 해소 종합대책’을 발표했지만, 현실적으로 이행이 어려운 ‘뺨뿔기 정책’이라는 비판을 받고 있는 실정이다. 정부는 이러한 문제의 근본적 근본적 원인을 파악해 본질적으로 문제를 해결할 수 있도록 해야 할 것이며, 민생의 갑갑함이 큰 만큼 그들의 목소리를 경청하고, 시민사회와 적극적으로 소통해 실질적 대책 마련을 해야 할 것이다. 다음으로, 사회 복지 측면에서의 개선이 필요하다. 사회 복지는 사회 유지에 기본적으로 요구되는 사회, 경제, 건강, 교육의 필요조건을 충족하도록 하는 국가의 혜택 및 서비스를 연구하는 인간의 행복을 위해서 필요한 실천 학문이므로 매우 중요하다. 현재 우리 나라에서는 빈부 격차 심화, 노령화 등으로 수많은 관점에서 복지 개선을 위한 제언점들이 쏟아져나오고 있으나 실제로는 이것들을 모두 충족시키지 못하고 있다. 이의 실제적 효력 발휘와 실현을 위해서는 사회복지공무원 일자리 개선 등 일련의 행정체계 개선이 시급하다. 사회복지공무원은 국민 전체의 복지를 담당하는 중책을 맡고 있지만, 열악한 근무 환경으로 ‘근무기피직군’이 되어 왔으며, 광범위한 업무 전담으로 전문성이 다소 떨어진다는 지적을 받아 왔다. 뿐만 아니라 우리나라의 4대 보험 제도는 각각 독자적 운영 기구에서 관리되고 있는 데다 비민주적인 자금 운용 방식을 채택하고 있어 매우 비효율적이며 형평성에 어긋난다는 비판을 듣고 있다. 이러

한 행정 체계가 우선적으로 개선된다면, 시민들의 사회적 불평등 또한 어느 정도 해소될 것은 자명한 일이다. 또한 우리나라는 세계에서 가장 빠른 속도의 ‘저출산과 고령화’를 맞이하며, 이에 따른 사회 복지 서비스의 발빠른 대처가 강력히 요구되고 있는 실정이다. 80세 이상 초고령 노인 증가는 경제 생활, 건강 문제 등을 초래 중이며 핵심적 사회 복지 서비스의 포커스가 노인이 되어야 함을 반증하고 있다 따라서 정부는 노인을 위한 양질 일자리 확충, 안정적 연금 계획 수립 등으로 국민 행복을 증진시킬 수 있도록 해야 한다(윤석명, 2014). 뿐만 아니라 타 사회 취약 계층, 즉 아동 또는 빈곤층을 위한 충실한 사회 복지 서비스 확충이 필요하다(노년복지연합, 2013).

다음으로 기업 측면의 행복 증진 및 불행 경감 방안 에 대해 살펴보고자 한다. 기업은 토지·노동·자본을 이용해 재화를 만드는 생산 활동 주체로서 주 목적은 ‘이윤 추구’이다. 기업이 얻는 이윤이 높아지면 국가경제 또한 탄탄해지고, 국민의 소득 수준 또한 증가하므로 큰 중요성을 가진다. 또한 대부분의 근로자들이 하루 1/3 이상을 일하며 보내고 있으므로, 일터에서의 행복은 굉장히 중요하다. 따라서 원활한 커뮤니케이션으로 갈등은 줄이고 상호 배려는 늘이는 기업문화를 도입해야 한다. 기업은 구성원들이 의욕적으로 일할 수 있는 환경을 조성해 그들이 기업의 발전을 위해 적극적으로 기여할 수 있도록 해야 한다. 다음으로 뜨거운 쟁점이 되고 있는 비정규직 실태에 대해 기업 차원의 실효성 있는 정책 수립이 필요하다. 비정규직은 2012년부터 증가하고 있으나, 상용직에 비해 낮은 임금, 부족한 복지와 일터 내 차별행위로 인해 사회적 이슈화 되고 있다. 그들을 위한 복지 확충, 정규직 이행 기회 확대 등 방안을 마련해, 근로자들이 행복을 누릴 수 있도록 해야 할 것이다(김복순과 정현상, 2016).

⁷ 서울경제신문(<http://m.post.naver.com/viewer/postView.nhn?volumeNo=1977563&memberNo=22213349&vType=VERTICAL>), 청년고용정책은 ‘우려먹는’ 사골정책

셋째, 가정 측면에서 행복 증진 및 불행 경감 방안을 이야기하고자 한다. 가정은 세상에 태어나 사회생활과 생활양식을 배우며 인격을 완성하며 행복을 추구하는 곳이다. 예로부터 ‘가화만사성(家和萬事成)’이라는 말이 있듯, 가정은 사회와 국가의 뿌리로 그 중요성이 크지만 현대에는 가족 간 공감대가 줄어들고 파편화되어 가정 분열이 생기는 등 문제점이 많다. 그러므로 우선적으로, 구성원 간 소통 노력으로 갈등을 개선해 행복을 증진시키려는 노력이 필요하다. 또한 가정에서부터 속 깊은 대화로 진정 소중한 ‘가치’를 탐색하는 시간을 많이 가진다면, 우리 사회의 팽배한 ‘물질주의’, ‘금전주의’

를 물리치고 차츰 행복에 대한 인식을 변화시킬 수 있으리라 생각된다. 더불어 가정 교육을 통한 습관은 굉장히 중요하다. 균형 있는 식생활, 바른 자세와 언어 사용 등 가정에서의 평상시 바른 생활습관을 통해 건강하고 올바른 삶의 자세를 가지는 것이 중요하다. 그리고, 가족은 동일한 종교를 믿으며 그들의 신념과 가치관을 공유한다. 종교는 인생의 의미와 목적을 제공하는 등 긍정적 역할을 하지만 자칫 과도한 종교에의 의존은 본업에 소홀해지거나 타 종교에 배타심을 갖게 되는 계기가 될 수 있다. 이에 가정에서는 ‘종교’에 대한 올바른 태도를 가질 수 있게끔 지도해야 한다.



<그림 2> 사회 각 분야에서의 행복 증진·불행 경감 방안

다음으로, 교육기관에서는 자아탐색교육을 활성화 해 학생들이 스스로에 대해 더 잘 알고, 행복하게 살아갈 수 있도록 하는데 기여해야 할 것이다. 그리고 어릴 때부터 체계적 경제 교육 시행해, 우리 사회의 물질주의적 인식에 변화를 주고, 다른 영역에서 행복을 탐색

할 수 있게끔 노력해야 한다. 유년기부터 금전을 통해 무엇을 얼마나 '소유'하느냐 보다, 어떤 '경험'하느냐가 중요하다는 인식을 갖도록 해야 한다. 또한 우리나라는 세계적 '사교육 공화국'으로 학생들은 오로지 명문대 진학을 위한 과다 경쟁에만 매진한다. 경쟁에서 승리하

면 출세할 수 있겠지만, 외려 본연의 모습은 알지 못하고, 진짜 행복이 무엇인지 알지 못한 채 한평생을 살게 된다. 따라서 교육 기관에서는 자아를 성취하여 행복을 누리는 방법이 무엇인지 교육해야 한다. 그리고 공교육을 내실화하여 사교육을 줄이기 위한 노력이 필요하다. 마지막으로, 대학·연구기관에서는 다양한 융합 다 학제적 연구를 통해 실질적 경제 회생 방안, 교육법 개선 등 다양한 방향으로 행복 증진 방안을 연구해야 할 것이다.

끝으로 사회복지기관에서는 불우이웃을 위한 멘토링 활동, 봉사 활동 등을 개최하여, 치열한 사회에서 나눌 틈 없이 살았던 사람들에게 서로 소통을 하고 마음을 열어 행복을 증진시킬 수 있는 기회를 가질 수 있도록 한다. 실제로 최근 많은 사회복지기관에서 ‘장기 기증’, ‘불우이웃 연탄 배달’ 등 행사를 활성화하며 나눔의 기쁨을 전파하고 있다. 봉사자들 또한 자신의 조그마한 희생과 나눔의 정성 자체가 타인에게 엄청난 힘이자 행복으로 돌아온다는 것을 깨닫고 봉사에 적극적으로 임하는 자세를 가지는 것이 중요하다. 또한, 사회복지기관에서 앞장서 여러 분야(인문학, 철학, 예술 등)에서의 강좌 및 공연 등을 마련하는 기회를 만들 수 있도록 해야 한다. 이는 우리 삶의 질을 고양하는 좋은 기회가 될 것이다. 마지막으로, 이미 고령화 사회에 진입해 노인 복지, 소외 문제 등 이슈가 발생하고 있는 시점에서, 사회복지기관이 앞장서 노인을 위한 건강과 여가 생활 등에 앞장선다면, 더욱 밝은 사회가 될 것이다.

VI. 시사점 및 결론

6.1. 학술적 시사점

본 연구의 학술적 시사점은 다음과 같다. 첫째, 본 연구는 행복 요인 도출을 위해 줄곧 활용되어온 정성적, 정량적 연구법과는 달리, 데이터 어널리틱스(Data

Analytics) 기반의 토픽 모델링 기법으로 새로운 행복과 불행의 요인을 탐색적으로 발견하고자 했다는 점에서 의의를 지닌다. 방대한 데이터로부터 핵심 주제를 찾아내고자 하는 토픽 모델링 기법을 통해, 민중들의 의견들을 반영하고 있는 웹 데이터를 기반으로 현 사회의 트렌드 마이닝을 행했다는 점에서 차별화되는 것이다. 이는 기존 OLAP와 더불어 여러 가지 단편적 데이터로부터 지식을 창출해 이를 현 사회가 당면한 문제에 적용, 해결하고자 하는 ‘지식경영’의 취지와도 부합된다는 측면에서도 의미가 있다. 행복 요인에 관련한 기존 연구들은 주로 설문 방식을 활용하거나, 소득, 직업, 거시적 경제 지표 등 객관적 요인에 한정해 분석되어 왔으며 (김명소 외 2003), 행복 지수 연구 또한 경험적으로 탐색된 한국인의 행복 요인을 기초로 제작되거나, 경제적 변수를 기반으로 행복지수를 공식화한 수준에 그친다 (김명소 외 2003; 국민통합 이슈 모니터링; 한국의 국민행복지수에 관한 연구, 2013). 하지만 본 연구는 이전 연구와 달리 토픽 모델링을 활용해 실제 한국 사회에서 살아가고 있는 민중들의 ‘행복’, ‘불행’ 관련 의견을 실제로 마이닝하고, 이로부터 행복 증대 및 불행 경감 방안을 논의하고자 했다는 점에서 학술적 시사점을 가진다.

둘째, 본 연구는 ‘행복’, ‘불행’ 각 영역에서의 토픽뿐만 아니라, 2차 세부 분석을 통한 ‘행복 요인’ 및 ‘불행 요인’에 대해 심층적으로 연구했다. 현 시대 한국인들이 생각하고 있는 ‘행복 요인’, ‘불행 요인’에 대해 깊숙이 탐구하고, 그를 통한 행복 증진 방안 및 불행 경감 방안을 모색하고자 했다는 점에서 중요성을 가진다.

끝으로, 본 연구는 ‘행복’, ‘불행’이라는 인문학적 가치와 데이터 어널리틱스 분야를 결합한 융복합적 연구의 성격을 지닌다. 고대부터 끊임없이 탐구되어온 ‘행복’이라는 근원적 가치에 데이터 어널리틱스라는 새로운 연구법을 접목한 융복합적 연구로, ‘삶의 질 증진’을 함께 추구했다는 점에서 기여점을 가지는 것이다.

6.2. 실무적 시사점

본 연구는 행복한 사회 건설을 위해 노력 중인 한국인들을 위해 다음과 같은 실무적 시사점을 제공한다. 첫째, 본 연구는 한국인들이 더욱 살기 좋은 사회를 만들어 나가기 위해 사회를 구성 중인 정부, 기업, 가정, 교육기관, 사회복지기관 측면에서 행복 증진 및 불행 경감 방안을 현실적, 구체적인 시각에서 제안했다. 김명소 et al. (2003)은 우리 국민의 행복을 위해서는 '성장'에 대한 주력이 필요하다 이야기했으나, 실질적 행복 증진 방안에 대해 논의하지 않았다. 본 연구는 사회 각계에서의 다양한 관점에서 행복 증진과 불행 경감 방안을 논의해, 한국인들의 행동 방향을 구체적으로 제안했다는 점에서 차별점을 지닌다.

다음으로, 본 연구는 세부 2차 토픽 모델링을 통해 행복 요인뿐만 아니라, 불행 요인을 분석했다는 측면에서 실무적 시사점을 가진다. 양 요인을 비교 분석한 결과, 공통적인 토픽으로 마음가짐, 경제력, 가족, 건강이 출현했다. 따라서 위 4가지 가치에 더욱 중점을 두어, 긍정적 마음가짐, 풍족한 경제력, 원만한 가족 관계, 건강을 목표로 살아간다면 행복한 삶 영위는 물론 삶의 질 향상에 큰 도움이 되리라 생각된다.

6.3. 결론

본 연구는 토픽 모델링을 통해 한국인이 실제로 생각하고 느끼고 있는 행복, 불행의 요인을 탐구하고, 이를 바탕으로 행복을 증진시키고 불행을 경감시키기 위한 방안을 제시하고자 하였다. 연구의 결과를 통해 분석된 행복 요인과 불행 요인 및 이를 통한 제안 사항을 요약하면 다음과 같다. 먼저, 1차 키워드 분석 결과, 행복 영역 주제로는 행복 요인, 행복 소통, 경제·물가 등이, 불행 영역의 주제로 불행 요인, 과도한 사교육, 이웃들의 불행 등이 분석되었고, 양 영역을 비교한 결과, 가장 큰 빈도로 행복 요인, 불행 요인이 추출되었다. 행복 또는 불행해지는 요인이라는 면에서 공통점이 존재했

기에, 행복 요인과 불행 요인의 데이터에 한해 2차 세부 토픽 모델링을 수행해 더욱 엄밀히 비교 분석하고자 했다. 그 결과, 행복 요인에서는 행복 마음가짐, 안정된 직장, 자녀 성장, 수입, 건강 등이, 불행 요인에 대해서는 가정 불화, 가난, 인간관계 문제, 건강 문제 등의 토픽들이 확인되었다. 이를 토대로 정부, 기업, 가정, 교육기관, 사회복지기관 측면에서 행복을 증진하고 불행을 경감시키기 위한 실질적 방안을 제안해(<그림 5>참조) 한국인의 삶의 질이 향상되는 계기를 마련코자 했다.

하지만 본 연구는 향후 보완 필요한 몇 가지 한계점을 가지고 있다. 첫째, 본 연구의 분석 데이터로 활용된 웹 문서는 체계적 전처리 과정을 수행했음에도 불구하고, 상품 프로모션과 같은 광고 성 데이터를 완벽히 제거하지 못했기에, 결과를 일반화하기에는 어느 정도 제약이 있다. 향후 상업적 데이터를 완벽히 정제해, 더욱 신뢰성 있는 탐구를 진행할 필요가 있다. 또한 본 연구는 웹 포털에 포스팅된 데이터를 추출하여, 그 결과를 분석하였다. 하지만 향후에는 대중들의 의견이 더욱 생생히 반영되어 있는 소셜 미디어 데이터를 추가적으로 분석한 심도 있는 연구를 꾀해야 할 것이다. 셋째, 본 연구에서는 Alderfer의 생존-관계-성장 욕구 위계 이론(ERG Theory)을 기초로 선정한 26가지 키워드로 연구를 진행했다. 하지만 토픽 모델링 이전 기 키워드에 대한 타당성 검증이 다소 미흡했다. 또한 토픽 모델링 수행 후 연구자들간의 토의를 토대로 선정한 주제에 대한 체계적, 객관적인 타당성 검증이 이루어지지 못한 부분이 있다. 따라서 추후 토픽 모델링 분석을 토대로 한 연구에서 타당성 검증에 더욱 힘을 기울여 연구의 완성도를 높여야 할 것이다. 추후 오픈한글과 같은 감성어사전을 이용하여 (안정국, 김희웅 2015), 감성어 분석도 필요하다. 활용마지막으로 본 연구는 특정 시점(2013년~2014년 12월)의 데이터를 통해 행복 및 불행 요인 분석을 했으나, 미래에는 다양한 시점의 데이터를 수집해 시계열적 분석을 한다면, 온라인 매체 상에서 생산되는 데

이터에 대한 트렌드 비교가 가능하므로 더욱 흥미로운 시사점을 제공하리라 기대한다.

참 고 문 헌

[국내 문헌]

1. 국가미래연구원, “한국의 국민행복지수에 관한 연구“, 2013, pp. 1-113
2. 국민대통합위원회, “국민통합 이슈모니터링“, 2014, Vol.3. pp.1-54
3. 강범일, 송민, 조화순, “토픽 모델링을 이용한 신문 자료의 오피니언 마이닝에 대한 연구“, 한국문헌정보학회지 (47:4), 2013, pp. 315-334.
4. 권석만, “심리학의 관점에서 본 욕망과 행복의 관계“, 철학사상, 2010
5. 김명소, 김혜원, 차경호, 임지영, 한영석, “한국 성인의 행복한 삶의 구성요인 탐색 및 척도개발“, 한국심리학회지: 건강 (8:2), 2003, pp. 415-442
6. 김명소, 한영석, “한국인의 행복지수 공식 개발“. 조사연구 7(2), 2006, pp. 1-38.
7. 김복순, 정현상, “최근 비정규직 노동시장의 변화“, 노동리뷰, 2016
8. 김영환, “창업 및 정부의 창업지원 정책에 대한 대중 인식 조사“. 과학기술정책, (25:6), 2015, pp. 36-39.
9. 김정선, 권은주, 송태민, “분석지의 확장을 위한 소셜 빅데이터 활용연구 - 국내 ‘빅데이터’ 수요공급 예측 -, 지식경영연구 (15:3), 2014, pp. 169-188
10. 노년복지연합, 사회복지정책의 문제점 및 개선 방안, 2013
11. 동아일보 특별 취재팀, “돈버는 기계 같아”... 행복 느낄 틈도 없는 고소득 30대男”, 동아일보, 2015
12. 박건숙, “빅데이터의 하위 주제어 의미 분석 연구“, 언어학, 2013
13. 박자현, 송민, “토픽 모델링을 활용한 국내 문헌 정

보학 연구 동향 분석”, 정보관리학회지 30(1), 2013, pp. 7-32.

14. 영국 BBC 다큐멘터리 행복 {How to be Happy}, 예담, 2006
15. 배정환, 손지은, 송민, “텍스트 마이닝을 이용한 2012 년 한국대선 관련 트위터 분석”, 지능정보연구 (19:3), 2013, pp. 141-156.
16. 안정국, 김희웅, “집단지성을 이용한 한글 감성어 사전 구축”, 지능정보연구, (21:2), 2015, pp.49-67.
17. 안정국, 이규현, 김희웅, “정보시스템 연구 트렌드 변화 분석: 토픽모델링과 네트워크 분석”, 한국경영정보학회 추계통합학술대회, 2015
18. 윤강재, 김계연, “OECD 국가의 행복지수 산정 및 비교”. 보건복지포럼 159, 2010, pp.86-98
19. 윤석명, “OECD 국가의 인구 고령화 대책 분석 -노후소득보장제도 개편이 주는 시사점 중심으로”. 지방재정 (2014:6), 2014, pp. 116-135.
20. 차윤정, 이지혜, 최지은, 김희웅, “소셜미디어 토픽모델링을 통한 스마트폰 마케팅 전략 수립 지원”, 지식경영연구 (16:4), 2015, pp. 69-87
21. 최돈정, 민무홍, 김재광, 이지형, “마이크로블로그를 이용한 토픽 추적에 관한 연구”. 한국지능시스템학회 2011 년도 춘계 학술대회, 2011
22. 추홍규, “개인의 행복지수 수준과 결정요인”. 인제대 대학원 박사논문, 2005
23. 한국 정보화 진흥원, 창조 경제 실현을 위한 ICT의 새로운 역할과 과제, 2013

[국외 문헌]

1. Blei, D., A. Ng, and M. Jordan, “Latent Dirichlet Allocation,” Journal of Machine Learning Research (3), 2003, pp. 993-1022.

2. Blei, D., "Probabilistic topic models" *Communications of the ACM*, 2012
3. Campbell, A., "Subjective Measures of Well-being", *American Psychologist* 2, 1976, pp. 117-124.
4. Chiasson, N. Dube, L., and Blondin, J., "A Look into the Folk Psychology of Four Cultural Groups." *Journal of Cross-Cultural Psychology*, (27), 1996, pp. 673-691.
5. Diener, E., and Emmons, R.A., "The Independence of Positive and Negative Affect." *Journal of Personality and Social Psychology* , (47), 1984, pp. 1105-1117.
6. Diener, E., Sandvik, E., Seidlitz, L., Diener, M., "The Relationship between Income and Subjective Well-Being: Relative or Absolute?" *Social Indicators Research* , (28:3), 1993, pp. 195-223.
7. Diener, E., Suh, E, Lucas, R., and Smith, H., "Subjective Well-being: Three Decades of Progress." *Psychological Bulletin* , (125), 1999, pp. 276-302
8. Gam, M. A. and M. Song., "A Study on Differences of Contents and Tones of Arguments among Newspapers Using Text Mining Analysis", *Journal of Intelligence and Information Systems* , (18:3), 2012pp. 53-77.
9. Gerrish, S., and Blei, D., "A language-based approach to measuring scholarly impact.", *The 27th International Conference on Machine Learning*, 2010, pp. 375-382.
10. Grimmer, J., "A Bayesian hierarchical topic model for political texts: Measuring expressed agendas in senate press releases." *Political Analysis*. (18:1), 2010, pp. 1-35.
11. Lee, D., Park, S., Uhlemann, M., and Patsula, P., "What Makes You Happy?: A Comparison of Self-reported Criteria of Happiness between Two Cultures." *Social Indicators Research* , (50), 1999, pp. 351-362.
12. Mimno, D., Wallach, H.M., McCallum, A., "Gibbs Sampling for Logistic Normal Topic Models with Graph-Based Priors", Unpublished papers, 2008
13. Newman, D., and Block, S., "Probabilistic Topic Decomposition of an Eighteenth-Century Newspaper.", *Journal of the American Society for Information Science and Technology*, (57:5), 2006, pp. 753-767.
14. HT Reis, SL Gable., "Toward a positive psychology of relationships.", *American Psychological Association*, 2003, pp.129-159
15. Ryu, W. J., J. W. Ha, Md. Hijbul Alam, and S. K. Sang., "Extracting Trends from Twitter using a Topic Modeling Technique," *Proceedings of Korea Computer Congress*, 2013, pp. 191-193.
16. Song, Min., and Kim, Suyeon., "Detecting the knowledge structure of bioinformatics by mining full-text collections." *Scientometrics*, (96:1), 2013, pp. 183-201.
17. Steyvers, M. and Griffiths, T., "Probabilistic topic models", *Handbook of latent semantic analysis*, 2007
18. P Shaver, J Freedman, "Your pursuit of happiness", *Psychology Today*, 1976
19. Zhao, W.X., Jiang, J., He, J.W.J., Lim, E.P., Yan.

H., Li. X., “Comparing Twitter and Traditional Media Using Topic Models”, Lecture Notes in Computer Science , 2011, pp 338-349

저 자 소 개



양 승 준 (Seung-Joon Yang)

‘연세대학교 정보대학원에서 디지털 경영 석사과정 졸업 후 현재 SLcoms 기업부설 연구소에서 근무 중이다. 주요 관심분야는 Big Data Analytics, Econometrics, Machine Learning 이다.’



이 보 연 (Bo-Yeon Lee)

‘중앙대학교 컴퓨터 과학과 졸업 후, 현재 연세대학교 정보대학원에서 디지털 비즈니스 석사 과정 재학 중이다. 롯데 정보통신에서 근무한 바 있으며, 주요 관심 분야는 Big Data Analytics, Social Media Mining, Digital Business이다.’



김 희 웅 (Hee-Woong Kim)

‘National University of Singapore 정보시스템학과에서 에서 근무 후, 현재 연세대학교 정보대학원 교수로 근무 중이다. 주요 관심분야는 디지털 비즈니스, 정보시스템 관리 및 활용 등이다. 관련 연구들은 MIS Quarterly, Information Systems Research, Journal of Management Information Systems, Journal of the Association for Information Systems, IEEE Transactions on Engineering Management, Journal of Retailing, European Journal of Operational Research, Communications of the ACM 등에 40여편의 논문이 게재되었다.’